



Optimisez les coûts des charges de travail Microsoft sur AWS

AWS Conseils prescriptifs



AWS Conseils prescriptifs: Optimisez les coûts des charges de travail Microsoft sur AWS

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Introduction	1
Présentation de	1
Public ciblé	1
Comment utiliser ce guide	2
Résultats commerciaux ciblés	3
Parcours d'optimisation des coûts	4
Les meilleures recommandations pour optimiser les coûts	7
Présentation	7
Les meilleures recommandations	7
AWS Optimisation et évaluation des licences	9
Présentation	9
Options d'évaluation	10
Évaluation complète	10
Élargir les charges de travail	11
Collectez des données	11
Analyser les données	12
Planifier les prochaines étapes	14
Impact de l'évaluation	16
Étapes suivantes	16
Ressources supplémentaires	16
Windows sur Amazon EC2	17
Automatisez les plannings d'arrêt et de démarrage	18
Présentation de	18
Études de cas	18
Scénario d'optimisation des coûts	19
Recommandations en matière d'optimisation des coûts	21
Ressources supplémentaires	34
Charges de travail Windows de la bonne taille	35
Présentation de	35
Scénario d'optimisation des coûts	35
Recommandations en matière d'optimisation des coûts	36
Recommandations	46
Ressources supplémentaires	47
Sélectionnez le type d'instance adapté aux charges de travail Windows	47

Présentation de	47
Recommandations en matière d'optimisation des coûts	48
Étapes suivantes	57
Ressources supplémentaires	59
Apportez des licences pour les charges de travail Windows et SQL Server	59
Présentation de	59
Hôtes dédiés Amazon EC2	60
AWS options de licence	65
Apporter des licences Windows Server	65
Scénarios d'optimisation des coûts	67
Recommandations en matière d'optimisation des coûts	73
Ressources supplémentaires	73
Optimisez les dépenses pour Windows sur Amazon EC2	74
Présentation de	74
Comprendre les Savings Plans	75
Scénarios d'optimisation des coûts	81
Recommandations en matière d'optimisation des coûts	85
Ressources supplémentaires	87
Surveillez les coûts à l'aide d' AWS outils	87
Présentation de	87
Recommandations en matière d'optimisation des coûts	88
Ressources supplémentaires	92
SQL Server	93
Choisissez une solution de haute disponibilité et de reprise après sinistre	94
Présentation de	94
Groupes de disponibilité SQL Server Always On	95
Instances de cluster Always On Failover de SQL Server	97
SIOS DataKeeper	99
Groupes de disponibilité Always On	101
Groupes de disponibilité distribués	102
Expédition de journaux	103
AWS Database Migration Service	105
Reprise après sinistre AWS Elastic	106
Comparaison des coûts	107
Recommandations en matière d'optimisation des coûts	112
Ressources supplémentaires	113

Comprendre les licences SQL Server	113
Présentation de	113
AWS options de licence	114
Impact financier de l'acquisition de licences	115
Optimisation des licences	115
Recommandations en matière d'optimisation des coûts	116
Ressources supplémentaires	47
Sélectionnez l'instance EC2 adaptée aux charges de travail SQL Server	122
Présentation de	123
Comparaison des coûts	123
Scénario d'optimisation des coûts	125
Recommandations en matière d'optimisation des coûts	126
Ressources supplémentaires	130
Consolider les instances	130
Présentation de	130
Scénario d'optimisation des coûts	131
Recommandations en matière d'optimisation des coûts	133
Ressources supplémentaires	134
Comparez les éditions de SQL Server	134
Présentation de	134
Impact sur les coûts	135
Recommandations en matière d'optimisation des coûts	137
Ressources supplémentaires	143
Évaluer l'édition SQL Server Developer	144
Présentation de	144
Impact sur les coûts	144
Ressources supplémentaires	47
Évaluer SQL Server sous Linux	148
Présentation de	148
Impact sur les coûts	149
Recommandations en matière d'optimisation des coûts	150
Ressources supplémentaires	151
Optimisation des stratégies de sauvegarde de SQL Server	152
Présentation de	152
Sauvegarde au niveau du serveur à l'aide de snapshots compatibles VSS	153
Sauvegarde de SQL Server à l'aide de AWS Backup	155

Sauvegarde au niveau de la base de données	157
Recommandations en matière d'optimisation des coûts	167
Ressources supplémentaires	170
Moderniser les bases de données SQL Server	171
Présentation de	171
Offres de bases de données	171
Comparaison entre Amazon RDS et Aurora	172
Recommandations en matière d'optimisation des coûts	174
Ressources supplémentaires	179
Optimisation du stockage pour SQL Server	180
Présentation de	180
Types de stockage SSD, performances et coûts pour Amazon EBS	181
Optimisation générale des coûts des SSD pour Amazon EBS	182
Ressources supplémentaires	184
Optimisez les licences SQL Server à l'aide de Compute Optimizer	185
Présentation de	185
Recommandations en matière d'optimisation des coûts	185
Configurer Compute Optimizer	186
Ressources supplémentaires	188
Optimisez le dimensionnement de SQL Server à l'aide de Compute Optimizer	188
Présentation de	188
Configurer Compute Optimizer	189
Ressources supplémentaires	190
Consultez Trusted Advisor les recommandations relatives aux charges de travail SQL Server .	190
Présentation de	190
Recommandations en matière d'optimisation des coûts	190
Configuration Trusted Advisor	191
Ressources supplémentaires	192
Containers	193
Déplacer les applications Windows vers des conteneurs	194
Présentation de	194
Avantages en termes de coûts	194
Recommandations en matière d'optimisation des coûts	196
Étapes suivantes	200
Ressources supplémentaires	200
Optimisez les coûts AWS Fargate des tâches sur Amazon ECS	201

Présentation de	201
Avantages en termes de coûts	201
Recommandations en matière d'optimisation des coûts	201
Étapes suivantes	208
Ressources supplémentaires	208
Gagnez en visibilité sur vos coûts Amazon EKS	209
Présentation de	209
Avantages en termes de coûts	209
Recommandations en matière d'optimisation des coûts	209
Étapes suivantes	213
Ressources supplémentaires	214
Replateforme des applications Windows avec App2Container	214
Présentation de	214
Avantages en termes de coûts	215
Recommandations en matière d'optimisation des coûts	216
Étapes suivantes	216
Ressources supplémentaires	216
Stockage	218
Amazon EBS	218
Migrer les volumes Amazon EBS de gp2 à gp3	219
Modifier les instantanés Amazon EBS	223
Supprimer des volumes Amazon EBS non attachés	226
Amazon FSx	230
Choisissez le bon système de stockage de fichiers pour PME	230
Activer la déduplication des données sur Amazon FSx	235
Comprendre le partitionnement des données FSx pour le serveur de fichiers Windows	238
Comprendre l'utilisation du volume du disque dur sur Amazon FSx	243
Utiliser une seule zone de disponibilité	246
AWS Storage Gateway	248
Passerelle de fichiers Amazon S3	249
Passerelle FSx de fichiers Amazon	249
Impact sur les coûts	249
Recommandations en matière d'optimisation des coûts	252
Ressources supplémentaires	254
Active Directory	255
Active Directory autogéré sur Amazon EC2	255

Présentation de	255
Impact sur les coûts	255
Recommandations en matière d'optimisation des coûts	256
Ressources supplémentaires	261
AWS Managed Microsoft AD	261
Présentation de	261
Impact sur les coûts	261
Recommandations en matière d'optimisation des coûts	262
Ressources supplémentaires	263
AD Connector	264
Présentation de	264
Impact sur les coûts	264
Recommandations en matière d'optimisation des coûts	264
Ressources supplémentaires	265
.NET	266
Passez au .NET moderne et passez à Linux	267
Présentation de	267
Impact sur les coûts	267
Recommandations en matière d'optimisation des coûts	268
Considérations et ressources supplémentaires	269
Conteneuriser les applications .NET	270
Présentation de	270
Impact sur les coûts	270
Recommandations en matière d'optimisation des coûts	272
Ressources supplémentaires	275
Utiliser des instances et des conteneurs Graviton	275
Présentation de	275
Impact sur les coûts	275
Recommandations en matière d'optimisation des coûts	277
Ressources supplémentaires	278
Support du dimensionnement dynamique pour les applications .NET Framework statiques	279
Présentation de	279
Impact sur les coûts	284
Recommandations en matière d'optimisation des coûts	285
Ressources supplémentaires	287
Utiliser la mise en cache pour réduire la demande de base de données	287

Présentation de	287
Impact sur les coûts	287
Recommandations en matière d'optimisation des coûts	288
Ressources supplémentaires	295
Envisagez le .NET sans serveur	295
Présentation de	295
Impact sur les coûts	296
Recommandations en matière d'optimisation des coûts	296
Ressources supplémentaires	300
Envisagez des bases de données spécialement conçues	300
Présentation de	300
Impact sur les coûts	304
Recommandations en matière d'optimisation des coûts	307
Ressources supplémentaires	309
Étapes suivantes	310
Historique du document	311
Glossaire	313
#	313
A	314
B	317
C	319
D	323
E	327
F	329
G	331
H	333
I	334
L	337
M	338
O	343
P	345
Q	348
R	349
S	352
T	356
U	357

V	358
W	358
Z	360
.....	ccclxi

Optimisez les coûts des charges de travail Microsoft sur AWS

Bill Pfeiffer, Chase Lindeman et Kevin Sookhan, Amazon Web Services (AWS)

octobre 2025 ([historique du document](#))

Présentation de

Ce guide fournit des recommandations, des meilleures pratiques et des stratégies pour vous aider à optimiser les coûts de vos charges de travail Microsoft sur AWS. Le guide inclut également des connaissances de base, des techniques d'optimisation des coûts et des architectures de référence pour vous aider à créer et à automatiser des charges de travail rentables et performantes qui répondent à vos objectifs commerciaux. Collectivement, ce guide est appelé Microsoft on AWS Cost Optimization (MACO). Les directives MACO ont été élaborées par des experts du secteur et sont basées sur des scénarios réels.

Ce guide couvre les charges de travail Microsoft suivantes :

- Windows sur Amazon Elastic Compute Cloud (Amazon EC2)
- SQL Server
- Containers
- Stockage
- Active Directory
- .NET

Public ciblé

Ce guide est destiné aux architectes, aux ingénieurs, aux administrateurs, aux directeurs CTOs, aux décideurs techniques et aux AWS partenaires. Il est utile mais pas nécessaire d'avoir une expérience préalable et une compréhension de base de la AWS facturation, des technologies Microsoft et de l'administration AWS des systèmes.

Comment utiliser ce guide

Vous pouvez utiliser ce guide pour planifier et mettre en œuvre votre transition vers le cloud avec MACO. Nous vous recommandons de lire ce guide du début à la fin afin de bien comprendre les options et les approches permettant d'optimiser les coûts de vos charges de travail Microsoft sur AWS. Vous pouvez consulter les sections suivantes relatives à la charge de travail en fonction des besoins de votre organisation :

- [Windows sur Amazon EC2](#)
- [SQL Server](#)
- [Conteneurs](#)
- [Stockage](#)
- [Active Directory](#)
- [.NET](#)

Important

Les exemples de code fournis dans ce guide sont uniquement destinés à des fins de démonstration. Il est recommandé de tester l'ensemble du code dans un environnement de développement avant de l'utiliser dans un environnement de production. Avant d'implémenter un code, nous vous recommandons de le tester par petits lots, puis de passer en revue les modifications de coûts qui en résultent en utilisant [AWS Cost Explorer](#). Cela peut vous aider à résoudre les cas extrêmes et d'autres problèmes susceptibles de devenir problématiques ultérieurement.

Important

Les exemples de prix présentés dans ce guide sont basés sur les prix au moment de la publication. Les prix sont susceptibles d'être modifiés. En outre, vos coûts peuvent varier en fonction de vos Région AWS Service AWS quotas et d'autres facteurs liés à votre environnement cloud.

Résultats commerciaux ciblés

Ce guide peut vous aider, vous et votre organisation, à atteindre les résultats commerciaux suivants :

- Découvrez comment utiliser une évaluation de AWS l'optimisation et des licences (AWS OLA) pour évaluer et optimiser vos environnements actuels sur site et dans le cloud, en fonction de l'utilisation des ressources, des licences tierces et des dépendances des applications.
- Développez une analyse de rentabilisation pour l'optimisation des coûts à l'aide du calculateur de AWS modernisation pour Microsoft Workloads.
- Optimisez les coûts pour vos charges de travail Microsoft spécifiques, notamment les charges de travail pour Windows sur Amazon Elastic Compute Cloud EC2 (Amazon), SQL Server, les conteneurs, le stockage, Active Directory et .NET.

Parcours d'optimisation des coûts

L'étendue, le calendrier et le parcours spécifique de votre migration vers le cloud dépendent de vos objectifs commerciaux, de vos exigences techniques et d'autres facteurs. Cette section fournit un exemple de parcours de migration vers le cloud axé sur la [gestion financière dans le cloud, AWS conformément](#) aux recommandations et aux meilleures pratiques de MACO. Vous pouvez utiliser cet exemple pour comprendre comment concevoir un parcours de migration vers le cloud pour les charges de travail Microsoft.

Les tâches de haut niveau suivantes illustrent l'approche qu'une organisation pourrait adopter pour mettre en œuvre les recommandations et les meilleures pratiques du MACO :

- Établissez une stratégie de balisage et activez les balises de répartition des coûts définies par l'utilisateur. Pour plus d'informations, consultez le AWS livre blanc sur les [meilleures pratiques en matière de balisage AWS des ressources](#).
- Définissez des budgets en fonction des applications, des équipes ou des départements. Pour plus d'informations, consultez [Managing your costs with AWS Budgets](#) dans le guide AWS de l'utilisateur de Billing and Cost Management.
- Réalisez une évaluation de AWS l'optimisation et des licences (AWS OLA) pour accélérer les économies. Pour plus d'informations, consultez la section [AWS Optimisation et évaluation des licences](#) dans la AWS documentation.
- Bring Your Own License (BYOL) pour les charges de travail Windows et SQL Server en utilisant Amazon Elastic Compute Cloud Dedicated Hosts Pour de plus amples informations, veuillez consulter la section [Apportez des licences pour les charges de travail Windows et SQL Server](#) de ce guide.
- Optimisez vos licences SQL Server sur AWS. Pour de plus amples informations, veuillez consulter la section [Comprendre les licences SQL Server](#) de ce guide.
- Sélectionnez le type d'instance adapté aux charges de travail Windows. Pour de plus amples informations, veuillez consulter la section [Sélectionnez le type d'instance adapté aux charges de travail Windows](#) de ce guide.
- Sélectionnez le type d'instance adapté aux charges de travail SQL. Pour de plus amples informations, veuillez consulter la section [Sélectionnez l'instance EC2 adaptée aux charges de travail SQL Server](#) de ce guide.

- Migrez Amazon Elastic Block Store (Amazon EBS) de gp2 à gp3. Pour de plus amples informations, veuillez consulter la section [Migrer les volumes Amazon EBS de gp2 à gp3](#) de ce guide.
- Contrôlez les charges de travail avec EC2 Instance Scheduler activé. AWS Pour de plus amples informations, veuillez consulter la section [Automatisez les plannings d'arrêt et de démarrage](#) de ce guide.
- Supprimez les coûts associés à SQL Server pour les charges de travail hors production en utilisant SQL Server Developer Edition. Pour de plus amples informations, veuillez consulter la section [Évaluer l'édition SQL Server Developer](#) de ce guide.
- Utilisez une zone de disponibilité unique pour le serveur de fichiers Amazon FSx pour Windows pour les charges de travail de développement et de test. Pour de plus amples informations, veuillez consulter la section [Utiliser une seule zone de disponibilité](#) de ce guide.
- Dimensionnez correctement vos charges de travail Windows en utilisant. Optimiseur de calcul AWS Pour de plus amples informations, veuillez consulter la section [Charges de travail Windows de la bonne taille](#) de ce guide.
- Optimisez les dépenses sur Windows sur Amazon EC2 en utilisant Savings Plans. Pour de plus amples informations, veuillez consulter la section [Optimisez les dépenses pour Windows sur Amazon EC2](#) de ce guide.
- Activez la déduplication FSx des données sur Windows File Server. Pour de plus amples informations, veuillez consulter la section [Activer la déduplication des données sur Amazon FSx](#) de ce guide.
- Utilisez le découpage des données pour les systèmes de fichiers sur FSx Windows File Server. Pour de plus amples informations, veuillez consulter la section [Comprendre le partitionnement des données FSx pour le serveur de fichiers Windows](#) de ce guide.
- Optimisez vos stratégies de sauvegarde SQL Server. Pour de plus amples informations, veuillez consulter la section [Optimisation des stratégies de sauvegarde de SQL Server](#) de ce guide.
- Faites en sorte que les applications .NET Framework statiques prennent en charge le dimensionnement dynamique. Pour plus d'informations, consultez le [Support du dimensionnement dynamique pour les applications .NET Framework statiques](#) contenu de ce guide.
- Utilisez des microservices .NET sans serveur. Pour de plus amples informations, veuillez consulter la section [Envisagez le .NET sans serveur](#) de ce guide.
- Déplacez vos applications Windows vers des conteneurs. Pour de plus amples informations, veuillez consulter la section [Conteneuriser les applications .NET](#) de ce guide.

- [Optimiseur de calcul AWS](#) À utiliser pour dimensionner correctement les conteneurs Windows exécutés sur AWS Fargate Amazon Elastic Container Service (Amazon ECS). Pour de plus amples informations, veuillez consulter la section [Activer Compute Optimizer](#) de ce guide.
- Passez au .NET moderne et passez à Linux. Pour de plus amples informations, veuillez consulter la section [Passez au .NET moderne et passez à Linux](#) de ce guide.
- Tirez parti des instances et des conteneurs Graviton. Pour de plus amples informations, veuillez consulter la section [Utiliser des instances et des conteneurs Graviton](#) de ce guide.
- Modernisez les bases de données SQL Server. Pour de plus amples informations, veuillez consulter la section [Moderniser les bases de données SQL Server](#) de ce guide.
- Concevez une infrastructure Active Directory. Pour de plus amples informations, veuillez consulter la section [Active Directory](#) de ce guide.

Pour plus d'informations sur un parcours client axé sur la gestion financière dans le cloud avec AWS, consultez le AWS livre blanc sur la fonctionnalité de [gestion financière dans le cloud](#).

Les meilleures recommandations pour optimiser les coûts

Présentation

L'optimisation des coûts est l'un des piliers du [AWS Well-Architected](#) Framework et joue un rôle essentiel dans vos plans de migration vers le cloud. Vous trouverez des recommandations pour l'optimisation des coûts tout au long de ce guide, mais cette section présente les recommandations ayant le plus d'impact. Vous pouvez mettre en œuvre ces recommandations rapidement et elles auront un impact significatif sur votre organisation. Ces recommandations peuvent vous aider à jeter les bases de l'ensemble de vos efforts d'optimisation des coûts.

Les meilleures recommandations

Le tableau suivant répertorie les principales recommandations pour les optimisations de coûts les plus efficaces. La colonne « Difficile à mettre en œuvre » évalue chaque optimisation en fonction d'une échelle allant de ce qui est le plus facile à mettre en œuvre (1) à ce qui est le plus difficile à mettre en œuvre (5). La colonne « Économies estimées » indique une estimation en pourcentage des économies que votre organisation peut réaliser pour chaque optimisation recommandée.

Optimisations	Difficulté à mettre en œuvre	Économies estimées
Charges de travail Windows de la bonne taille	3	25%
Apportez des licences pour les charges de travail Windows et SQL Server	3	30 %
Évaluer l'édition SQL Server Developer	2	20 %
Comprendre les licences SQL Server	2	Jusqu'à 50 %
Automatisez les plannings d'arrêt et de démarrage	3	Jusqu'à 40 %

Optimisations	Difficulté à mettre en œuvre	Économies estimées
Sélectionnez le type d'instance adapté aux charges de travail Windows	1	10 à 30 %
Passez au .NET moderne et passez à Linux	5	10 à 20 %
Optimisez les dépenses pour Windows sur Amazon EC2	3	Jusqu'à 20 à 40 %
Migrer les volumes Amazon EBS de gp2 à gp3	4	Jusqu'à 20 %

Important

Les économies estimées dans le tableau précédent s'appliquent à chaque domaine technique individuel, et non aux AWS dépenses globales d'un compte. Par exemple, vous pouvez implémenter le planificateur d'instances dans différents types et tailles d'environnements susceptibles de modifier les économies potentielles. Les estimations s'appliquent spécifiquement aux coûts des EC2 instances Amazon et n'impliquent aucune économie globale pour les autres Services AWS. Ces estimations sont fournies à titre d'indicateur et non à titre de garantie.

Les experts de MACO sont disponibles pour discuter plus en profondeur de l'optimisation des coûts. Pour organiser une réunion afin d'étudier en détail votre cas d'utilisation, contactez l'équipe chargée de votre compte ou envoyez un e-mail à optimize-microsoft@amazon.com.

AWS Optimisation et évaluation des licences

Présentation

Une [évaluation de AWS l'optimisation et des licences \(AWS OLA\)](#) peut vous aider à évaluer et à optimiser vos environnements cloud actuels et existants, en fonction de l'utilisation des ressources, des licences tierces et des dépendances des applications. Vous pouvez utiliser AWS OLA pour aider votre entreprise à élaborer une stratégie de migration et de licence permettant de réaliser des économies lors de la migration vers AWS ou de l'évaluation des charges de travail Microsoft existantes sur. AWS Un AWS OLA peut également vous aider à atteindre les objectifs suivants :

- Comprenez les déploiements existants, les performances des applications et les contrats.
- Dimensionnez correctement vos ressources.
- Élaborez une feuille de route pour le AWS Cloud.
- Réduisez ou éliminez les coûts en utilisant les investissements existants et en ne payant que ce que vous utilisez.

Nous vous recommandons de faire d'un AWS OLA la première étape de votre [parcours d'optimisation des coûts](#). Vous pouvez utiliser le AWS Partner Network pour compléter une AWS OLA. Ils vous aideront à recueillir des données d'évaluation et vous fourniront des recommandations pour optimiser vos coûts de licence et d'instance.

Le schéma suivant donne un aperçu du processus d'évaluation.



Options d'évaluation

Vous pouvez choisir entre deux options AWS OLA pour vos charges de travail Microsoft sur AWS :

- **Version allégée** : dans ce cas d'utilisation, toutes vos charges de travail sont activées VMware. Vous pouvez fournir AWS une sortie à partir de [RVTools](#). Ensuite, AWS peut offrir un délai de traitement de 1 à 5 jours. Cette approche utilise des point-in-time informations extraites directement de VMware vCenter pour élaborer des recommandations de dimensionnement et proposer des options de tarification à la demande.
- **Version complète** : dans ce cas d'utilisation, vous disposez d'un environnement mixte exécuté par différents fournisseurs de cloud, serveurs physiques et serveurs virtuels. AWS utilise des agents du système d'exploitation pour collecter des données d'utilisation sur une période de 14 à 30 jours. Cela permet de AWS prendre des décisions éclairées en matière de dimensionnement des instances en fonction des modèles d'utilisation de vos applications. AWS utilise plusieurs outils tiers, tels que Cloudamize, pour effectuer l'analyse. AWS travaille avec its AWS Partner Network pour aider à réaliser l'évaluation finale du coût total de possession (TCO) avec de multiples options de tarification qui tiennent compte des modèles de tarification et des différentes architectures.

Évaluation complète

L'évaluation complète de l' AWS OLA commence par un appel téléphonique d'une heure. Au cours de cet appel, il vous AWS aide à déterminer l' AWS infrastructure la plus optimale pour prendre en charge votre migration, à choisir une méthode de collecte de données et à établir un calendrier de réalisation. La mise en œuvre d'outils de découverte dans votre organisation dépend de la méthode de collecte des données, de la taille de votre organisation et des outils utilisés par votre organisation pour gérer son parc de serveurs. La collecte des données d'utilisation prend généralement deux semaines.

Le processus AWS OLA complet prend entre 30 et 45 jours et comprend les phases suivantes :

- Élargir les charges de travail
- Collectez des données
- Analyser les données
- Planifier les prochaines étapes

Élargir les charges de travail

Tout d'abord, il AWS travaille avec vous et votre équipe pour déterminer la portée de l'évaluation. Cela est généralement ventilé par type d'environnement (par exemple, hors production et production). Le champ d'application inclut l'emplacement des charges de travail. Il peut s'agir de charges de travail vers lesquelles vous migrez AWS, de charges de travail déjà en cours d'exécution AWS (par exemple, OLA AWS pour Amazon EC2) ou de charges de travail exécutées par d'autres fournisseurs de cloud.

Collectez des données

AWS Déploie ensuite des outils pour faciliter la découverte des ressources et collecter les données de performance de vos serveurs. Cet outillage se décline en quatre options de déploiement :

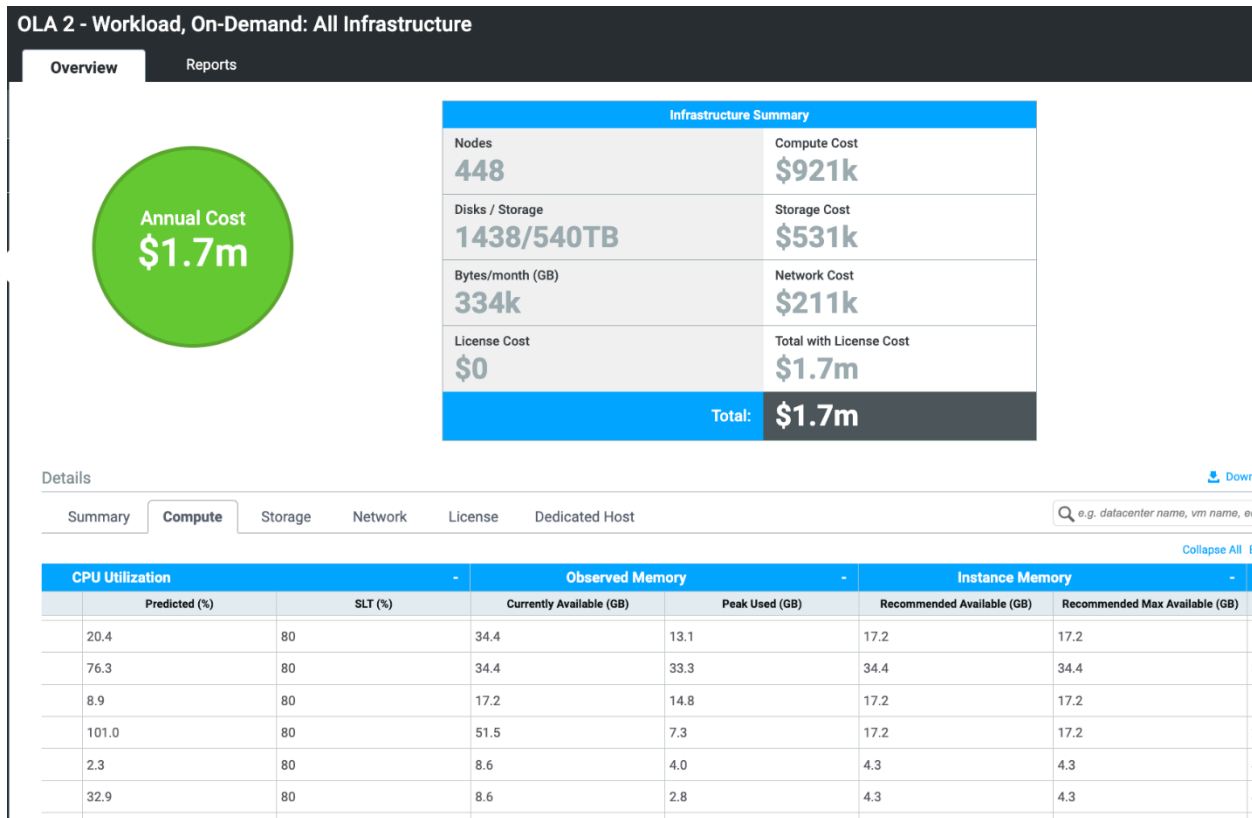
- Outils permettant d'interroger l'hyperviseur (nécessite uniquement des informations d'identification VMware vCenter ou Hyper-V)
- Agents pouvant être déployés sur des machines physiques ou virtuelles
- Découverte sans agent à l'aide de SSH, de Windows Remote Management (WinRM) ou de Windows Management Instrumentation (WMI) en fonction de votre environnement et de votre système d'exploitation
- Collecte et analyse de données de fichiers plats

Pour le déploiement de vos outils, vous pouvez combiner chaque option et consolider les résultats. Il est essentiel de veiller à ce que l'option que vous choisissez ne sollicite pas vos ressources informatiques. AWS s'efforce de rendre le processus d'évaluation aussi clé en main que possible. Au-delà d'un bref appel téléphonique pour aider à la configuration, l'équipe AWS OLA et les architectes de solutions spécialisés de Microsoft prépareront l'analyse du coût total de possession (TCO) et les recommandations pour examen.

La collecte des données prend généralement deux à trois semaines lorsque l'utilisation du processeur, l'utilisation de la RAM, le débit de stockage, les IOPS et le débit du réseau sont analysés. Idéalement, cette collecte a lieu pendant les périodes de pointe de votre mois d'affaires (par exemple, pendant les rapports end-of-month financiers). AWS souhaite capturer le pic d'utilisation, car cela fournit de bons échantillons statistiques indiquant la taille idéale de l' AWS instance, tout en garantissant des performances supérieures à celles disponibles sur site. AWS fusionne les indicateurs d'utilisation avec les heuristiques de performance des différentes générations de processeurs afin de cibler exactement la quantité de CPU et de RAM requise par une charge de

travail donnée. Ces objectifs sont généralement inférieurs à ce qui est alloué sur place. Cela permet non seulement de réduire les coûts de calcul liés à la taille de l'instance, mais également d'optimiser les coûts de licence.

La vue du tableau de bord suivante montre un exemple de coûts d'infrastructure qui peuvent être pris en compte par une évaluation.

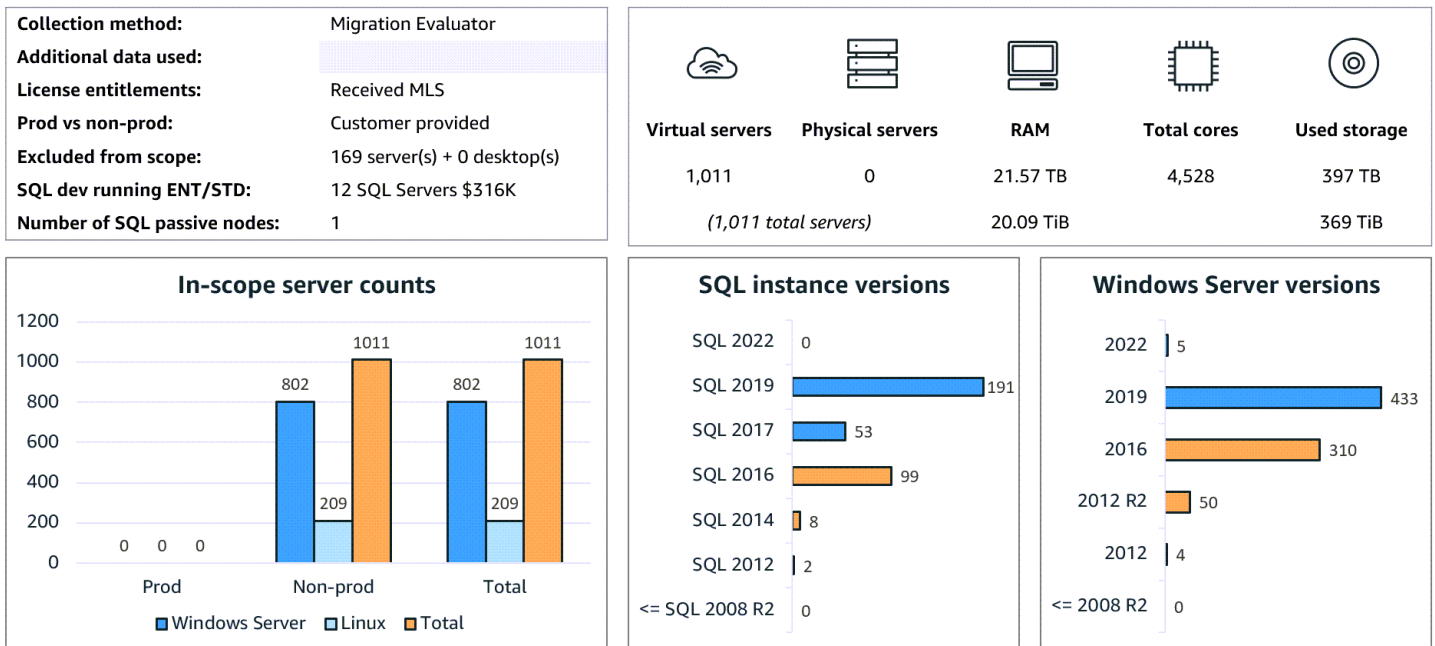


Analyser les données

AWS fournit un compte rendu une fois la collecte des données terminée. AWS examine les données, résume les résultats, puis formule des recommandations pour l'utilisation sur site et la migration vers le cloud. Vous pouvez réduire les coûts de calcul et de licence en examinant les opportunités de consolidation, les gains d'élasticité (lorsque les charges de travail peuvent être désactivées ou corrigées en fonction des variations saisonnières), les opportunités d'utilisation du bon SKU (par exemple, l'édition SQL Server Enterprise est utilisée, mais les besoins en ressources et l'utilisation des fonctionnalités suggèrent que l'édition SQL Server Standard convient). Pour les produits tels que SQL Server dont la licence est octroyée par le noyau, il est souvent rentable de placer les charges de travail dans une instance de calcul plus coûteuse. En effet, si le profil du processeur et le ratio entre

la RAM et le processeur virtuel ont pour effet net de réduire le nombre de cœurs sous licence, à la fois pour les cas d'utilisation avec licence et pour les applications BYOL (Bring Your Own License).

Vous trouverez ci-dessous un exemple d'analyse basé sur les données collectées lors de l'évaluation.



Les scénarios d'optimisation courants incluent l'identification des opportunités d'optimisation des AWS ressources et des économies de licences tierces.

Exemples d'opportunités d'optimisation des AWS ressources :

- Évitez le surprovisionnement en cas de pic d'utilisation.
- Évitez de surspécifier et de sous-utiliser les ressources.
- Dimensionnez correctement vos instances et migrez vers les nouvelles générations d'EC2 instances.
- Réduisez les coûts d'exploitation en optant pour des bases de données gérées.

Exemples d'économies réalisées sur des licences tierces :

- Réduisez le nombre de cœurs nécessaires pour exécuter la même charge de travail.
- Débarrassez-vous de l'édition Enterprise et des packs complémentaires inutiles de SQL Server.
- Supprimez les serveurs zombies et remplacez le matériel obsolète.
- Utilisez le BYOL et les options incluses dans les licences pour réduire les futurs accords commerciaux.

- Modernisez-vous vers des solutions open source et cloud natives.

Planifier les prochaines étapes

Enfin, AWS utilise les données de performance collectées pour estimer le dimensionnement et le coût spécifiques de la charge de travail. AWS peut également examiner globalement votre environnement défini et fournir une analyse quantitative. Cela peut vous aider à déterminer si la meilleure option est une actualisation sur site ou une migration vers AWS. Vous pouvez élaborer une analyse de rentabilisation économique du cloud en utilisant le résumé de l'analyse du coût total de possession (comme indiqué dans l'exemple suivant) fourni à la fin d'un AWS OLA.

	Option 1: Amazon EC2 shared	Option 1a: Amazon EC2 shared + power management	Option 2: Amazon EC2 mixed	Option 2a: Amazon EC2 mixed + power management
<i>Option details: compute</i>	100% Reserved Instances (RIs)	RIs + on-demand power management	100% RIs	RIs + on-demand power management
<i>Option details: Microsoft licenses</i>	WS LI and SQL BYOL	WS LI and SQL BYOL	WS BYOL or LI+SQL BYOL	WS BYOL or LI+SQL BYOL
Compute costs¹				
Year 1 compute cost	\$414,546	\$482,623	\$504,019	\$513,941
Year 1 vendor license included cost	\$392,858	\$244,415	\$9,804	\$4,783
	\$807,404	\$727,038	\$513,823	\$518,724
<i>Total compute savings in year 1, compared to Option 1</i>	—	10% (\$80,366)	36% (\$293,581)	36% (\$288,680)
Storage and networking costs²				
Annual estimated storage cost	\$336,494	\$336,494	\$336,494	\$336,494
Annual estimated networking cost	\$41,455	\$41,455	\$41,455	\$41,455
	\$377,949	\$377,949	\$377,949	\$377,949
Microsoft license costs**				
WS/CIS annual Software Assurance (SA) + current SPLA/Subs cost	\$0	\$0	\$0	\$0
WS/CIS license + SA + SPLA/Subs true-up cost	\$0	\$0	\$0	\$0
SQL annual SA + current SPLA/Subs cost	\$0	\$0	\$0	\$0
SQL license SA + current SPLA/Subs true-up cost	\$0	\$0	\$0	\$0
	\$0	\$0	\$0	\$0
Total estimated costs	\$1,185,353	\$1,104,987	\$891,772	\$896,673
<i>Annual TCO savings in year 1, compared to Option 1</i>	—	7% (\$80,366)	25% (\$293,581)	24% (\$288,680)

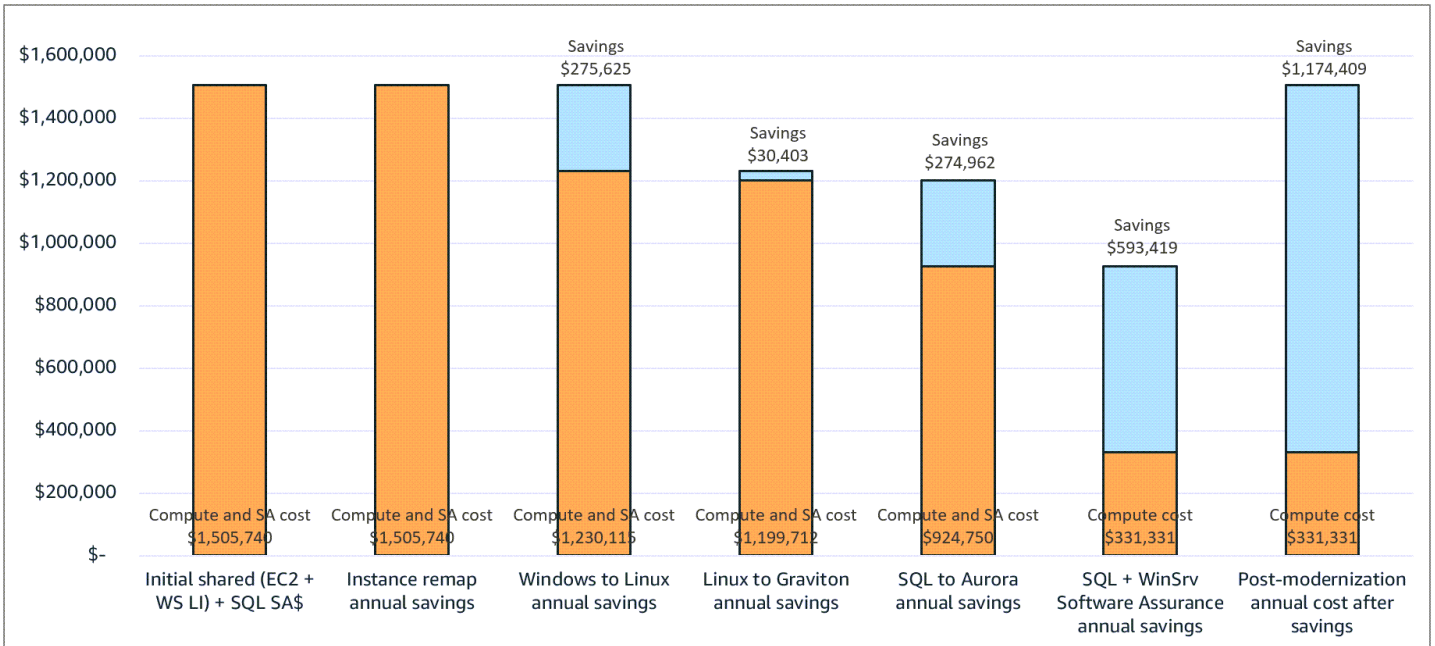
¹ Pricing model used: 3-year, no upfront RI

² Software Assurance and true-up costs provided by Microsoft

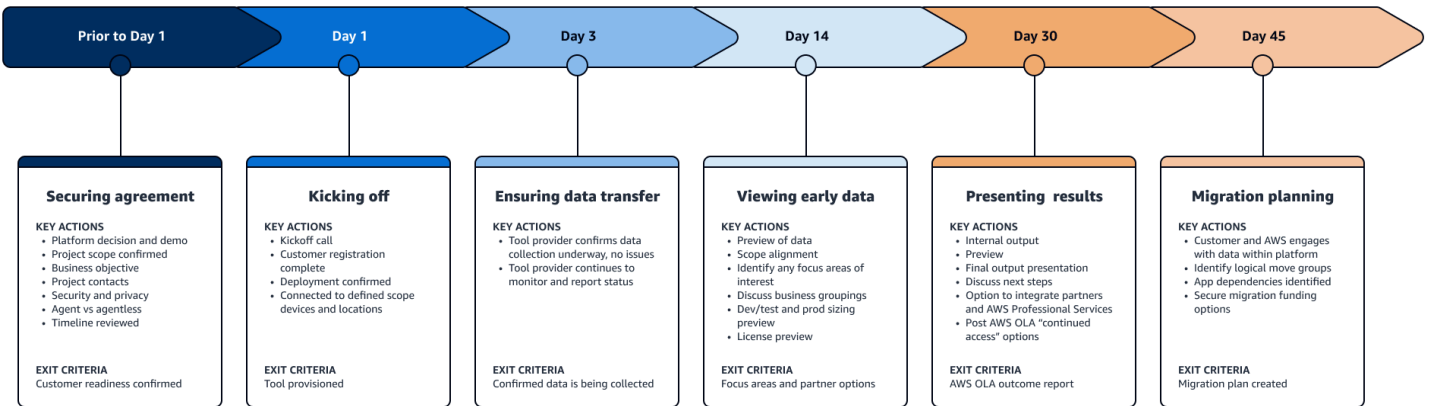
Un AWS OLA fournit également un aperçu de l'impact que la modernisation peut avoir sur vos charges de travail existantes en faisant des suggestions telles que les suivantes :

- Passez à un système d'exploitation Linux.
- Ajoutez la prise en charge des applications pour les processeurs ARM (AWS Graviton).
- Déplacez les charges de travail SQL Server vers Amazon Aurora.
- Supprimez l'assurance logicielle en transférant les charges de travail Windows et SQL Server vers des technologies open source.

Le schéma suivant montre les économies qui peuvent être réalisées grâce à des techniques de modernisation telles que le passage de Windows à Linux ou de SQL Server à Aurora.



Le processus AWS OLA complet prend environ 45 jours du début à la fin. Le schéma suivant montre un exemple de chronologie.



Si vous disposez d'un VMware environnement pur et que vous pouvez fournir des résultats RVTools, vous pouvez réduire ce délai à une semaine ouvrable. En outre, il AWS peut analyser un fichier plat contenant des données relatives aux actifs et à l'utilisation, telles que la moyenne du processeur, le pic du processeur, la moyenne de la RAM et le pic de RAM.

Impact de l'évaluation

Le client moyen bénéficie généralement d'une réduction de 20 à 30 % de ses coûts grâce à un effort de dimensionnement approprié. Le bon dimensionnement fait correspondre la charge de travail source aux AWS instances les mieux dimensionnées en fonction des données d'utilisation. Ces ajustements de taille permettent non seulement de réduire le coût mensuel de l' AWS environnement, mais également de réaliser des économies dans d'autres secteurs de l'organisation. Par exemple, un gain de 20 à 30 % en termes de licences Windows ou SQL Server peut réduire le prochain rapprochement avec Microsoft ou libérer des licences pour des applications supplémentaires. line-of-business C'est généralement grâce à la consolidation et au dimensionnement correct des charges de travail SQL Server que les gains financiers les plus spectaculaires sont réalisés.

AWS peut vous aider à classer les systèmes dans des compartiments de modernisation. Certains systèmes sont anciens et ne sont pas financièrement viables, tandis que d'autres peuvent être modernisés pour devenir des conteneurs ou des applications sans serveur, ce qui permet de réaliser les économies les plus importantes. La conversation avec votre AWS équipe passe de sujets généraux sur ce que permet le cloud à des discussions plus spécifiques sur comment et pourquoi des charges de travail spécifiques devraient être modernisées. AWS vous aide également à explorer les opportunités d'innovation potentielles.

Étapes suivantes

Si vous commencez votre parcours d'optimisation des coûts pour les charges de travail Microsoft exécutées dans des environnements sur site ou sur site AWS, contactez l'équipe chargée de votre AWS compte et demandez un AWS OLA. AWS les membres de l'équipe peuvent répondre à vos questions et vous aider à décider si une AWS LLO est finalement le bon choix pour vous et votre organisation. Vous pouvez également [demander un AWS OLA en ligne](#).

Ressources supplémentaires

- [AWS Optimisation et évaluation des licences](#) (AWS documentation)
- [AWS re:Invent 2022 - Comment réduire les coûts et optimiser les charges de travail Microsoft sur AWS \(ENT205\) \(\)](#) YouTube

Windows sur Amazon EC2

[Amazon Elastic Compute Cloud \(Amazon EC2\)](#) est une plateforme de cloud computing hautement flexible et évolutive, idéale pour exécuter vos charges de travail Windows. Vous pouvez utiliser Amazon EC2 pour déployer, gérer et dimensionner vos charges de travail Windows Server sur l'infrastructure sécurisée, fiable, hautement disponible et adaptable du. AWS Cloud Tenez compte des principaux avantages suivants liés à l'exécution de charges de travail Windows sur Amazon EC2 :

- **Évolutivité** — Amazon EC2 vous permet d'adapter facilement vos charges de travail Windows en fonction de l'évolution des exigences. Vous pouvez créer rapidement de nouvelles instances EC2 pour répondre à l'augmentation de la demande, et tout aussi facilement mettre fin aux instances lorsqu'elles ne sont plus nécessaires. Vous ne payez que pour les ressources que vous utilisez réellement.
- **Flexibilité** — Windows sur Amazon EC2 prend en charge un large éventail de types d'instances conçus pour répondre aux différentes exigences de charge de travail, qu'il s'agisse d'instances à usage général, de mémoire ou d'instances optimisées pour le calcul. Cette flexibilité vous permet de choisir le type d'instance le mieux adapté à vos applications Windows spécifiques, d'optimiser les performances et de minimiser les coûts.
- **Sécurité** : AWS fournit plusieurs niveaux de sécurité pour vos charges de travail Windows, notamment des pare-feux réseau, le chiffrement des données et des contrôles d'accès sécurisés. Cela signifie que vous pouvez être sûr que vos applications et vos données sont protégées, tout en gardant un contrôle total sur vos paramètres et configurations de sécurité.
- **Rentabilité** — Le modèle de pay-as-you-go tarification vous permet de ne payer que pour les ressources que vous utilisez, éliminant ainsi le besoin d'investissements initiaux en matériel et en logiciels. Ce modèle vous permet également d'optimiser vos coûts, de réduire les dépenses d'investissement et d'accroître l'efficacité opérationnelle. Il s'agit d'un modèle de tarification idéal pour les entreprises de toutes tailles.

Cette section du guide couvre les sujets suivants :

- [Automatisez les plannings d'arrêt et de démarrage](#)
- [Charges de travail Windows de la bonne taille](#)
- [Sélectionnez le type d'instance adapté aux charges de travail Windows](#)
- [Apportez des licences pour les charges de travail Windows et SQL Server](#)

- [Optimisez les dépenses pour Windows sur Amazon EC2](#)
- [Surveillez les coûts à l'aide d' AWS outils](#)

Automatisez les plannings d'arrêt et de démarrage

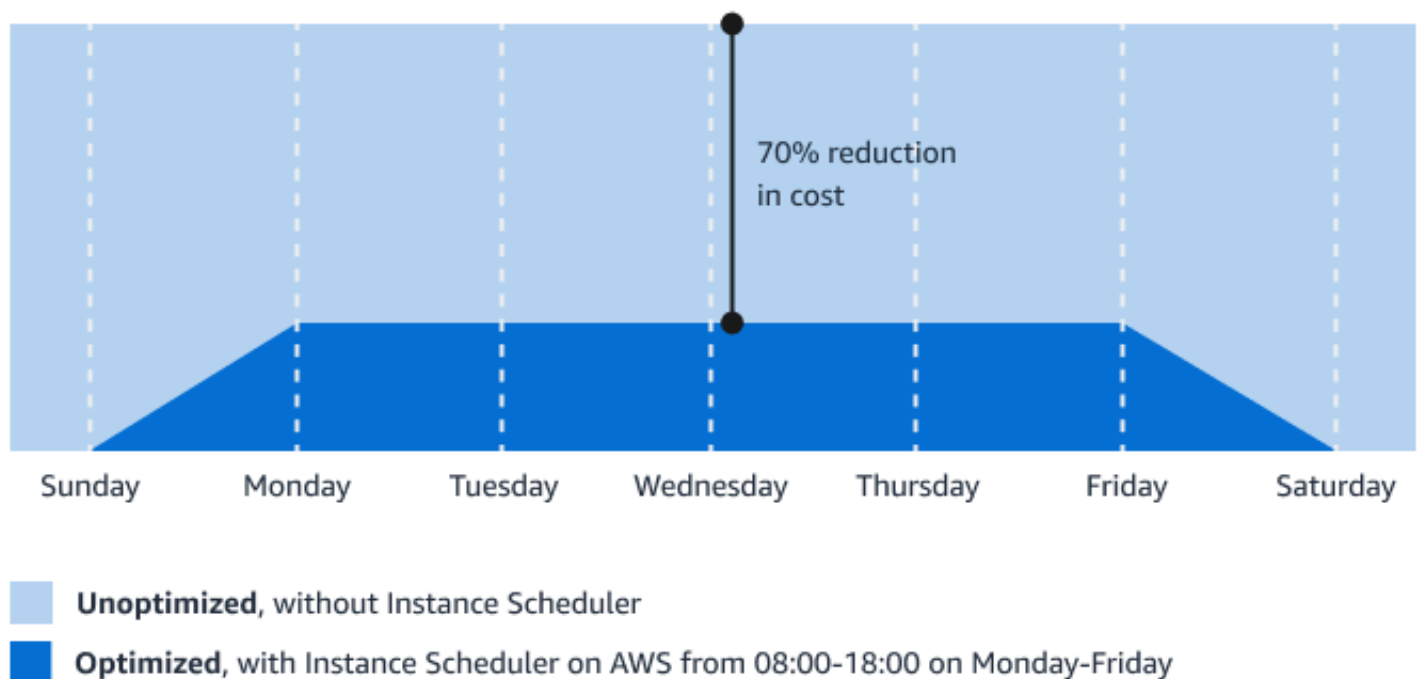
Présentation de

Le [planificateur d'instance activé AWS](#) peut vous aider à réduire les coûts opérationnels en automatisant le démarrage et l'arrêt des instances Amazon EC2 et [Amazon Relational Database Service](#) (Amazon RDS). Si vous laissez toutes vos instances fonctionner à pleine capacité en continu, vous pourriez finir par payer pour des ressources qui ne sont pas utilisées. Le planificateur d'instances activé vous AWS permet de désactiver les instances lorsqu'elles ne sont pas nécessaires, par exemple en dehors des heures de bureau, le week-end ou à d'autres périodes où l'utilisation est faible. Cela peut permettre de réaliser d'importantes économies au fil du temps.

Le planificateur d'instances activé permet AWS également la planification d'instances entre comptes, le balisage automatique et la possibilité de configurer des plannings ou des périodes à l'aide d'une interface de ligne de commande ou de la fenêtre de maintenance. [AWS Systems Manager](#) Ces fonctionnalités peuvent vous aider à gérer vos instances de manière plus efficace et à suivre et répartir les coûts avec précision entre différents projets ou équipes.

Études de cas

Prenons l'exemple d'une entreprise qui utilise Instance Scheduler AWS dans un environnement de production pour arrêter automatiquement les instances en dehors des heures de bureau tous les jours. Si cette entreprise laisse toutes ses instances fonctionner à pleine capacité, elle peut réaliser jusqu'à 70 % d'économies sur les instances qui ne sont nécessaires que pendant les heures normales de bureau. Le graphique suivant montre comment l'utilisation hebdomadaire est réduite de 168 heures à 50 heures.

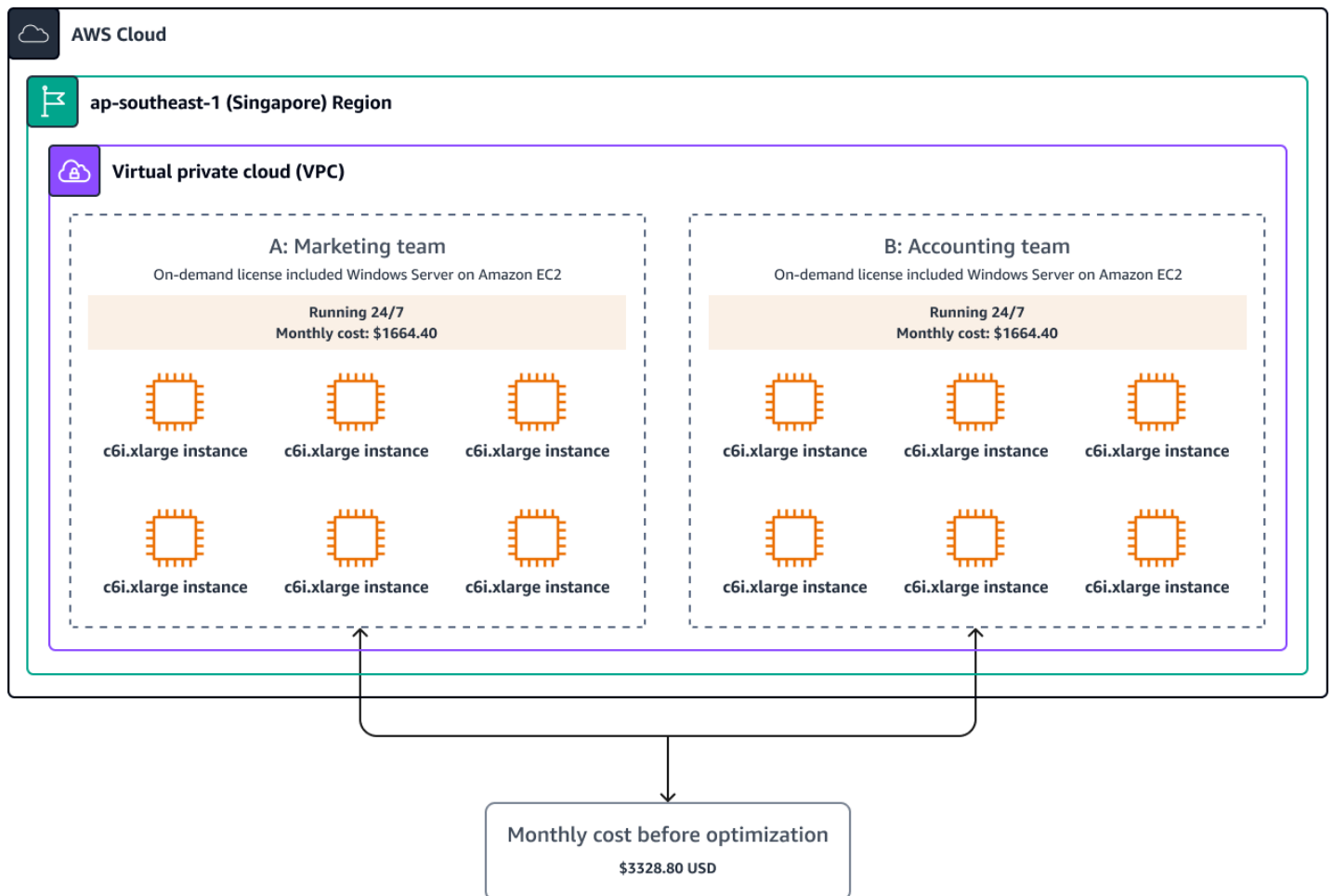


Prenons un autre exemple. La société d'électricité Jamaica Public Service Company Limited (JPS) a migré sa base de données vers Amazon RDS. JPS utilise désormais Amazon EC2 pour héberger des services d'API et exécuter d'autres applications. Pour JPS, Instance Scheduler AWS est devenu l'outil clé pour gérer les environnements hors production. JPS a utilisé le planificateur d'instances AWS pour réduire les coûts de développement et gérer les instances EC2 en fonction des besoins de l'équipe et des horaires de travail. Cela a permis à JPS de réduire ses coûts de 40 %. Pour plus d'informations, consultez l'étude de cas : le [service public jamaïcain migre efficacement vers le cloud et réduit ses coûts de 40 % AWS grâce à Instance Scheduler](#).

Scénario d'optimisation des coûts

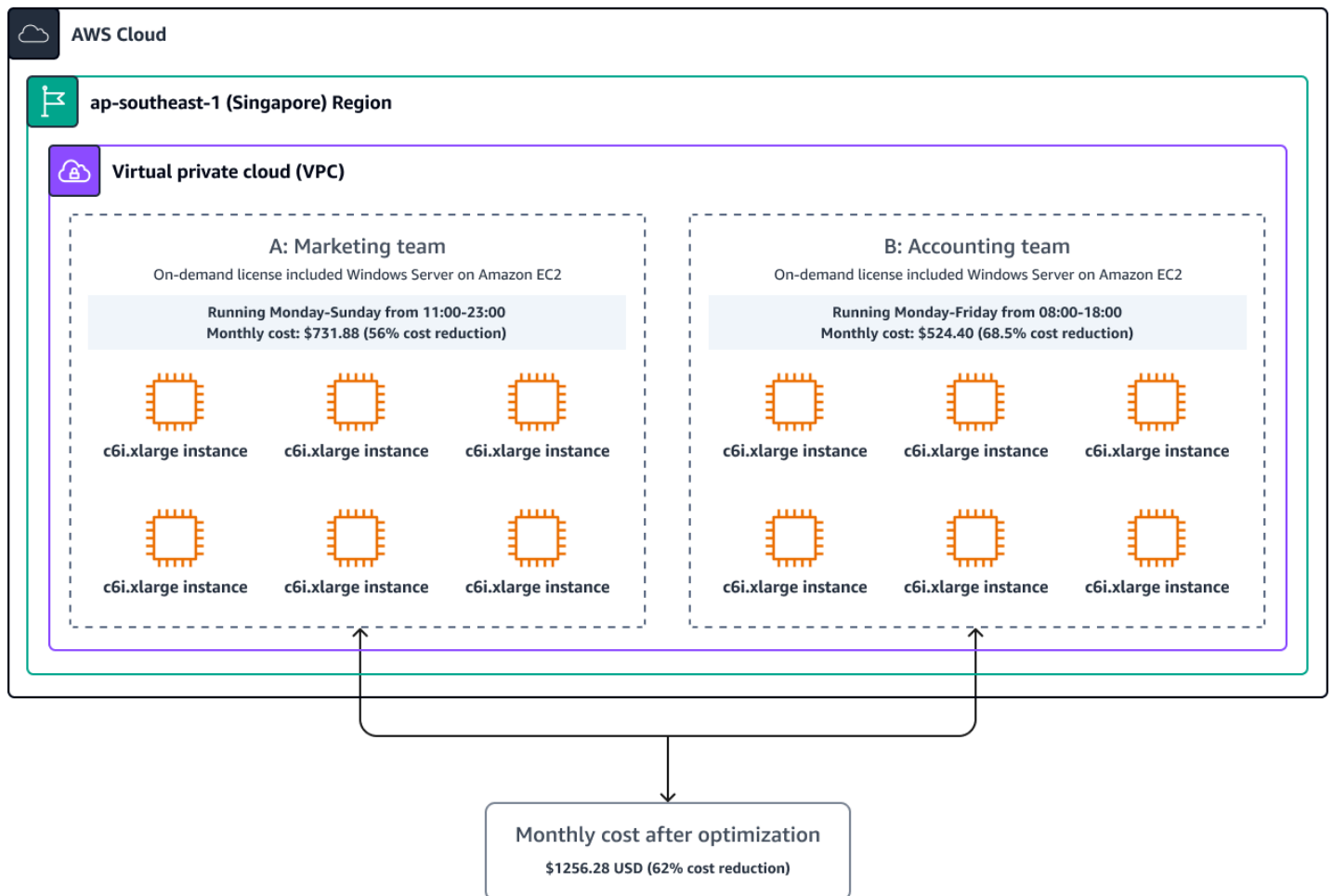
L'exemple de scénario suivant permet d'illustrer les avantages financiers liés à l'utilisation d'Instance Scheduler sur AWS. Dans ce scénario, une grande entreprise de vente au détail de Singapour déploie deux environnements Windows sur Amazon EC2. Le premier environnement, connu sous le nom de charge de travail A, est utilisé par l'équipe marketing pour analyser les transactions en temps réel en magasin lorsque les magasins sont ouverts. Le second environnement, appelé charge de travail B, est réservé à l'équipe comptable, qui travaille uniquement pendant les heures normales de bureau. Le calendrier d'exploitation actuel des deux environnements (24 heures sur 24, 7 jours sur 7) n'est pas idéal compte tenu des modèles d'utilisation actuels et doit être optimisé pour réduire les coûts d'exploitation de l'entreprise.

Le schéma suivant montre le coût mensuel avant optimisation.



Par exemple, le mois de mars compte 31 jours, dont 23 jours de semaine. Si l'équipe marketing utilise Instance Scheduler AWS et ne fait fonctionner ses instances qu'en cas de besoin (c'est-à-dire 321 heures par mois au lieu de 730 heures par mois), elle pourrait potentiellement économiser 932,52\$ par mois. Cela représente une réduction de 56 % des coûts d'exploitation. L'équipe comptable peut également bénéficier d'avantages significatifs, la durée d'utilisation de ses instances passant de 730 heures par mois à 230 heures. Cela se traduit par une réduction de 1 140\$, soit 68,5 %. L'entreprise pourrait économiser un total combiné de 2 072,52\$ par mois (soit une réduction de 62 %), soit 24 870,24\$ par an.

Le schéma suivant montre le coût mensuel après optimisation.



Note

Le prix de cet exemple a été déterminé en utilisant le [Calculateur de tarification AWS](#) en mars 2023.

Recommandations en matière d'optimisation des coûts

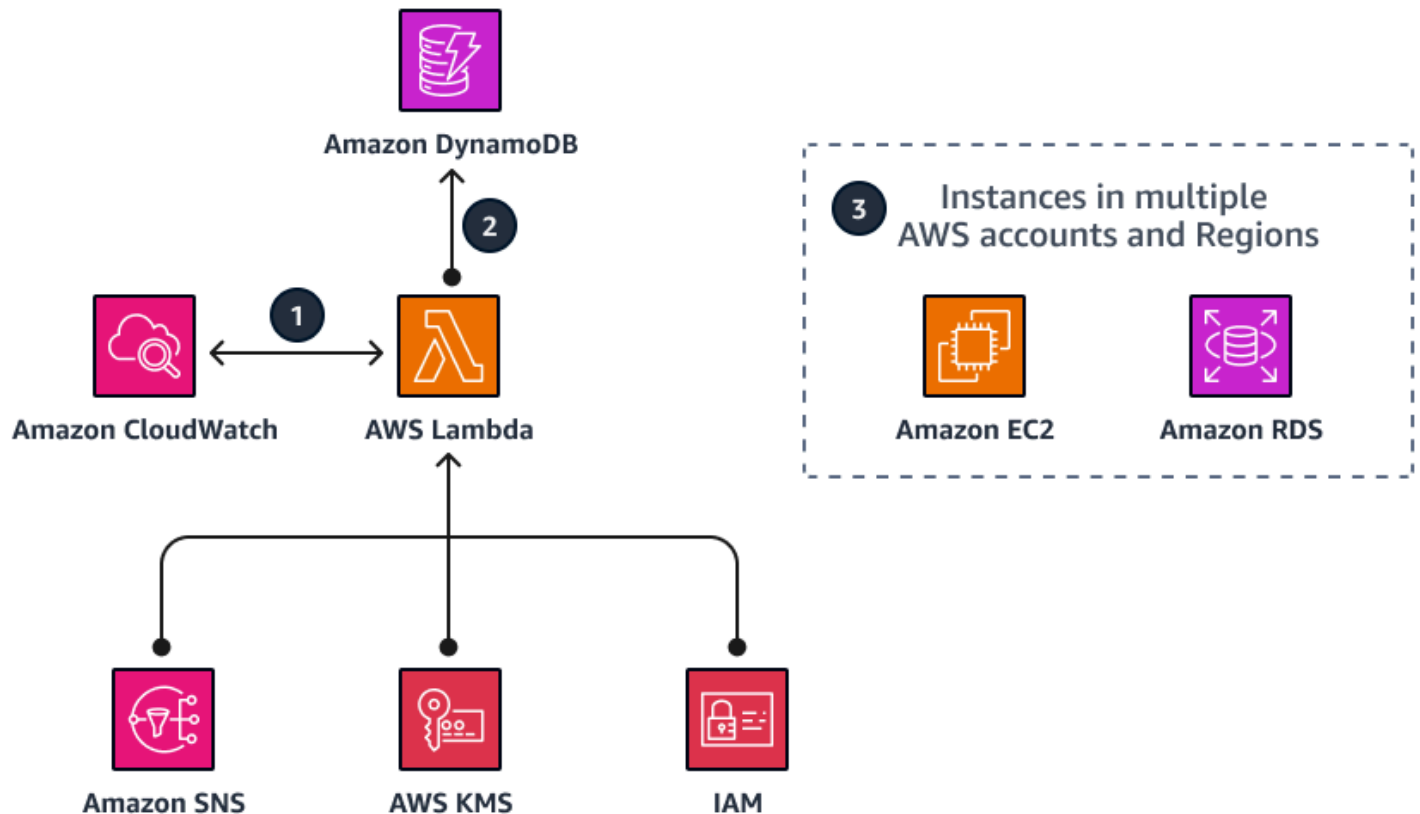
Cette section explique comment déployer et configurer le planificateur d'instances sur la AWS base de l'exemple de scénario décrit dans la section précédente consacrée aux scénarios d'optimisation des coûts. Nous vous recommandons de suivre les étapes suivantes pour optimiser vos coûts en utilisant le planificateur d'instance sur : AWS

1. Lancez la pile Instance Scheduler
2. Configuration des périodes

3. Configuration des plannings

4. Baliser les instances de

Le schéma d'architecture suivant montre ce qui est créé dans la pile AWS Cloud by the Instance Scheduler.



Le diagramme montre les étapes du flux de travail suivantes :

1. Un AWS CloudFormation modèle configure un CloudWatch événement Amazon selon un intervalle que vous définissez. Cet événement invoque une AWS Lambda fonction. Lors de la configuration, vous définissez les comptes Régions AWS et. Vous définissez également une balise personnalisée qu'Instance Scheduler AWS utilise pour associer les plannings aux instances Amazon EC2, aux instances Amazon RDS et aux clusters applicables.
2. Les valeurs de configuration du planning sont stockées dans Amazon DynamoDB, et la fonction Lambda les récupère à chaque exécution. Vous pouvez ensuite appliquer la balise personnalisée aux instances applicables.
3. Lors de la configuration initiale de l'Instance Scheduler, vous définissez une clé de balise pour identifier les instances Amazon EC2 et Amazon RDS applicables. Lorsque vous créez un

calendrier, le nom que vous spécifiez est utilisé comme valeur de balise identifiant le calendrier que vous souhaitez appliquer à la ressource balisée.

Lancez la pile Instance Scheduler

Cette section explique comment lancer la CloudFormation pile pour le planificateur d'instances sur AWS.

Note

Vous êtes responsable du coût de l' Services AWS utilisation lors de l'exécution d'Instance Scheduler sur AWS. En janvier 2023, le coût d'exécution de cette solution avec les paramètres par défaut dans la us-east-1 région était d'environ 9,90\$ par mois pour les frais Lambda, ou moins si vous disposez d'un crédit d'utilisation mensuel pour le niveau gratuit Lambda. Pour plus d'informations, consultez la section Coûts du [guide d' AWS implémentation du planificateur d'instance](#) dans la bibliothèque de AWS solutions.

Pour lancer la pile de planificateurs d'instances, procédez comme suit.

1. Connectez-vous au [AWS Management Console](#) et choisissez [Launch solution](#) (modèle téléchargeable) pour lancer le `instance-scheduler-on-aws` template CloudFormation modèle.

Note

Vous pouvez également [télécharger le modèle](#) comme point de départ pour votre propre implémentation.

2. Le modèle est lancé par défaut dans la région USA Est (Virginie du Nord). Pour lancer le planificateur d'instance dans une autre région, utilisez le sélecteur de région dans la barre de navigation de la console.

Note

Cet exemple utilise la région Asie-Pacifique (Singapour).

3. Sur la page Créer une pile, dans la section Prérequis - Préparer le modèle, vérifiez que l'option Le modèle est prêt est sélectionnée. Dans la section Source du modèle, vérifiez que l'option URL Amazon S3 est sélectionnée.
4. Vérifiez que l'URL du modèle est correcte dans la zone de texte URL Amazon S3, puis choisissez Next.
5. Sur la page Spécifier les détails de la pile, attribuez un nom à votre pile de solutions. Pour plus d'informations sur les limites de dénomination des caractères, consultez la section [Limites IAM et STS](#) dans la Gestion des identités et des accès AWS documentation (IAM). Le nom de la pile pour l'exemple présenté dans ce guide est appelé MyInstanceScheduler.

Note

Le nom de la pile ne peut pas contenir plus de 28 caractères.

6. Sous Paramètres, passez en revue les paramètres du modèle et modifiez-les si nécessaire.
7. Choisissez Next (Suivant). Sur la page Configurer les options de pile, choisissez Suivant.
8. Sur la page Révision, vérifiez et confirmez les paramètres. Cochez la case indiquant que le modèle créera des ressources IAM.
9. Choisissez Créer pour déployer la pile.

Configuration des périodes

Après avoir déployé le CloudFormation modèle, la solution crée une table DynamoDB qui contient des exemples de règles et de calendriers de période que vous pouvez utiliser comme référence pour créer vos propres règles et calendriers de période personnalisés. Pour un exemple de configuration de période, consultez la section [Exemples de plannings](#) dans le planificateur d'instances sur AWS la documentation.


Pour terminer l'étape de ce scénario, vous devez générer des périodes correspondant à chaque charge de travail et à leurs besoins spécifiques. Par exemple :

```
Period 1 (Workload A):  
  Name: retail-hours  
  Days: Monday to Sunday  
  Hours: 1100 - 2300  
Period 2 (Workload B):  
  Name: office-hours  
  Days: Monday to Friday
```

Hours: 0800 - 1800

Pour configurer les périodes, procédez comme suit :

1. Connectez-vous à la console [DynamoDB](#) et assurez-vous que vous vous trouvez dans la même région que celle dans laquelle vous avez lancé le modèle pour CloudFormation le planificateur d'instances. AWS
2. Dans le volet de navigation, choisissez Tables, puis sélectionnez la table nommée ConfigTable.
3. Choisissez Explorer les éléments de table.
4. Pour créer une période pour les heures de bureau, sélectionnez une période pour l'élément des heures de bureau.
5. Sur la page Modifier l'élément, remplacez la valeur de begintime par 0800 et celle de endtime par 1800. Laissez la valeur par défaut en place pour les jours de semaine.

 Note

Les valeurs de début et de fin déterminent le moment où les instances doivent être démarrées et arrêtées, tandis que la valeur des jours de la semaine détermine les jours de la semaine auxquels ce calendrier s'applique (du lundi au vendredi pour cet exemple).

6. Sélectionnez Enregistrer les modifications.
7. Pour dupliquer la période des heures de bureau et l'utiliser pour créer une nouvelle période pour les heures de vente au détail, sélectionnez une période pour l'élément des heures de bureau. Ensuite, dans le menu Actions, choisissez Dupliquer l'élément.
8. Modifiez les attributs en fonction de vos besoins. Les attributs suivants sont utilisés pour répondre aux exigences de l'exemple de scénario :

```
type: period
name: retail-hours
begintime: 11:00
description: Retail hours
endtime: 23:00
weekdays: mon-sun
```

9. Choisissez Créer un élément.
10. Dans ConfigTableDynamoDB, identifiez les deux périodes que vous venez de créer répertoriées dans les listes d'éléments.

Configuration des plannings

Dans le contexte d'Instance Scheduler activé AWS, les plannings font référence à l'application d'une ou de plusieurs périodes et au fuseau horaire correspondant. Ces plannings sont ensuite affectés à vos instances sous forme de balises. Cette section explique comment créer deux plannings (illustrés ci-dessous) adaptés aux différents modèles temporels des deux exemples de charges de travail, puis comment associer les plannings aux périodes que vous avez créées dans la section précédente.

```
Schedule 1:  
  Name: singapore-office-hours  
  Period: office-hours  
  Timezone: Asia/Singapore  
Schedule 2:  
  Name: singapore-retail-hours  
  Period: retail-hours  
  Timezone: Asia/Singapore
```

Pour créer et configurer des plannings, procédez comme suit :

1. Connectez-vous à la console [DynamoDB](#) et assurez-vous que vous vous trouvez dans la même région que celle dans laquelle vous avez lancé le modèle pour CloudFormation le planificateur d'instances. AWS
2. Dans le volet de navigation, choisissez Tables, puis sélectionnez la table nommée ConfigTable.
3. Choisissez Explorer les éléments de table.
4. Pour dupliquer le calendrier des heures de bureau au Royaume-Uni et l'utiliser pour créer un nouveau calendrier pour vos heures de bureau (heures de bureau à Singapour, par exemple), sélectionnez le calendrier pour uk-office-hours l'article. Ensuite, dans le menu Actions, choisissez Dupliquer l'élément.
5. Modifiez les attributs en fonction de vos besoins. Les attributs suivants sont utilisés pour répondre aux exigences de l'exemple de scénario :

```
type: schedule  
name: singapore-office-hours  
description: Office hours in Singapore  
periods: office-hours  
timezone: Asia/Singapore
```

6. Choisissez Créer un élément.

7. Répétez les étapes 4 à 6 pour créer un calendrier pour les heures de vente au détail à Singapour en utilisant les valeurs d'attribut suivantes :

```
type: schedule
name: singapore-retail-hours
description: Retail hours in Singapore
periods: retail-hours
timezone: Asia/Singapore
```

8. Dans ConfigTableDynamoDB, identifiez les deux plannings et les deux périodes que vous avez créés.

Baliser les instances de

Après avoir établi vos plannings, vous devez utiliser des tags pour allouer les plannings aux instances spécifiques que vous souhaitez utiliser. Vous pouvez utiliser l'éditeur de balises intégré [Groupes de ressources AWS](#) pour générer et attribuer des balises à vos instances Amazon EC2.

1. Connectez-vous au [AWS Management Console](#) et assurez-vous que vous vous trouvez dans la même région que celle dans laquelle vous avez lancé le CloudFormation modèle.
2. Ouvrez la [console Resource Groups](#). Dans le volet de navigation, développez Tagging, puis choisissez Tag Editor.
3. Dans la section Trouver des ressources à étiqueter, pour Régions, choisissez vos régions. Pour les types de ressources, choisissez Amazon EC2 ou Amazon RDS. Ce scénario se concentre sur les instances Amazon EC2 dans la charge de travail A. L'équipe marketing utilise la charge de travail A dans la région de Singapour. Les ressources pour cette charge de travail sont déjà étiquetées avec une clé de département et une valeur marketing. Vous pouvez utiliser cette balise pour rechercher les instances.
4. Choisissez Rechercher des ressources.
5. Sélectionnez les instances que vous souhaitez inclure dans le calendrier dans la liste des résultats de recherche, puis choisissez Gérer les balises des ressources sélectionnées.
6. Dans la section Modifier les balises de toutes les ressources sélectionnées, choisissez Ajouter une balise pour ajouter les balises de planification du planificateur d'instances à vos instances EC2. Vous pouvez utiliser les clés de balise et les valeurs correspondant à schedulea (créées précédemment dans DynamoDB).
7. Pour la touche Tag, ajoutez Schedule. Dans le champ Valeur du tag, entrez singapore-retail-hours.
8. Choisissez Vérifier et appliquer les modifications apportées à la balise.

9. Pour appliquer la balise à toutes les instances EC2 que vous avez sélectionnées, choisissez Appliquer les modifications à toutes les instances sélectionnées.
10. Répétez les étapes 3 à 9 pour tous les programmes supplémentaires que vous souhaitez appliquer.

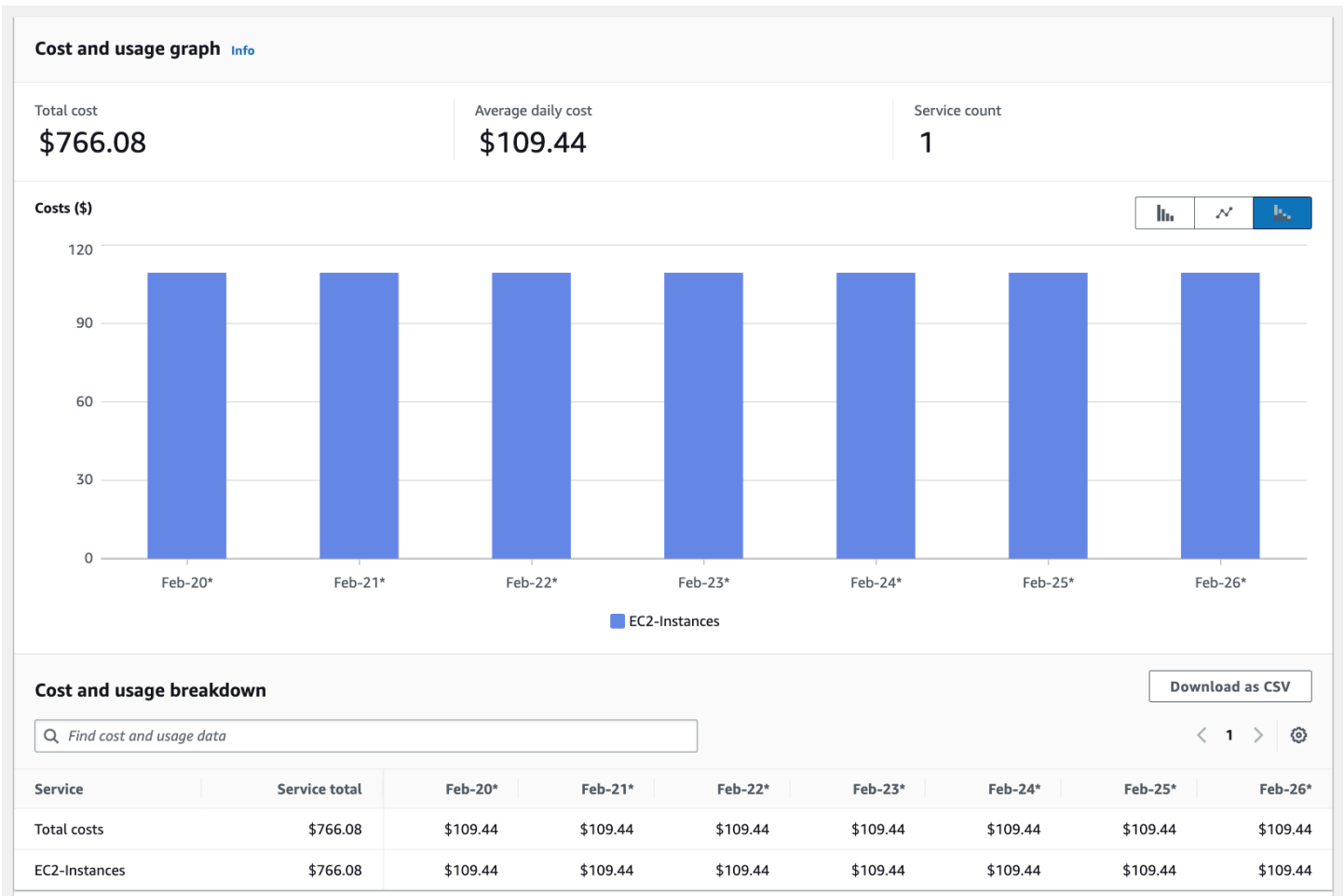
Validation des résultats

Nous vous recommandons de l'utiliser [AWS Cost Explorer](#) pour mesurer les avantages en termes de coûts liés à l'utilisation d'Instance Scheduler sur AWS. Vous pouvez utiliser Cost Explorer pour effectuer les opérations suivantes :

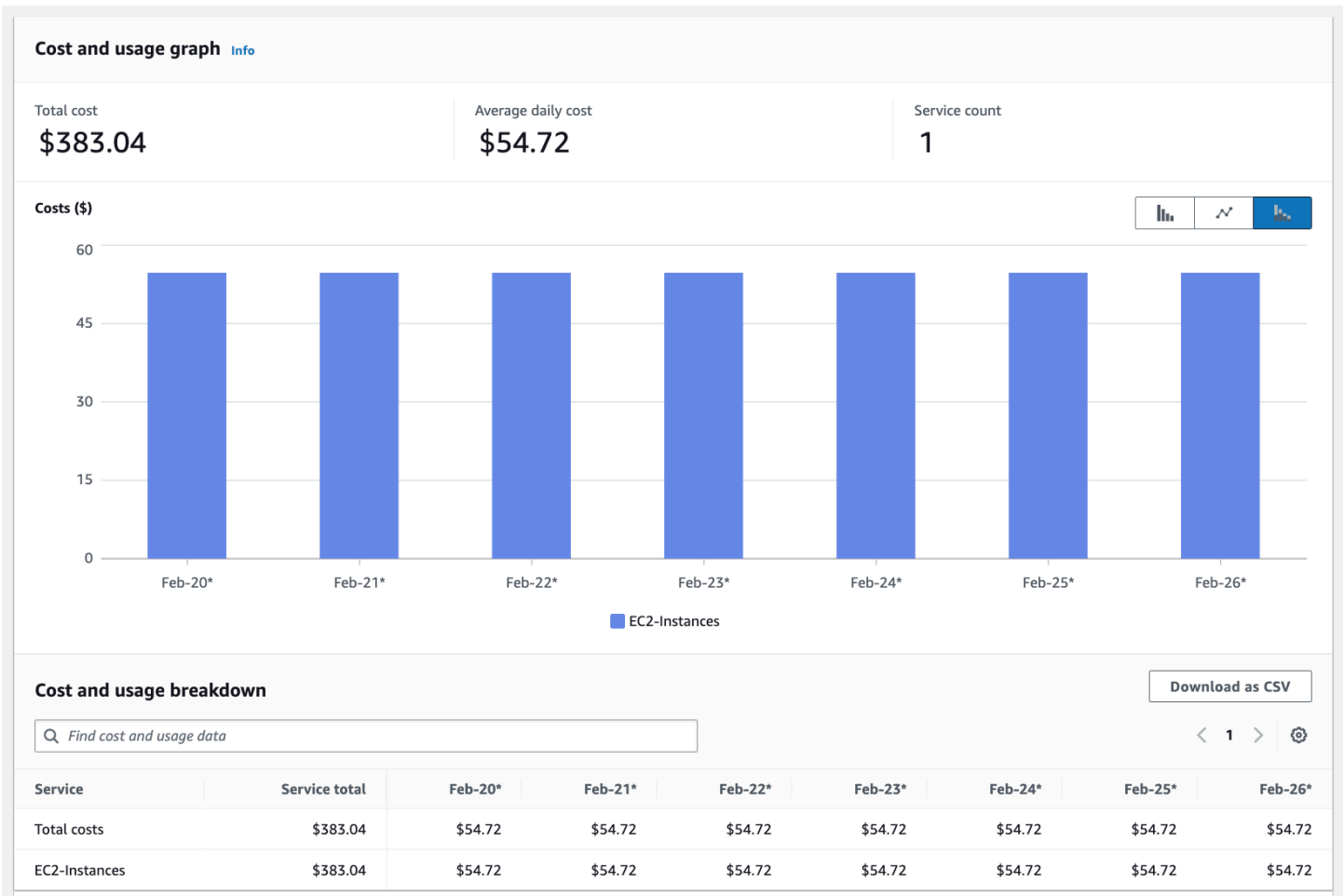
- Consultez et analysez les coûts associés à vos instances EC2, y compris les instances gérées par Instance Scheduler.
- Filtrez votre affichage Cost Explorer par balises afin de vous concentrer sur des charges de travail spécifiques et d'obtenir une vue détaillée des économies réalisées grâce à Instance Scheduler.
- Découvrez l'impact financier de l'utilisation d'Instance Scheduler.
- Identifiez les opportunités d'optimisation des coûts et prenez des décisions basées sur les données pour optimiser vos AWS dépenses.

Les graphiques suivants illustrent le coût d'exploitation des charges de travail A et B sur une période de sept jours (du lundi au dimanche) avant l'optimisation à l'aide du planificateur d'instance.

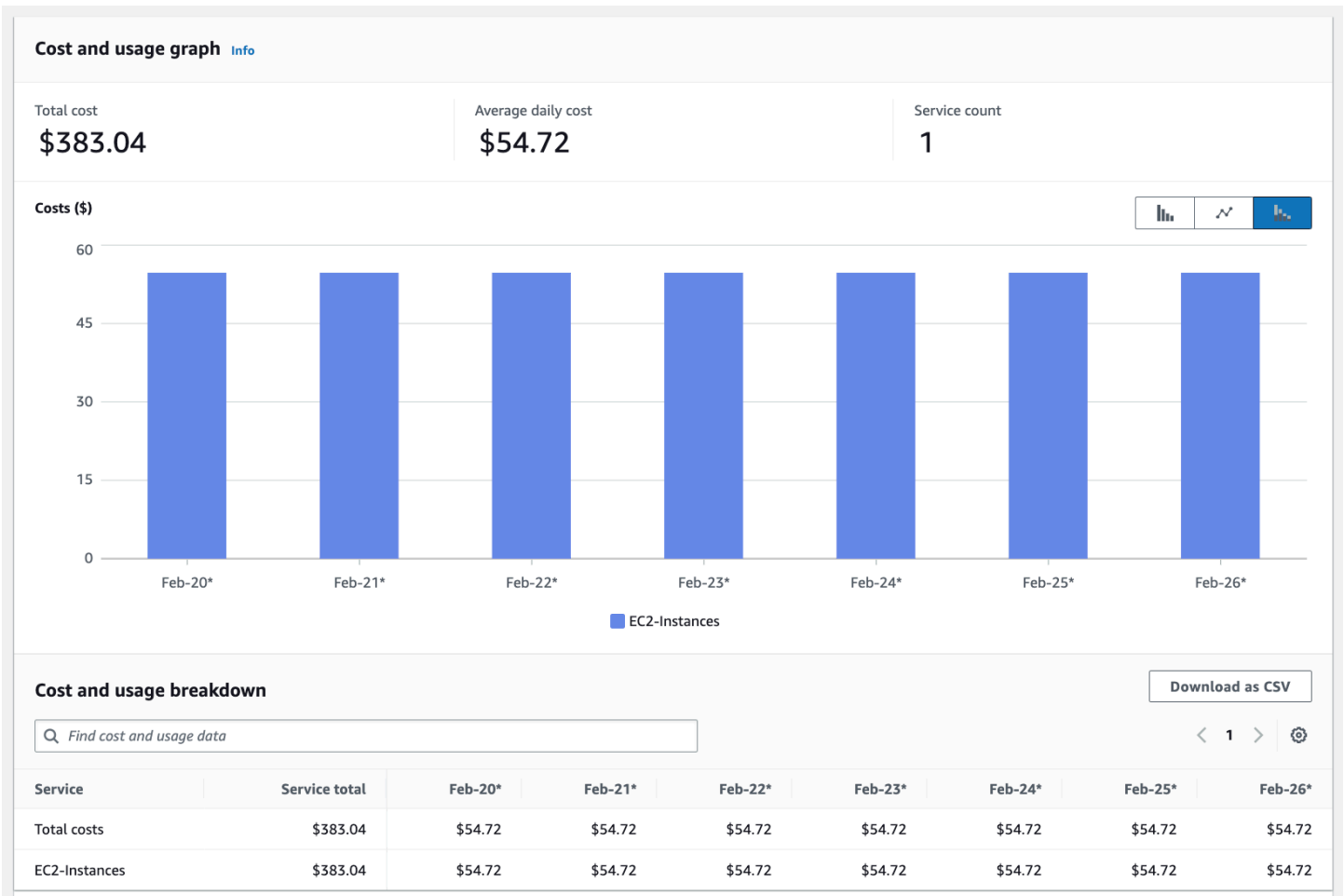
Total des dépenses combinées liées aux charges de travail A et B



Charges de travail A

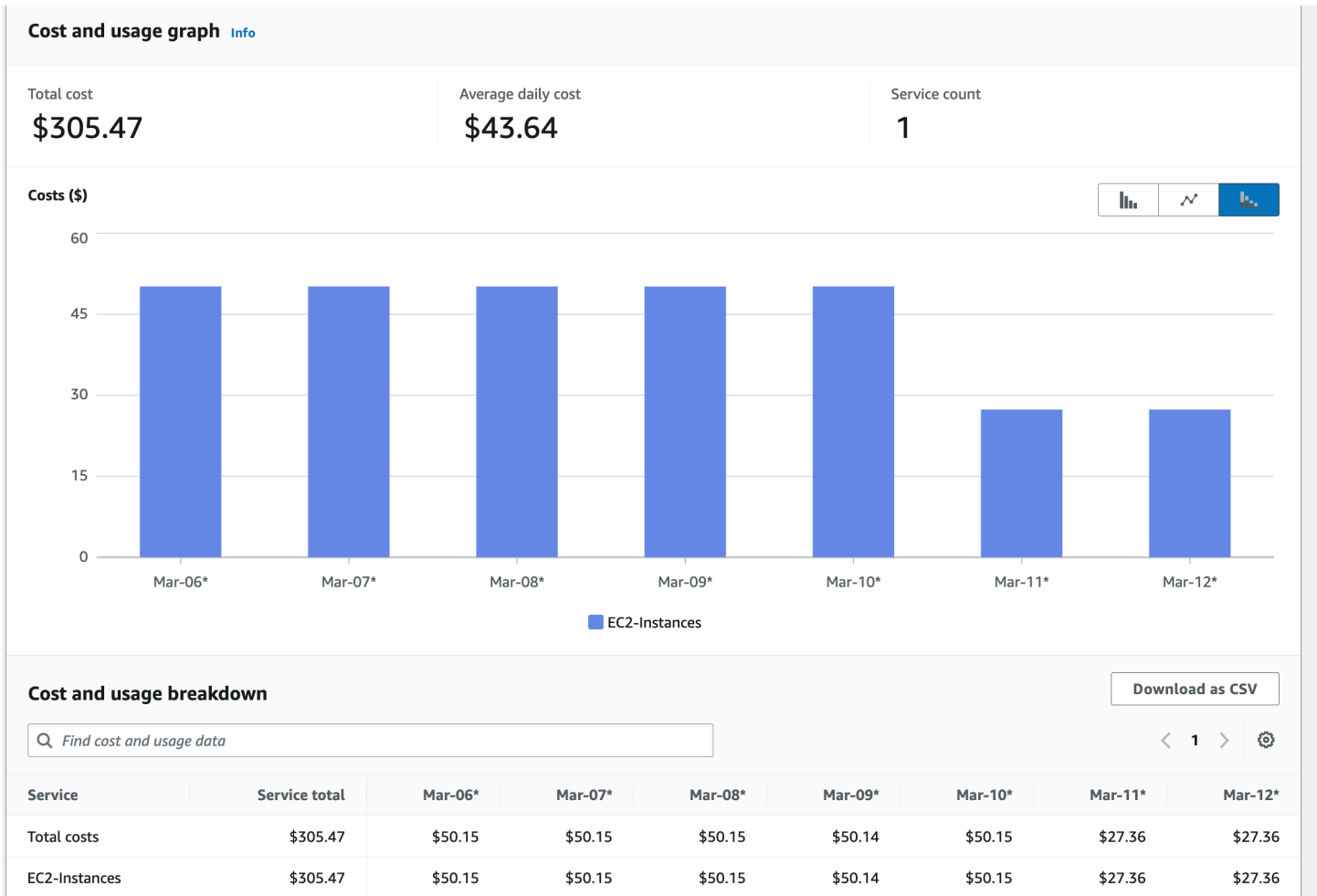


Charge de travail B et dépenses

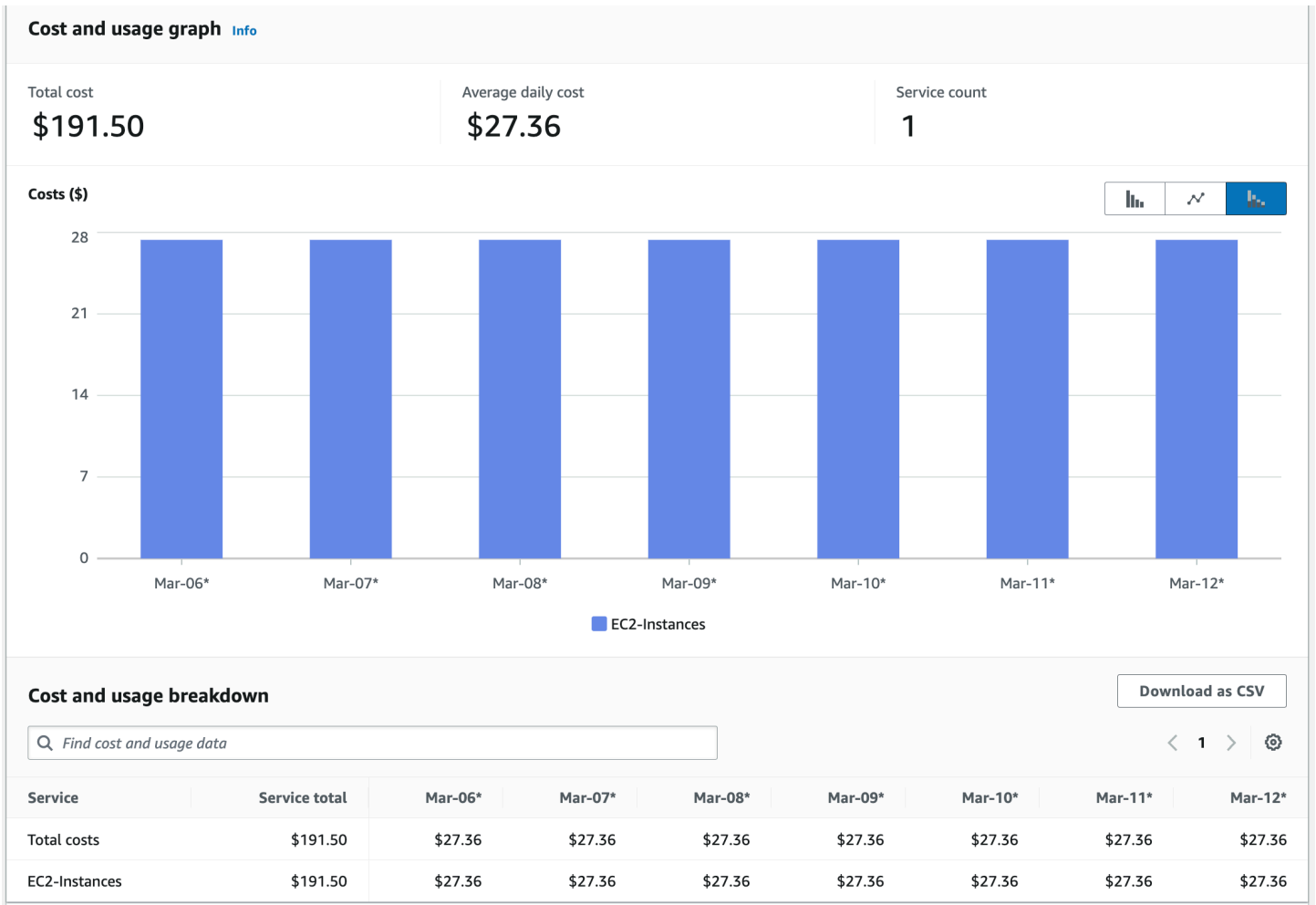


Dans ce scénario, Cost Explorer montre les réductions de coûts résultant de l'implémentation d'Instance Scheduler sur AWS. Les graphiques suivants montrent les coûts opérationnels des charges de travail A et B pour une période de sept jours (du lundi au dimanche) après l'optimisation.

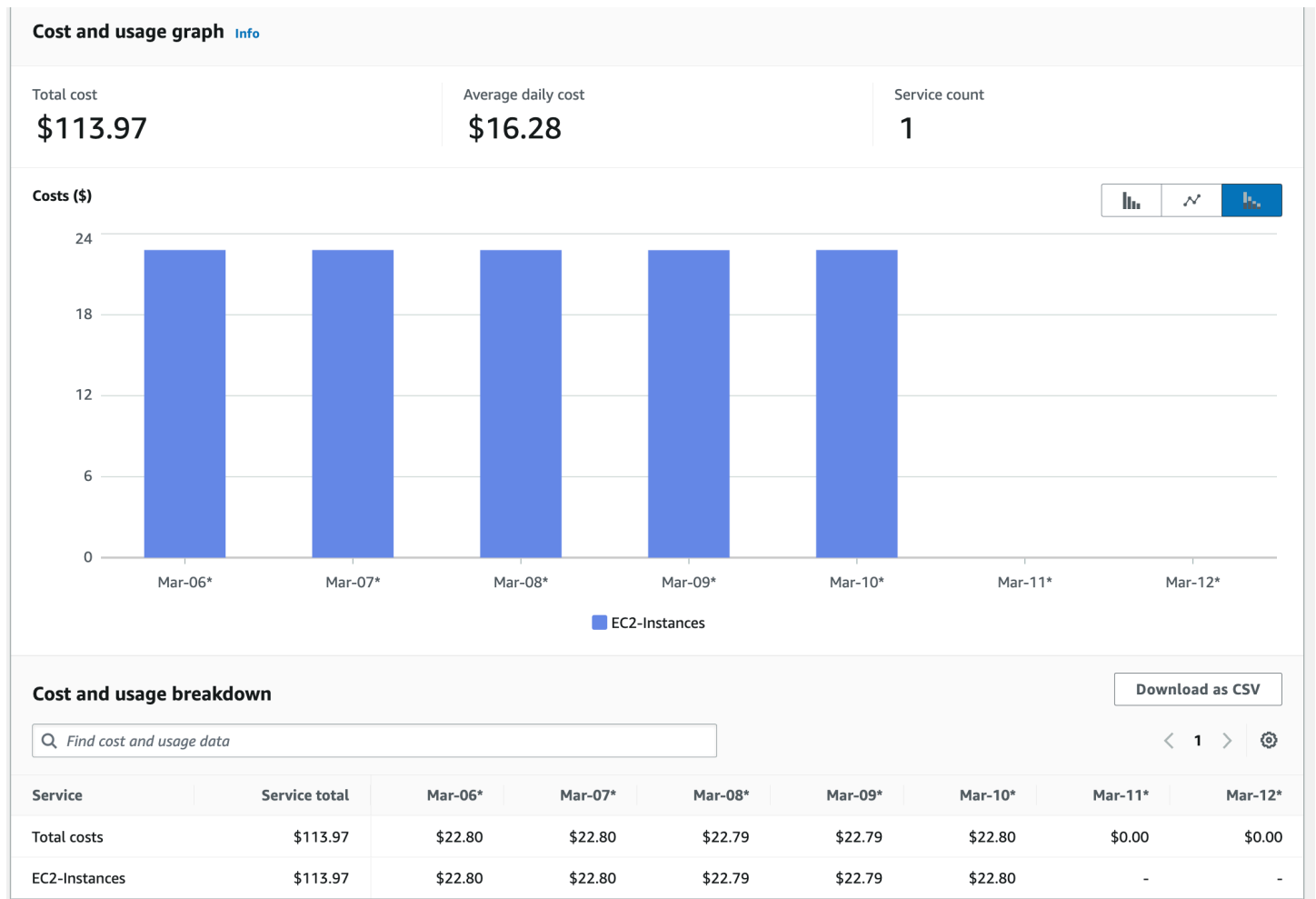
Dépenses totales combinées liées aux charges de travail A et B



Charges de travail A



Charge de travail B et dépenses



Ressources supplémentaires

- [Automatisez le démarrage et l'arrêt AWS des instances](#) (planificateur d'instances sur la AWS documentation)
- [Retour à l'essentiel : utilisation d'un planificateur d'instance pour contrôler les coûts des ressources Amazon EC2 et Amazon RDS](#) () YouTube
- [Marquer vos AWS ressources](#) (Guide de l'utilisateur du balisage AWS des ressources)
- [Analyser vos coûts avec AWS Cost Explorer](#) (AWS Billing and Cost Management documentation)

Charges de travail Windows de la bonne taille

Présentation de

Le bon dimensionnement est l'un des outils les plus efficaces pour réduire les coûts. AWS propose différentes méthodes pour collecter des informations sur le dimensionnement correct, allant de l'examen des charges de travail potentielles à l'aide d'une [évaluation de l'AWS optimisation et des licences \(AWS OLA\)](#) à l'examen des charges de travail existantes à l'aide de [AWS Cost Explorer](#)

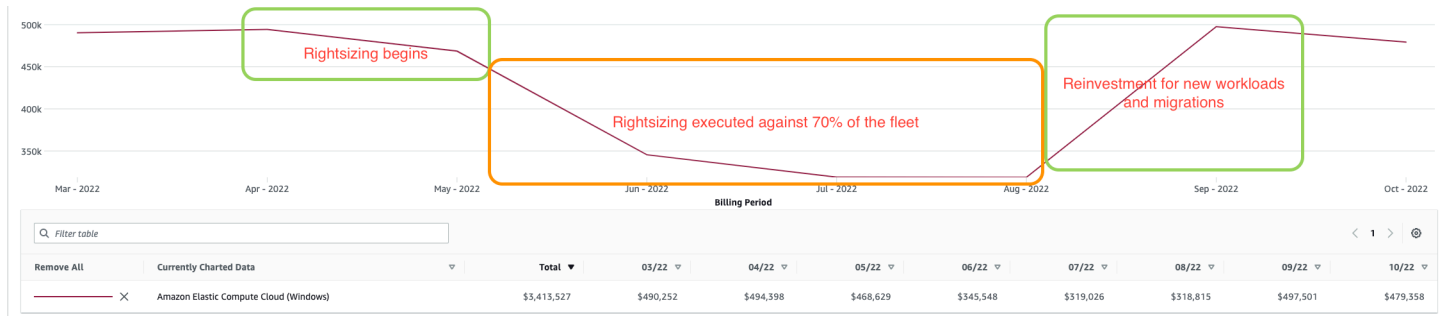
Cette section explique comment identifier les opportunités [Optimiseur de calcul AWS](#) de dimensionnement adaptées à Amazon EC2. Compute Optimizer permet d'éviter le surprovisionnement et le sous-provisionnement pour les types de ressources suivants : AWS

- [Types d'instances Amazon Elastic Compute Cloud \(Amazon EC2\)](#)
- [Volumes d'Amazon Elastic Block Store \(Amazon EBS\)](#)
- Services [Amazon Elastic Container Service \(Amazon ECS\)](#) sur AWS Fargate
- [AWS Lambda](#) fonctions basées sur les données d'utilisation fournies par [Amazon CloudWatch](#)

Scénario d'optimisation des coûts

Mesurer l'efficacité de la bonne taille peut s'avérer difficile, car les efforts de dimensionnement appropriés peuvent être dirigés vers une application spécifique, une équipe ou l'ensemble de l'organisation. Prenons l'exemple d'une organisation qui migre plusieurs milliers d'instances vers AWS laquelle 90 % de son parc est composé de charges de travail Windows. L'organisation peut utiliser Compute Optimizer pour analyser son parc et découvrir un surprovisionnement important sur ses comptes et Régions AWS. Ils peuvent ensuite utiliser l'[AWS Systems Manager automatisation](#) pour dimensionner correctement leur parc pendant plusieurs fenêtres de maintenance. L'entreprise parvient ainsi à ajuster le type d'instance à la bonne taille pour 70 % de son parc et réalise des économies de 35 %.

Le tableau de bord suivant illustre les économies réalisées en plusieurs mois grâce à la mise en œuvre stratégique par cet exemple d'organisation des recommandations de Compute Optimizer en matière de dimensionnement. Leur objectif était d'exploiter leurs charges de travail existantes aussi efficacement que possible afin de reprendre une migration bloquée depuis un centre de données en colocation approchant de la fin de son contrat.



Recommandations en matière d'optimisation des coûts

Nous vous recommandons de suivre les étapes suivantes pour optimiser vos coûts à l'aide de Compute Optimizer :

- Activer Compute Optimizer
- Activer la collecte de métriques de mémoire pour les nœuds Windows
- Suivez les recommandations de Compute Optimizer
- Étiquetez les instances pour les dimensionner correctement
- Activez l'étiquette de répartition des coûts pour qu'elle fonctionne avec les outils AWS de facturation
- Mettez en œuvre des recommandations de dimensionnement appropriées grâce AWS Systems Manager à l'automatisation
- Envisagez d'autres méthodes de redimensionnement
- Révision des coûts avant et après dans Cost Explorer

Activer Compute Optimizer

Vous pouvez activer [Compute Optimizer](#) au niveau de l'organisation ou au niveau du compte unique dans AWS Organizations. La configuration à l'échelle de l'organisation fournit des rapports permanents pour les instances nouvelles et existantes sur l'ensemble de votre flotte pour tous les comptes membres. Cela permet de faire de la bonne taille une activité récurrente plutôt qu'une point-in-time activité.

Niveau de l'organisation

Pour la plupart des entreprises, le moyen le plus efficace d'utiliser Compute Optimizer est au niveau de l'organisation. Cela fournit une visibilité multicompte et multirégionale de votre organisation et

centralise les données dans une seule source à des fins de révision. Pour activer cette fonctionnalité au niveau de l'organisation, procédez comme suit :

1. Connectez-vous à votre [compte de gestion Organizations](#) avec un rôle disposant des [autorisations requises](#) et choisissez d'activer l'accès à tous les comptes de cette organisation. [Toutes les fonctions doivent être activées](#) pour votre organisation.
2. Après avoir activé le compte de gestion, vous pouvez vous connecter au compte, voir tous les autres comptes membres et consulter leurs recommandations.

Note

Il est recommandé de configurer un [compte d'administrateur délégué](#) pour Compute Optimizer. Cela vous permet d'appliquer le principe du moindre privilège. Ainsi, vous pouvez minimiser l'accès au compte de gestion de l'organisation tout en fournissant un accès au service à l'échelle de l'organisation.

Niveau de compte unique

Si vous ciblez un compte dont les coûts sont élevés mais que vous n'y avez pas accès AWS Organizations, vous pouvez toujours activer Compute Optimizer pour ce compte et cette région. Pour en savoir plus sur le processus d'opt-in, consultez [Getting started with Optimiseur de calcul AWS](#) dans la documentation de Compute Optimizer.

Activer la collecte de métriques de mémoire pour les nœuds Windows

Les métriques de mémoire fournissent à Compute Optimizer les indicateurs essentiels nécessaires pour formuler des recommandations de dimensionnement bien informées au sein de votre organisation. Cela est dû à l'analyse du processeur, de la mémoire, du réseau et du stockage effectuée avant de proposer une recommandation.

Pour transmettre les métriques de mémoire des instances Windows EC2 à Compute Optimizer, vous devez activer CloudWatch l'agent et configurer les métriques de mémoire afin qu'elles soient collectées toutes les 60 secondes. L'utilisation des métriques de mémoire avec CloudWatch.

Activer l' CloudWatch agent et configurer les métriques de mémoire

Téléchargez le [ComputeOptimizefichier .yaml](#). Vous pouvez utiliser ce fichier pour activer la collecte de mémoire pour toutes les instances de votre compte. Le fichier modèle génère les composants suivants :

- [AWS Systems Manager Parameter Store](#) : il stocke la configuration de l' CloudWatch agent requis pour collecter les métriques de mémoire.
- Gestion des identités et des accès AWS rôle (IAM) AWS Systems Manager associé à des [politiques AWS gérées](#) — Ceci est destiné au document Systems Manager Automation.
- [AWS Systems Manager documents](#) — Ceci installe et configure l' CloudWatchagent (en remplaçant toute CloudWatch configuration existante).
- AWS Systems Manager Association [State Manager](#) : cela permet aux documents Systems Manager de s'exécuter sur toutes les instances de votre compte.

Important

L'exécution de ce modèle remplace toute CloudWatch configuration existante sur les instances.

Procédez ensuite comme suit :

1. Connectez-vous à la [CloudFormation console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation, choisissez Stack (Piles).
3. Choisissez Créer une pile, puis sélectionnez Avec les ressources existantes (ressources d'importation).
4. Choisissez Suivant.
5. Pour Ressource de modèle, sélectionnez Télécharger un fichier modèle.
6. Choisissez un fichier, puis importez le ComputeOptimize .yaml fichier.
7. Choisissez Suivant.
8. Sur la page Spécifier les détails de la pile, dans Nom de la pile, entrez le nom de votre pile, puis choisissez Next.
9. Sur la page Identifier les ressources, entrez les valeurs d'identification des ressources que vous importez.

10. Choisissez Importer des ressources.

11. Une fois la pile déployée, choisissez l'onglet Sorties pour trouver la clé, la valeur et la description de votre association.


Suivez les progrès de l'association

1. Une fois le déploiement de la CloudFormation pile terminé, ouvrez la [console Systems Manager](#).
2. Dans le volet de navigation, dans la section Gestion des nœuds, choisissez State Manager.
3. Sur la page Associations, choisissez l'identifiant de votre association.
4. Sélectionnez l'onglet Historique d'exécution.
5. Dans la colonne ID d'exécution, choisissez l'ID d'exécution de votre association. Le statut doit être Réussite.

Afficher les statistiques dans CloudWatch

Nous vous recommandons d'attendre au moins cinq minutes pour que les statistiques soient renseignées. CloudWatch

1. Ouvrez la [CloudWatch console](#).
2. Dans le volet de navigation, développez la section Mesures, puis choisissez Toutes les mesures.
3. Vérifiez que les métriques apparaissent sous l'espace de CWAgentnoms.

 Note

Pour appliquer les paramètres à toute nouvelle instance, réexécutez l'association.

Suivez les recommandations de Compute Optimizer

Prenons un exemple qui met l'accent sur les modifications de taille appropriées au sein d'un seul compte et d'une seule région. Dans cet exemple, Compute Optimizer est activé au niveau de l'organisation sur tous les comptes. N'oubliez pas que le bon dimensionnement est un processus perturbateur qui, dans la plupart des cas, est effectué avec précision par les propriétaires de l'application au cours d'une période de maintenance planifiée de plusieurs semaines.

Si vous accédez à Compute Optimizer depuis le compte de gestion d'une organisation (comme indiqué dans les étapes suivantes), vous pouvez choisir le compte que vous souhaitez examiner. Dans cet exemple, six instances sont exécutées dans un seul compte dans la us-east-1 région. Les six instances sont surprovisionnées. L'objectif est de redimensionner les instances en fonction des recommandations de Compute Optimizer.

Identifiez les instances surprovisionnées et exportez les détails des recommandations

1. Connectez-vous à la console [Compute Optimizer AWS Management Console](#) et ouvrez-la.
2. Dans le panneau de navigation, sélectionnez Dashboard (Tableau de bord).
3. Dans le champ de recherche de la page Tableau de bord, saisissez Region=US East (Virginie du Nord). Entrez ensuite Findings=OverProvisioned. Ces filtres vous permettent de voir toutes les instances surprovisionnées dans la us-east-1 région.
4. Pour consulter les recommandations détaillées concernant les instances EC2 surprovisionnées, faites défiler la page vers le bas jusqu'à la fiche des instances EC2, puis choisissez Afficher les recommandations.
5. Choisissez Exporter et enregistrez le fichier pour une utilisation future.
6. Pour le compartiment S3, entrez le nom du compartiment Amazon S3 que vous souhaitez utiliser comme destination pour le fichier d'exportation.

Note

Pour enregistrer des recommandations en vue d'une future révision, vous devez disposer d'un compartiment S3 dans lequel Compute Optimizer pourra écrire dans chaque région. Pour plus d'informations, consultez la [politique relative aux compartiments Amazon S3 Optimiseur de calcul AWS dans la](#) documentation de Compute Optimizer.

7. Dans la section Filtres d'exportation, cochez la case Inclure les recommandations pour tous les comptes membres de l'organisation.
8. Pour le type de ressource, choisissez les instances EC2.
9. Dans la section Colonnes à inclure, cochez la case Tout sélectionner.
10. Cliquez sur Exporter.

Choisissez des instances en fonction des recommandations

Les recommandations relatives aux instances sont basées sur les indicateurs de performance collectés et analysés par Compute Optimizer. Il est essentiel de connaître les charges de travail exécutées sur l'instance afin de choisir la meilleure instance. [Cet exemple suppose que vous pouvez choisir parmi la dernière génération d'instances Amazon EC2 R6i, R5 et T3](#). Les instances T3 sont évolutives et ont des capacités de bande passante réseau plus faibles. Les instances R5 et R6 ont le même coût horaire et sont quasiment identiques. Cependant, l'instance R6 dispose d'une capacité de bande passante réseau supérieure, intègre la dernière génération de processeurs Intel et offre le même encombrement de calcul que l'instance R5. Dans cet exemple, R6 est la meilleure option à choisir pour le redimensionnement.

1. Dans la [console Compute Optimizer](#), sélectionnez Recommendations for EC2 instances dans la barre de navigation. Cette page présente une comparaison du type d'instance actuel avec les options recommandées pour le remplacer.
2. Pour obtenir l'ID de l'instance que vous souhaitez dimensionner correctement, ouvrez la [console Amazon S3](#) depuis le compte de gestion AWS Organizations.
3. Dans le volet de navigation, choisissez Buckets, puis choisissez le bucket que vous utilisez pour stocker vos résultats exportés.
4. Dans l'onglet Objets, sélectionnez votre fichier d'exportation dans la liste des objets, puis choisissez Télécharger.
5. Pour extraire les informations d'instance du fichier, vous pouvez utiliser le bouton Texte en colonnes de l'onglet Données de Microsoft Excel.

Note

IDs Les instances sont représentées par Amazon Resource Names (ARNs). Assurez-vous de définir le délimiteur sur «/» et d'extraire l'ID de l'instance. Vous pouvez également écrire un script ou utiliser un environnement de développement intégré (IDE) pour découper l'ARN.

6. Dans Excel, filtrez la colonne de recherche pour afficher uniquement les instances OVER_PROVISIONED. Il s'agit des instances que vous ciblez pour obtenir la bonne taille.
7. Enregistrez l'instance IDs dans un éditeur de texte pour y accéder facilement ultérieurement.

Étiquetez les instances pour les dimensionner correctement

Le balisage de vos charges de travail est un outil puissant pour organiser vos ressources. AWS Les tags vous permettent d'obtenir une visibilité précise des coûts et de faciliter la rétrofacturation. Pour plus d'informations sur les stratégies et les méthodes d'ajout de balises aux AWS ressources, consultez le AWS livre blanc [Meilleures pratiques en matière de balisage AWS des](#) ressources. Dans cet exemple, vous pouvez utiliser l'[éditeur de AWS balises](#) pour effectuer des ajustements de balisage sur les instances surprovisionnées que vous souhaitez cibler pour les redimensionner pendant une période de maintenance. Vous pouvez également utiliser cette balise pour consulter les coûts avant et après le changement.

1. Connectez-vous au AWS Management Console et ouvrez la [Groupes de ressources AWS console](#) du compte contenant les instances à redimensionner.
2. Dans la barre de navigation, dans la section Balisage, choisissez Tag Editor.
3. Pour Régions, sélectionnez votre région cible.
4. Pour les types de ressources, sélectionnez `AWS::EC2::Instance`.
5. Choisissez Rechercher des ressources.
6. Sur la page des résultats de recherche de ressources, sélectionnez toutes les instances que vous souhaitez dimensionner correctement, puis choisissez Gérer les balises des ressources sélectionnées.
7. Choisissez Ajouter une balise.
8. Pour Tag key, saisissez Rightsizing. Dans le champ Valeur du tag, entrez activé. Choisissez ensuite Réviser et appliquez les modifications de balise.

Note

Vous pouvez inclure des métadonnées supplémentaires, telles que Team ou Business Unit, pour faciliter le filtrage ultérieur dans Cost Explorer.

Après avoir créé et appliqué des balises définies par l'utilisateur à vos ressources, l'activation des balises peut prendre jusqu'à 24 heures sur la page des balises de répartition des coûts. Une fois que vous avez sélectionné les balises à activer, l'activation des balises peut prendre 24 heures supplémentaires.

Pour les utilisateurs avancés, vous pouvez l'utiliser [AWS CloudShell](#) dans le compte cible et dans la région pour étiqueter plusieurs instances. Par exemple :

```
bash
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="type-m5"
# Get a list of instance IDs
INSTANCE_IDS=$(aws ec2 describe-instances --query
  "Reservations[].Instances[].InstanceId" --output text)
# Loop through each instance ID and add the tag
for INSTANCE_ID in $INSTANCE_IDS; do
  aws ec2 create-tags --resources $INSTANCE_ID --tags Key=$TAG_KEY,Value=$TAG_VALUE
done
```

Activez l'étiquette de répartition des coûts pour qu'elle fonctionne avec les outils AWS de facturation

Nous vous recommandons d'activer la balise de répartition des coûts définie par l'utilisateur. Cela permet à la balise Rightsizing d'être reconnue et filtrable dans les outils AWS de facturation (par exemple, Cost Explorer et) AWS Cost and Usage Report. Si vous ne l'activez pas, l'option de filtrage des balises et les données ne seront pas disponibles. Pour plus d'informations sur l'utilisation des balises de répartition des coûts, consultez la section [Activation des balises de répartition des coûts définies par l'utilisateur](#) dans la AWS Billing and Cost Management documentation.

1. Connectez-vous à la [AWS Billing console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation, dans la section Facturation, choisissez les balises de répartition des coûts.
3. Dans l'onglet Balises de répartition des coûts définies par l'utilisateur, saisissez Rightsizing.
4. Sélectionnez la touche de balise Rightsizing, puis sélectionnez Activer.

Au bout de 24 heures, le tag devrait apparaître dans Cost Explorer.

Mettez en œuvre les recommandations de dimensionnement appropriées avec Systems Manager Automation

Le redimensionnement est un scénario qui nécessite l'arrêt et le démarrage d'une instance. Dans ce scénario, vous devrez peut-être gérer cette interruption pendant une période de maintenance et avoir besoin de différentes équipes pour gérer leur propre redimensionnement. Avant de modifier

un type d'instance, consultez les [considérations relatives aux types d'instances compatibles](#) dans la documentation Amazon EC2.

Les exemples d'étapes présentés dans cette section mettent en œuvre des recommandations de dimensionnement adaptées par compte et par région à l'aide d'un document d'automatisation de Systems Manager appelé [AWS- ResizeInstance](#). Cette approche est typique de la plupart des entreprises, car la plupart d'entre elles ont besoin de différents types d'instances pour des objectifs différents. Vous pouvez également utiliser le même document AWS-ResizeInstance d'automatisation pour cibler les déploiements à un ou plusieurs comptes.

1. Connectez-vous à la [console Systems Manager AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation, dans la section Ressources partagées, sélectionnez Documents.
3. Dans la barre de recherche, saisissez AWS- ResizeInstance, puis choisissez AWS- dans les résultats ResizeInstance de recherche.
4. Choisissez Execute automation (Exécuter l'automatisation).
5. Sur la page Exécuter le runbook d'automatisation, sélectionnez Exécution simple.
6. Dans la section Paramètres d'entrée, entrez InstanceId et InstanceType. Conservez le reste des valeurs par défaut.
7. Choisissez Execute, puis attendez que l'automatisation passe par les étapes permettant de modifier le type d'instance.

Envisagez d'autres méthodes de redimensionnement

Si vous utilisez un modèle de lancement pour déployer vos instances, vous pouvez mettre à jour le modèle de lancement avec le type d'instance de la bonne taille, puis effectuer une actualisation de l'instance pour remplacer les instances par la version de la bonne taille.

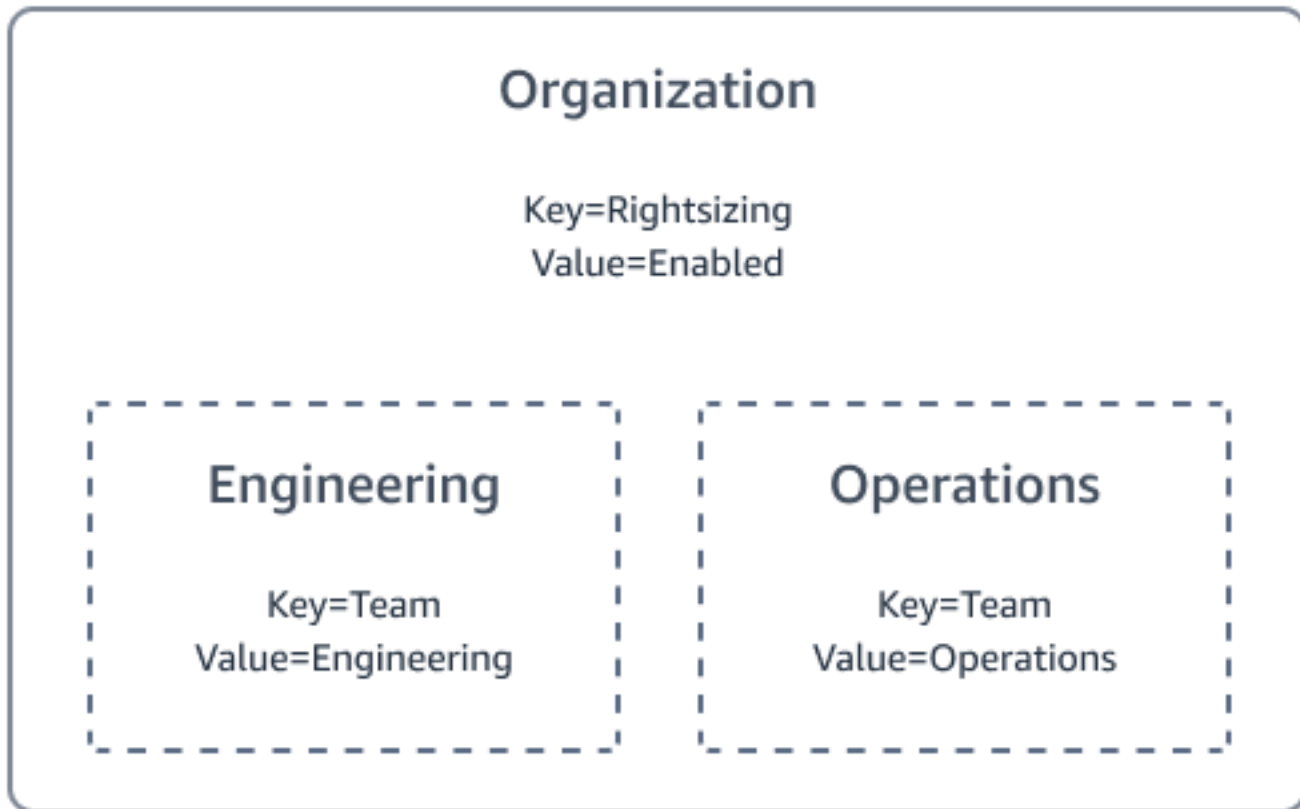
Si vous envisagez d'utiliser le processus de dimensionnement approprié pour plusieurs comptes et régions, vous devez créer un document personnalisé d'automatisation de Systems Manager. Ce document vous permet d'introduire plusieurs instances en tant que paramètre et de cibler des instances se déplaçant vers le même type d'instance de destination (par exemple, toutes les instances passant à t3a.medium, quel que soit le type d'instance source).

Révision des coûts avant et après dans Cost Explorer

Une fois que vous avez correctement dimensionné vos ressources, vous pouvez utiliser Cost Explorer pour afficher les coûts avant et après en utilisant la balise Rightsizing. N'oubliez pas que

vous pouvez utiliser [des balises de ressources](#) pour suivre les coûts. En utilisant plusieurs couches de balises, vous pouvez obtenir une visibilité précise de vos coûts. Dans l'exemple présenté dans ce guide, la balise Rightsizing est utilisée pour appliquer une balise générique à toutes les instances ciblées. Ensuite, un tag d'équipe est utilisé pour mieux organiser les ressources. L'étape suivante consiste à introduire des balises d'application pour mieux montrer l'impact financier de l'exploitation d'une application spécifique.

Le schéma suivant montre la structure des balises d'une organisation.



Prenons l'exemple d'une entreprise qui dimensionne correctement les serveurs Web de production appartenant à l'équipe des opérations. Dans Cost Explorer, le tag Rightsizing est défini sur activé, et le tag Team est défini sur operations. Dans cet exemple, un effort de dimensionnement approprié réduit les coûts d'exploitation de 0,89 cent à 0,28 cent de l'heure. En supposant 744 heures par mois, le coût annuel avant de choisir la bonne taille est de 7 945,92\$. Après le bon dimensionnement, le coût annuel tombe à 2 499,84\$. Cela se traduit par une diminution de 68,5 % des coûts annuels liés à la charge de travail. Imaginez l'impact de cette situation sur une grande organisation. N'oubliez pas que cela se fait dans un environnement d'exemple et que les instances sont généralement inactives. Dans un environnement de production, vous pouvez réaliser des économies de 10 à 35 %.

À présent, considérez l'impact d'un dimensionnement correct de l'hôte du bastion de production appartenant à l'équipe d'ingénierie. Dans Cost Explorer, le tag Rightsizing est défini sur activé, et le tag Team est défini sur Engineering. Dans cet exemple, un effort de dimensionnement adapté permet de réduire les coûts de 0,75 centime à 0,44 centime de l'heure. En supposant 744 heures par mois, le coût annuel avant de choisir la bonne taille est de 6 696\$. Après le bon dimensionnement, le coût annuel tombe à 3 928,32\$.

Si vous utilisez plusieurs balises, vous pouvez filtrer les données pour obtenir des détails précis sur les coûts. Dans cet exemple, le tag Team réduit le bruit afin que vous puissiez visualiser l'impact au niveau de l'équipe. La balise Rightsizing étant activée, vous pouvez également filtrer les instances dont la valeur est activée ou aucune valeur n'est présente pour cette balise. Cela peut fournir une vue globale de vos efforts de dimensionnement appropriés, en particulier lorsqu'ils sont visualisés dans le compte de gestion (payeur) au niveau de Cost Explorer. Cette vue vous permet de voir tous les comptes et instances.

Prenons un exemple au niveau du compte unique où la balise Rightsizing est définie sur Activé.

Les coûts d'exploitation passent de 1,64 dollar de l'heure à 0,72 cents de l'heure. En supposant 744 heures par mois, le coût annuel avant de choisir la bonne taille est de 14 641,92\$. Après le bon dimensionnement, le coût annuel tombe à 6 428,16\$. Cela se traduit par une diminution de 56 % des coûts de calcul pour ce compte.

Avant de vous lancer dans l'aventure de la bonne taille, tenez compte des points suivants :

- AWS offre de nombreuses options de réduction des coûts. Cela inclut [AWS OLA](#), où AWS passe en revue vos instances sur site avant de passer à AWS. L' AWS OLA vous fournit également des recommandations de dimensionnement et des conseils de licence appropriés.
- Choisissez toutes les bonnes tailles avant de souscrire à [Savings Plans](#). Cela peut vous aider à éviter les surachats dans le cadre de votre engagement Savings Plans.

Recommandations

Nous recommandons les prochaines étapes suivantes :

1. Passez en revue votre environnement existant et envisagez de convertir les volumes Amazon EBS gp2 en volumes gp3.
2. Passez en revue [Savings Plans](#).

Ressources supplémentaires

- [Optimiseur de calcul AWS](#)(AWS documentation)
- [Bonnes pratiques pour le balisage des AWS ressources](#) (livres AWS blancs)
- [Comment collecter des données depuis Optimiseur de calcul AWS et AWS Trusted Advisor entre votre AWS Organizations](#) (YouTube)
- [Optimisation des performances et réduction des coûts de licence : tirer parti Optimiseur de calcul AWS des instances Amazon EC2 SQL Server](#) (Microsoft Workloads sur le blog) AWS

Sélectionnez le type d'instance adapté aux charges de travail Windows

Présentation de

La pratique du surprovisionnement constitue une différence significative entre les charges de travail fonctionnant dans le cloud et celles exécutées dans des environnements sur site. Lorsque vous achetez du matériel physique pour une utilisation sur site, vous effectuez une dépense d'investissement qui devrait durer pendant une durée prédéterminée, généralement de 3 à 5 ans. Pour faire face à la croissance prévue pendant la durée de vie du matériel, celui-ci est acquis avec plus de ressources que ce dont votre charge de travail a actuellement besoin. Par conséquent, le matériel physique est souvent surapprovisionné bien au-delà des besoins de votre charge de travail réelle.

La technologie des machines virtuelles (VM) est apparue comme un moyen efficace d'utiliser les ressources matérielles excédentaires. Les administrateurs ont VMs surapprovisionné en v CPUs et en RAM, ce qui a permis à l'hyperviseur de gérer l'utilisation des ressources physiques entre les serveurs occupés et inactifs en allouant les ressources inutilisées à chaque machine virtuelle. Lors de la gestion VMs, les ressources vCPU et RAM allouées à chaque machine virtuelle fonctionnaient davantage comme des gouverneurs de ressources que comme des indicateurs de l'utilisation réelle. La surallocation des ressources des machines virtuelles peut facilement dépasser le triple des ressources de calcul disponibles.

[Amazon Elastic Compute Cloud \(Amazon EC2\)](#) évite de VMs surprovisionner le matériel sous-jacent, car cela n'est pas nécessaire. Le cloud computing est une dépense opérationnelle, pas une dépense en capital, et vous ne payez que pour ce que vous utilisez. Si votre charge de travail nécessite

davantage de ressources à l'avenir, provisionnez-les lorsque vous en avez réellement besoin, plutôt que de le faire de manière préventive.

Il existe des centaines d'options pour choisir les bons types [d'instances Amazon EC2](#). Si vous envisagez de migrer une charge de travail Windows vers le cloud, AWS propose un [AWS OLA](#) pour vous aider à mieux comprendre votre charge de travail actuelle et fournir un exemple de ses performances sur AWS. L'analyse AWS OLA vise à faire correspondre le type et la taille d'instance EC2 appropriés à votre utilisation réelle sur site.

Si des charges de travail s'exécutent déjà sur Amazon EC2 et que vous recherchez des stratégies d'optimisation des coûts, cette section du guide vous aide à identifier les différences entre les instances Amazon EC2 et leur applicabilité aux charges de travail Windows classiques.

Recommandations en matière d'optimisation des coûts

Pour optimiser les coûts de vos types d'instances EC2, nous vous recommandons de procéder comme suit :

- Choisissez la famille d'instances adaptée à votre charge de travail
- Comprendre les variations de prix entre les architectures de processeurs
- Comprenez les différences entre le prix et les performances entre les générations EC2
- Migrer vers de nouvelles instances
- Utiliser des instances burstables

Choisissez la famille d'instances adaptée à votre charge de travail

Il est important de choisir la famille d'instances adaptée à votre charge de travail.

Les instances Amazon EC2 sont réparties dans les différents groupes suivants :

- Usage général
- Calcul optimisé
- Mémoire optimisée
- Calcul accéléré
- Stockage optimisé
- Optimisé pour le HPC

La plupart des charges de travail Windows entrent dans les catégories suivantes :

- Usage général
- Calcul optimisé
- Mémoire optimisée


Pour simplifier encore davantage les choses, considérez une instance EC2 de base dans chaque catégorie :

- Optimisé pour le calcul — C6i
- Usage général — M6i
- Mémoire optimisée — R6i

La génération précédente d'instances EC2 présentait des différences mineures dans les types de processeurs. Par exemple, les instances optimisées pour le calcul C5 disposent de processeurs plus rapides que les instances M5 à usage général ou les instances optimisées pour la mémoire R5. La dernière génération d'instances EC2 (C6i, M6i, R6i, C6a, M6a et R6a) utilise toutes le même processeur pour toutes les familles d'instances. Comme le processeur est le même pour les instances de dernière génération, la différence de prix entre les familles d'instances dépend désormais davantage de la quantité de RAM. Plus une instance possède de RAM, plus elle coûte cher.

L'exemple suivant illustre le tarif horaire pour une instance à 4 vCPU Intel exécutée dans us-east-1 la région.

Instance	v CPUs	RAM	Tarif horaire
c6i.xlarge	4	8	0,17\$
m6i.xlarge	4	16	0,19\$
r6i.xlarge	4	32	0,25\$

 Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Instances éclatables

Bien qu'il soit recommandé en matière de cloud computing de désactiver les ressources informatiques inutilisées pour éviter des frais, toutes les charges de travail ne peuvent pas être désactivées et activées chaque fois qu'elles sont nécessaires. Certaines charges de travail restent inactives pendant de longues périodes mais doivent être accessibles 24 heures sur 24.

Les instances Burstable (T3) permettent de maintenir en ligne des charges de travail exigeantes ou peu utilisées tout au long de la journée, tout en réduisant les coûts de calcul. Les instances Burstable EC2 disposent d'une quantité maximale de ressources vCPU qu'elles peuvent utiliser pendant de brèves périodes. Ces instances utilisent un système basé sur des [crédits CPU extensibles](#). Ces crédits sont accumulés pendant les périodes d'inactivité de la journée. Les instances Burstable offrent des vCPU-to-RAM ratios variables, ce qui en fait des alternatives aux instances optimisées pour le calcul dans certains cas et à d'autres instances à usage général dans d'autres.

L'exemple suivant illustre le tarif horaire pour une instance T3 (c'est-à-dire une instance burstable) exécutée dans la us-east-1 région.

Instance	v CPUs	RAM (GO)	Tarif horaire
t3.nano	2	0.5	0,0052\$
t3.micro	2	1	0,0104\$
t3.small	2	2	0,0208\$
t3.medium	2	4	0,0416\$
t3.large	2	8	0,0832\$
t3.xlarge	4	16	0,1664\$
t3.2xlarge	8	32	0,3328\$

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Comprendre les variations de prix entre les architectures de processeurs

Les processeurs [Intel](#) sont la norme pour les instances EC2 depuis leur création. Les générations précédentes d'instances EC2, telles que C5, M5 et R5, n'indiquaient pas Intel comme architecture de processeur (car c'était l'architecture par défaut). Les nouvelles générations d'instances EC2, telles que C6i, M6i et R6i, incluent un « i » pour indiquer l'utilisation d'un processeur Intel.

La modification de l'annotation de l'architecture du processeur est due à l'introduction d'options de processeur supplémentaires. Le processeur le plus comparable à Intel est [AMD](#) (désigné par un « a »). Les processeurs AMD EPYC utilisent la même architecture x86 et offrent des performances similaires à celles des processeurs Intel, mais à un prix inférieur. Comme le montrent les exemples de tarification suivants, les instances AMD EC2 offrent une réduction d'environ 10 % sur les coûts de calcul par rapport à leurs homologues Intel.

instance Intel	Tarif horaire	Instance AMD	Price	% de différence
c6i.xlarge	0,17\$	c6a.xlarge	0,153\$	10 %
m6i.xlarge	0,192\$	m6a.xlarge	0,1728\$	10 %
r6i.xlarge	0,252\$	r6a.xlarge	0,2268\$	10 %

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.


La troisième option d'architecture de processeur majeure est [AWS celle des processeurs Graviton](#) (désignés par un « g ») sur les instances EC2. Conçus par AWS, les processeurs Graviton offrent le meilleur rapport prix/performance sur Amazon EC2. Non seulement les processeurs Graviton actuels sont 20 % moins chers que leurs homologues Intel, mais ils offrent également une amélioration des performances de 20 % ou plus. La prochaine génération de processeurs Graviton devrait encore accroître cette différence de performances, les tests montrant une augmentation supplémentaire de 25 % des performances.

Windows Server ne peut pas fonctionner sur les processeurs Graviton, qui sont basés sur l'architecture ARM. En fait, Windows Server fonctionne uniquement sur des processeurs x86. Bien

que vous ne puissiez pas augmenter le rapport prix/performance de 40 % en utilisant des instances basées sur Graviton pour Windows Server, vous pouvez toujours utiliser des processeurs Graviton pour des charges de travail Microsoft spécifiques. Par exemple, [les nouvelles versions de .NET peuvent être exécutées sous Linux](#). Cela signifie que ces charges de travail peuvent utiliser des processeurs ARM et bénéficier d'instances Graviton EC2 plus rapides et plus abordables.

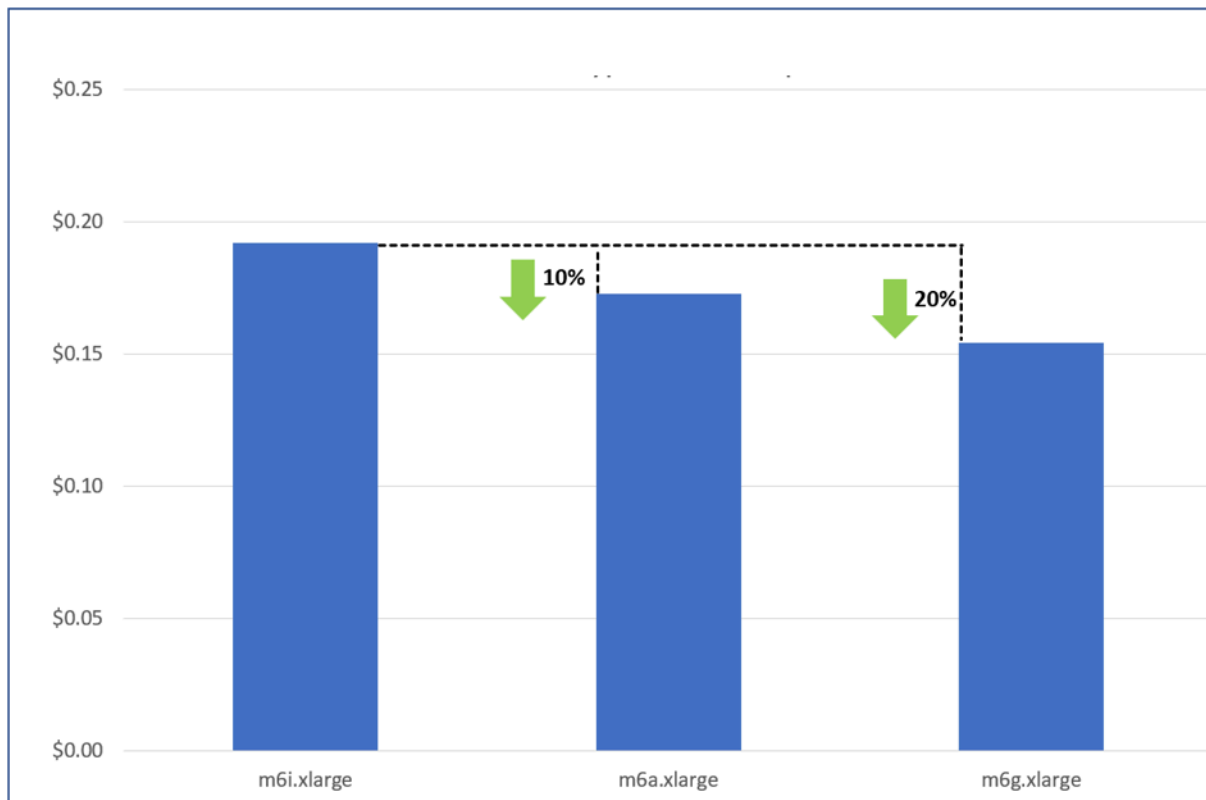
L'exemple suivant illustre la tarification horaire pour une instance Graviton exécutée dans la us-east-1 région.

instance Intel	Tarif horaire	Instance de Graviton	Tarif horaire	% de différence
c6i.xlarge	0,17\$	c6g.xlarge	0,136\$	20 %
m6i.xlarge	0,192\$	m6g.xlarge	0,154\$	20 %
r6i.xlarge	0,252\$	r6g.xlarge	0,2016\$	20 %

 Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Le graphique suivant compare les prix des instances de la série M.



Comprenez les différences de prix et de performances entre les générations EC2

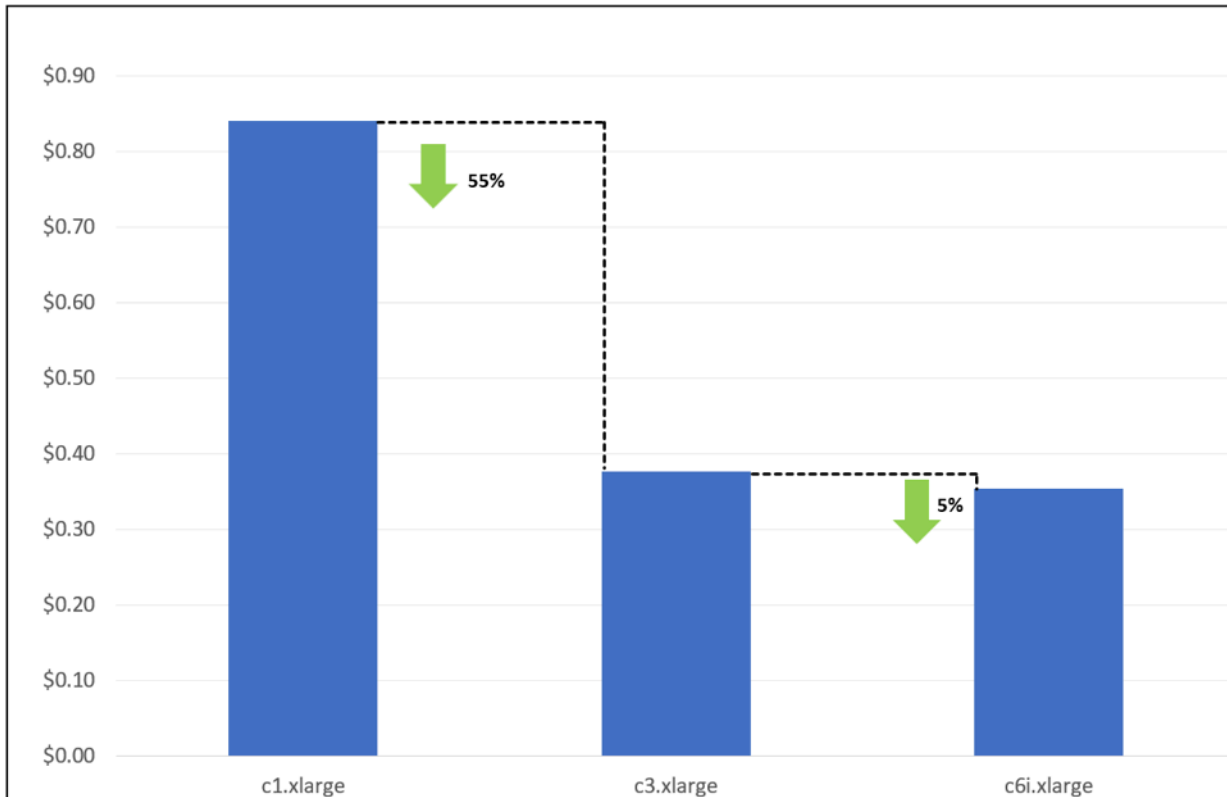
L'une des caractéristiques les plus constantes d'Amazon EC2 est que chaque nouvelle génération offre un meilleur rapport prix/performance que la précédente. Comme le montre le tableau suivant, le prix des instances EC2 de nouvelle génération diminue à chaque nouvelle version.

Instance optimisée pour le calcul	Tarif horaire	Instance à usage général	Tarif horaire	Instance optimisée pour la mémoire	Tarif horaire
C1.xLarge	0,52\$	M1.xLarge	0,35\$	r1.xlarge	N/A
C3.xLarge	0,21\$	M3.xLarge	0,266\$	r3.xlarge	0,333\$
C5.xLarge	0,17\$	M5.xLarge	0,192\$	r5.xlarge	0,252\$

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Le graphique suivant compare les coûts des différentes générations d'instances de la série C.



Toutefois, le prix des instances de 6e génération est identique à celui de la 5e génération, comme le montre le tableau suivant.

Instance optimisée pour le calcul	Tarif horaire	Instance à usage général	Tarif horaire	Instance optimisée pour la mémoire	Tarif horaire
C5.xLarge	0,17\$	M5.xLarge	0,192\$	r5.xlarge	0,252\$
C6i.xLarge	0,17\$	M6 x Large	0,192\$	r6i.xlarge	0,252\$

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Malgré le même coût, la nouvelle génération offre un rapport prix/performance supérieur grâce à des processeurs plus rapides, à un débit réseau amélioré et à une augmentation du débit et des IOPS d'Amazon Elastic Block Store (Amazon EBS).

L'une des améliorations les plus significatives en termes de rapport prix/performance est l'amélioration de l'instance [X2i](#). Cette génération de l'instance offre un rapport prix/performance jusqu'à 55 % supérieur à celui de la génération précédente. Comme le montre le tableau suivant, le x2iedn montre une amélioration de tous les aspects de performance (le tout au même prix que la génération précédente).

Instance	Tarif horaire	v CPUs	RAM	Vitesse du processeur	Stockage d'instances	Réseaux	Débit Amazon EBS	IOPS EBS
x1e.2xlarge	1,66\$	8	244	2.3 GHz	SSD DE 237 GO	10 Gbit/s	125 Mo/s	7400
x1iedn.2xlarge	1,66\$	8	256	3,5 GHz	SSD de 240 Go NVMe	25 Gb/s	2500 Mo/s	65000

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Exemples de scénarios

Prenons l'exemple d'une société d'analyse qui suit les véhicules de livraison et souhaite améliorer les performances de SQL Server. Une fois qu'une PME MACO a examiné les problèmes de performance

de cette entreprise, celle-ci passe des instances x1e.2xlarge aux instances x2iedn.xlarge. La nouvelle taille d'instance est plus petite, mais les améliorations apportées aux instances x2 permettent d'améliorer les performances et l'optimisation de SQL Server grâce à l'utilisation d'extensions de pool de mémoire tampon. Cela permet à l'entreprise de passer de l'édition SQL Server Enterprise à l'édition SQL Server Standard. Cela permet également à l'entreprise de réduire ses licences SQL Server de 8 V CPUs à 4 CPUs V.

Avant l'optimisation :

Serveur	instance EC2	Edition SQL Server	Coût mensuel
Prod DB1	x1e.2xlarge	Enterprise	3 918,64\$
Prod DB2	x1e.2xlarge	Enterprise	3 918,64\$
Total			7 837,28\$

Après optimisation :

Serveur	instance EC2	Edition SQL Server	Coût mensuel
Prod DB1	x2iedn.xlarge	Standard	1 215,00\$
Prod DB2	x2iedn.xlarge	Standard	1 215,00\$
Total			2 430,00\$

Dans l'ensemble, le passage des instances x1e.2xlarge aux instances x2iedn.xlarge permet à l'entreprise, dans le scénario d'exemple, d'économiser 5 407\$ par mois sur ses serveurs de base de données de production. Cela réduit le coût total de la charge de travail de 69 %.

Note

La tarification est basée sur la tarification horaire à la demande dans la us-east-1 région.

Migrer vers de nouvelles instances

[Les anciennes générations d'Amazon EC2 s'exécutent sur l'hyperviseur Xen, tandis que les nouvelles générations fonctionnent sur le système Nitro.AWS](#) Le système Nitro fournit la quasi-totalité des ressources de calcul et de mémoire du matériel hôte à vos instances. Cela se traduit par une amélioration des performances globales. Des considérations particulières doivent être prises en compte lors [de la migration d'instances basées sur Xen vers Nitro](#). Par exemple, [AWS Windows](#) est configuré avec AMIs les paramètres et personnalisations par défaut utilisés par le support d'installation Microsoft. Les personnalisations incluent des pilotes et des configurations compatibles avec les types d'instances de dernière génération ([instances basées sur le système Nitro](#)).

Si vous lancez des instances à partir de Windows personnalisé AMIs ou de Windows AMIs fournis par Amazon qui ont été créés avant août 2018, nous vous recommandons de suivre les étapes décrites dans la section [Migration vers les types d'instances de dernière génération](#) dans la documentation Amazon EC2.

Utiliser des instances burstables

Bien que les instances burstables soient un bon moyen de réduire les coûts de calcul, nous vous recommandons de les éviter dans les scénarios suivants :

- Les [spécifications minimales pour Windows Server](#) with the Desktop Experience nécessitent 2 Go de RAM. Évitez d'utiliser des instances t3.micro ou t3.nano avec Windows Server car elles ne disposent pas d'un minimum de RAM.
- Si votre charge de travail est chargée mais qu'elle ne reste pas inactive assez longtemps pour générer des crédits en rafale, il est plus efficace d'utiliser des instances EC2 normales que des instances de rafale. Nous vous recommandons [de surveiller les crédits de votre processeur](#) pour vérifier cela.
- Nous vous recommandons d'éviter d'utiliser des instances burstables avec SQL Server dans la plupart des scénarios. La licence pour SQL Server est basée sur le nombre de v CPUs attribués à une instance. Si SQL Server est inactif la majeure partie de la journée, vous devrez payer pour des licences SQL que vous n'utilisez pas pleinement. Dans ces scénarios, nous vous recommandons de consolider plusieurs instances de SQL Server sur un serveur plus grand.

Étapes suivantes

Nous vous recommandons de suivre les étapes suivantes afin d'optimiser vos coûts pour les instances Windows Amazon EC2 :

- Utilisez l'instance EC2 de dernière génération pour bénéficier du meilleur rapport qualité/prix.
- Utilisez des instances EC2 dotées de processeurs AMD pour réduire de dix pour cent les coûts de calcul.
- Optimisez l'utilisation des ressources en choisissant un type d'instance EC2 adapté à votre charge de travail.

Le tableau suivant présente des exemples de points de départ typiques pour les charges de travail Windows. Des options supplémentaires sont disponibles, telles que des volumes de stockage d'instance pour améliorer les charges de travail SQL Server ou des instances EC2 avec des ratios beaucoup plus vCPU-to-RAM élevés. Nous vous recommandons de tester vos charges de travail de manière approfondie et d'utiliser des outils de surveillance Optimiseur de calcul AWS pour vous aider à effectuer les ajustements nécessaires.

Charge de travail	Typique	Facultatif
Active Directory	T3, M6i	R6i
Serveurs de fichiers	T3, M6i	C6i
Serveurs Web	T3, C6i	M6i, R6i
SQL Server	R6i	x2iedn, x2ieZn

Si vous devez modifier le type d'instance EC2, le processus implique généralement un simple redémarrage du serveur. Pour plus d'informations, consultez [Modifier le type d'instance](#) dans la documentation Amazon EC2.

Avant de modifier le type d'instance, nous vous recommandons de prendre en compte les points suivants :

- Vous devez arrêter vos instances soutenues par Amazon EBS avant de pouvoir modifier leur type d'instance. Assurez-vous de prévoir les interruptions de service pendant l'arrêt de votre instance. L'arrêt d'une instance et la modification de son type peuvent prendre quelques minutes, et la durée du redémarrage de votre instance peut varier en fonction des scripts de démarrage de votre application. Pour plus d'informations, consultez la section [Arrêter et démarrer votre instance](#) dans la documentation Amazon EC2.

- Lorsque vous arrêtez et démarrez une instance, AWS déplacez-la vers un nouveau matériel. Si votre instance possède une IPv4 adresse publique, AWS publie l'adresse et attribue une nouvelle IPv4 adresse publique à votre instance. Si vous avez besoin d'une IPv4 adresse publique qui ne change pas, utilisez une [adresse IP élastique](#).
- Vous ne pouvez pas modifier le type d'instance si [l'hibernation](#) est activée sur l'instance.
- Vous ne pouvez pas modifier le type d'instance d'une [instance Spot](#).
- Si votre instance fait partie d'un groupe Auto Scaling, Amazon EC2 Auto Scaling indique que l'instance arrêtée est défectueuse et peut la mettre hors service et lancer une instance de remplacement. Pour empêcher que cela ne se produise, vous pouvez suspendre les processus de mise à l'échelle pour le groupe pendant que vous modifiez le type d'instance. Pour plus d'informations, consultez [Suspendre et reprendre un processus pour un groupe Auto Scaling](#) dans la documentation Amazon EC2 Auto Scaling.
- Lorsque vous modifiez le type d'instance d'une instance avec des volumes de stockage d' NVMe instance, l'instance mise à jour peut avoir des volumes de stockage d'instance supplémentaires, car tous les volumes de stockage d' NVMe instance sont disponibles même s'ils ne sont pas spécifiés dans l'Amazon Machine Image (AMI) ou dans le mappage des périphériques par blocs d'instance. Autrement, l'instance mise à jour a le même nombre de volumes de stockage d'instances que celui spécifié lors du lancement de l'instance initiale.

Ressources supplémentaires

- [Types d'instances Amazon EC2 \(documentation\)](#) AWS
- [AWS Optimisation et évaluation des licences](#) (AWS documentation)

Apportez des licences pour les charges de travail Windows et SQL Server

Présentation de

Si vous avez investi des sommes importantes dans les charges de travail Microsoft et dans les contrats de licence d'entreprise existants, vous pouvez choisir parmi plusieurs AWS options pour prendre en charge ces charges de travail, notamment les options [licence incluse \(fournie par AWS\)](#) et [Bring Your Own License \(BYOL\)](#). Vous pouvez utiliser les [hôtes dédiés Amazon EC2](#) pour tirer pleinement parti des contrats de licence Microsoft existants et intégrer Windows Server à AWS. Cela

peut vous faire économiser jusqu'à 50 % sur les coûts des instances Amazon EC2. Étant donné que les licences Windows représentent environ la moitié des coûts des instances, le transfert de Windows Server AWS sur des hôtes dédiés peut permettre de réaliser des économies substantielles. Windows Server ne pouvant pas être mis en [location par défaut \(partagée\)](#), les hôtes dédiés constituent le choix idéal si vous souhaitez utiliser vos licences existantes pour Windows Server sur AWS.

Les hôtes dédiés ne sont pas réservés aux instances BYOL de Windows Server. Ils vous offrent également la flexibilité nécessaire pour adapter vos licences locales aux charges de travail SQL Server existantes. Les hôtes dédiés exposent les cœurs physiques du serveur sous-jacent et vous permettent d'octroyer une licence à SQL Server au niveau du cœur physique. Cela n'est pas possible dans le cadre de la location par défaut (partagée) où les licences SQL Server sont basées sur le nombre de machines virtuelles CPUs allouées à l'instance. Cette fonctionnalité vous permet d'octroyer des licences pour les charges de travail SQL Server conformément à votre stratégie de licence sur site. Par conséquent, vous pouvez économiser jusqu'à 50 % sur les coûts de licence SQL Server par rapport à la location (partagée) par défaut, en plus des économies sur les coûts d'instance, en utilisant des licences Windows éligibles. Pour plus d'informations sur ce scénario, consultez la section [Comprendre les licences SQL Server](#) de ce guide.

Hôtes dédiés Amazon EC2

Un hôte dédié Amazon EC2 est essentiellement le même hôte EC2 que celui AWS utilisé pour exécuter ses offres de calcul EC2. La différence est que ces hôtes sont entièrement dédiés à un seul client et fournissent un accès exclusif à l'infrastructure physique sous-jacente. Vous pouvez utiliser des hôtes dédiés pour exécuter vos instances sur du matériel entièrement dédié à votre usage, au lieu de partager des ressources avec d'autres AWS clients. Cela vous permet de mieux contrôler les ressources du cloud et de réduire les coûts en transférant vos propres licences logicielles, telles que Windows Server et SQL Server, sur AWS.

Gardez à l'esprit les points suivants :

- Un hôte dédié est un serveur physique entièrement dédié à un seul client. Vous bénéficiez d'une visibilité sur les sockets et les cœurs physiques de l'hôte dédié, ce qui vous permet de répondre aux exigences de conformité en matière de licences, telles que les contrats de licence logicielle par socket, par cœur ou par machine virtuelle.
- Les hôtes dédiés qui peuvent prendre en charge plusieurs tailles d'instance d'une même famille d'instances sont appelés hôtes dédiés hétérogènes. Ces [familles d'instances](#) incluent T3, A1, C5, M5, R5, C5n, R5n et M5n. En revanche, les autres familles d'instances ne prennent en charge

qu'une seule taille d'instance sur le même hôte dédié. C'est ce qu'on appelle des hôtes dédiés homogènes.

- Les hôtes dédiés sont facturés par hôte. Cela signifie que vous êtes facturé par hôte dédié, quel que soit le nombre d'instances exécutées sur celui-ci. La tarification des hôtes dédiés varie en fonction de la famille d'instances, de la région et de l'option de paiement sélectionnée. Vous pouvez choisir la configuration optimale pour votre charge de travail afin d'obtenir les performances et les coûts souhaités.

Ce schéma illustre les différences entre les instances de location partagée et les hôtes dédiés.



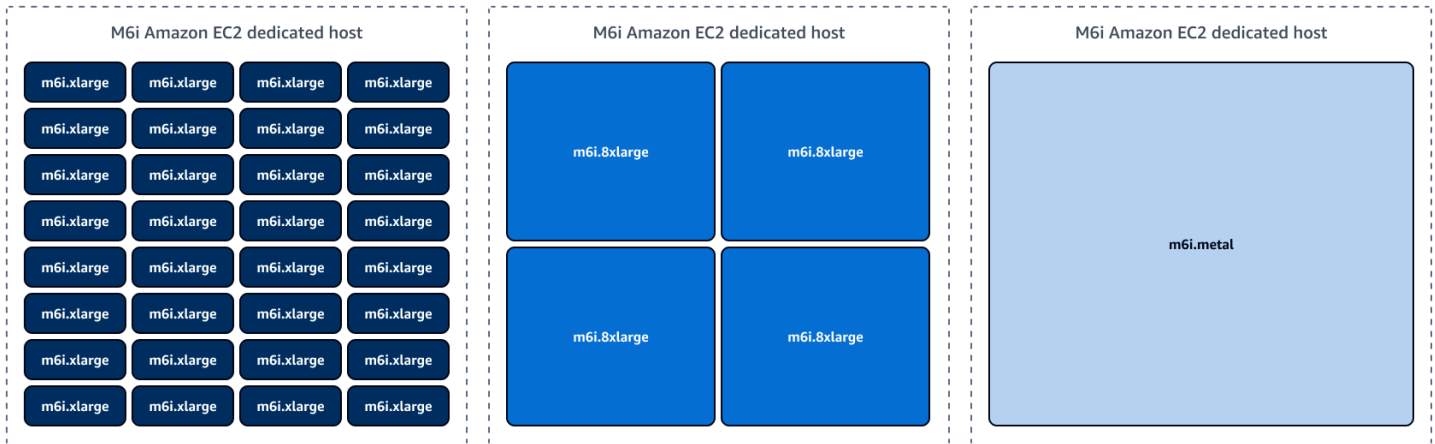
Hôtes dédiés homogènes

Imaginons un scénario dans lequel un hôte dédié M6i est utilisé. Les hôtes dédiés M6i et R6i possèdent deux sockets, 64 cœurs physiques et prennent en charge des types d'instances de même taille. C'est ce qu'on appelle des hôtes dédiés homogènes. Cela signifie que le nombre d'instances que vous pouvez lancer sur un seul hôte dédié M6i dépend de la taille de l'instance.

Par exemple :

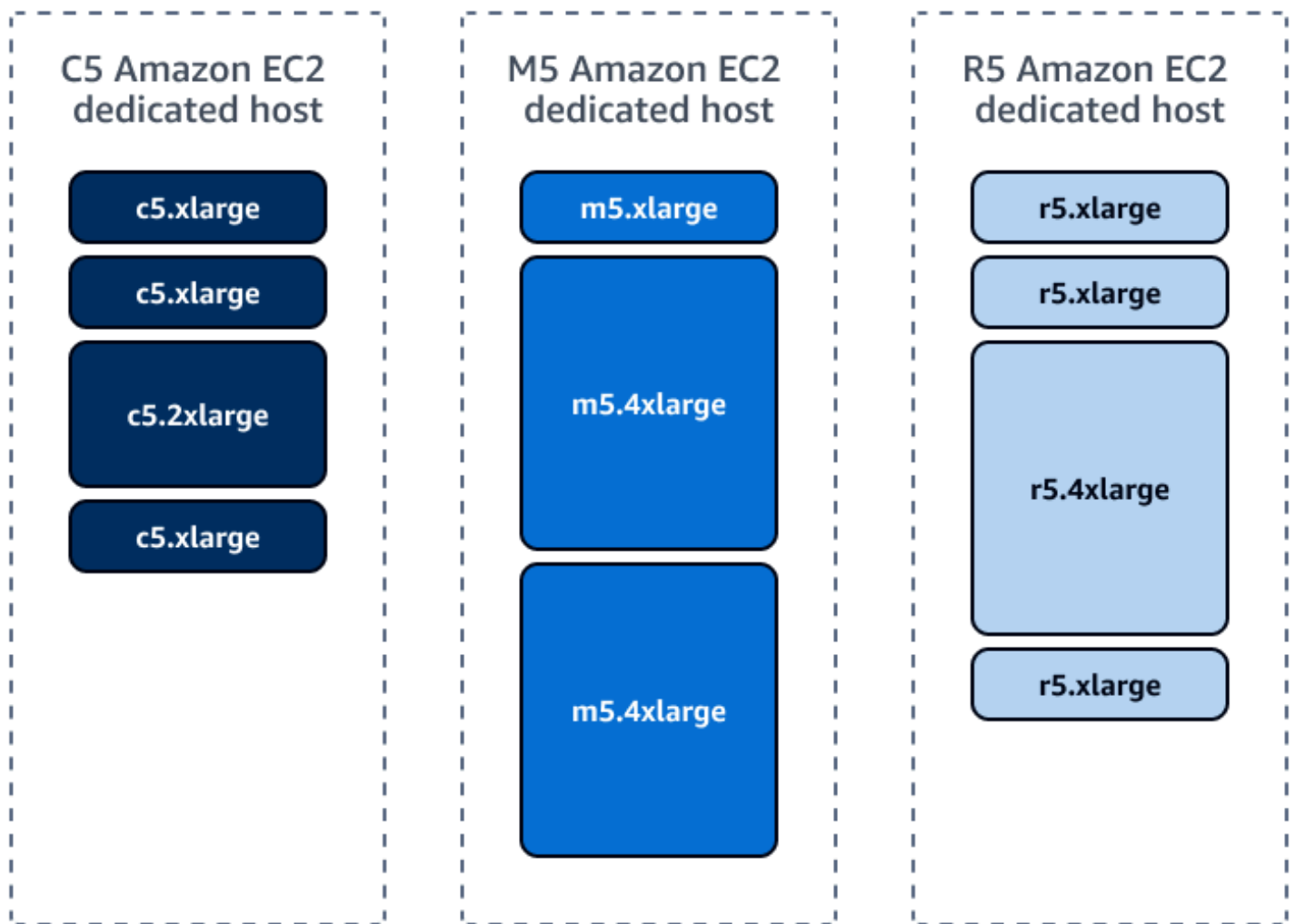
- Dans le cas de xlarge (4 vCPUs), vous pouvez lancer un maximum de 32 instances m6i.xlarge sur cet hôte dédié.
- Dans le cas de 8xlarge (32 vCPUs), vous pouvez lancer un maximum de 4 instances m6i.8xlarge sur cet hôte dédié.
- Dans le cas de metal (128 vCPUs), vous pouvez lancer au maximum 1 instance m6i.metal sur cet hôte dédié.

Le schéma suivant montre les options d'hôte dédié pour les instances M6.



Hôtes dédiés hétérogènes

Les hôtes dédiés qui prennent en charge plusieurs tailles d'instance sur le même hôte sont appelés hôtes dédiés Amazon EC2 hétérogènes. Le schéma suivant montre un exemple d'hôtes dédiés C5, M5 et R5 avec différentes tailles d'instance, telles que 2xlarge, xlarge et 4xlarge.



Gestion d'hôtes dédiés

Nous vous recommandons de prendre en compte les points suivants en ce qui concerne la gestion des hôtes dédiés Amazon EC2 :

- Pour tirer pleinement parti des hôtes dédiés, vous pouvez [partager un hôte unique entre plusieurs comptes au sein de votre organisation](#). Le partage d'hôtes permet d'optimiser les ressources et de réaliser des économies en utilisant tous les emplacements disponibles sur l'hôte. En partageant un hôte dédié entre les unités commerciales, vous pouvez centraliser votre infrastructure informatique et améliorer l'utilisation des ressources, tout en maintenant la séparation entre les charges de travail. Si vous faites partie d'une organisation AWS Organizations et que le partage est activé au sein de votre organisation, les clients de votre organisation ont automatiquement accès à l'hôte dédié partagé. Dans le cas contraire, les consommateurs reçoivent une invitation à rejoindre

le partage de ressources et bénéficie d'un accès à l'Hôte dédié partagé après avoir accepté l'invitation.

- Vous pouvez exécuter Windows Server 2022 sur des hôtes dédiés dans le cadre du modèle avec licence incluse, car Windows Server 2019 est la dernière version où vous pouvez utiliser le BYOL. Si vous souhaitez utiliser Windows Server 2022 sur des hôtes dédiés, vous devez utiliser des instances Windows Server 2022 incluses dans une licence.
- [AWS License Manager](#) est une solution complète pour gérer les licences logicielles de différents fournisseurs dans des environnements sur site AWS et sur site. Si vous [utilisez License Manager](#), vous pouvez bénéficier d'une visibilité et d'un contrôle accrus sur la manière dont les licences logicielles sont utilisées, ce qui vous permet de réaliser des économies et d'améliorer la conformité. Vous pouvez utiliser License Manager pour définir des règles afin d'émuler vos conditions de licence uniques. Cela vous permet de faire appliquer ces règles et d'empêcher toute utilisation abusive des licences. Cela peut réduire le risque de non-conformité et améliorer les processus de gestion des licences.
- Vous pouvez utiliser License Manager pour automatiser le placement, la libération et la restauration des hôtes à l'aide de [groupes de ressources d'hôtes](#). Cela peut augmenter la productivité et réduire les frais de gestion. License Manager fournit également une vue centralisée de l'utilisation des licences dans les environnements locaux AWS et dans les environnements sur site en fonction des règles de licence, ce qui facilite la gestion des achats de licences incrémentiels, de la conformité et des audits des fournisseurs au sein de votre organisation. En outre, License Manager s'intègre à AWS Organizations et AWS Resource Access Manager (AWS RAM) pour partager les configurations de licence entre les comptes et les régions. Cela vous permet de créer des rapports pour l'ensemble de votre environnement selon un calendrier et de gérer les règles de licence de manière centralisée en un seul outil Compte AWS. En fin de compte, cela peut améliorer la gouvernance et réduire la complexité.
- Lorsque vous concevez la haute disponibilité pour des hôtes dédiés au sein d'une même région, assurez-vous d'avoir alloué au moins deux hôtes dédiés dans au moins deux zones de disponibilité pour les charges de travail critiques pour la production. Pour plus d'informations, consultez les [hôtes dédiés Amazon EC2 pour Microsoft Windows sur le déploiement](#) de AWS référence.
- Pour chaque famille d'instances d'hôte dédié, le nombre d'instances que vous pouvez exécuter est limité pour chaque taille d'instance. Pour plus d'informations, consultez le [tableau de configuration des hôtes dédiés](#) dans la documentation Amazon EC2.

AWS options de licence

Les licences sont classées dans les catégories principales suivantes :

- Licence incluse — Cette option de licence vous permet d'acheter et d'utiliser des licences à la demande, en payant uniquement pour ce que vous utilisez. Il est idéal pour les cas d'utilisation où vous recherchez de la flexibilité dans l'utilisation de vos licences et souhaitez éviter les coûts initiaux. Vous pouvez choisir parmi une variété de produits Windows Server, SQL Server et d'autres produits Microsoft.
- Produits BYOL avec mobilité des licences : si vous possédez déjà des licences existantes et que vous souhaitez les utiliser dans le cloud, cette option de licence vous permet de transférer vos propres licences dans le cloud via le [programme Microsoft License Mobility](#). Les produits dotés d'une mobilité de licence, tels que SQL Server with Software Assurance (SA), peuvent être proposés en location partagée ou dédiée. Cela réduit les coûts d' AWS instance.
- Produits BYOL sans mobilité de licence — Pour les produits Microsoft tels que Windows Server dépourvus de mobilité de licence, AWS des options dédiées pour l'utilisation de ces produits dans le cloud sont proposées. En outre, les hôtes dédiés permettent d'octroyer des licences au niveau du cœur physique, ce qui permet d'économiser 50 % ou plus sur les licences nécessaires à l'exécution de vos charges de travail. Les hôtes dédiés constituent un excellent choix pour des charges de travail stables et prévisibles qui s'exécutent la plupart du temps.

Apporter des licences Windows Server

Apporter vos propres licences Windows est l'une des stratégies les plus efficaces pour optimiser les licences, car elle vous permet de tirer parti des investissements existants et de réduire vos AWS dépenses. Les scénarios BYOL spécifiques ne nécessitent pas les avantages de la SA ou de la mobilité des licences, mais une infrastructure dédiée Amazon EC2 est toujours nécessaire. Pour être éligible, vous devez avoir acheté des licences perpétuelles avant le 1er octobre 2019 ou les avoir ajoutées à titre indicatif dans le cadre d'un abonnement Enterprise actif en vigueur avant le 1er octobre 2019. Dans ces scénarios BYOL spécifiques, vous ne pouvez mettre à niveau les licences que vers les versions disponibles avant le 1er octobre 2019. Par exemple, si vous avez abandonné SA en 2017, vous avez le droit de déployer uniquement jusqu'à Windows Server 2016, et non jusqu'en 2019. Cependant, 2019 est la dernière version éligible au BYOL to AWS. Pour plus d'informations, consultez la section [Licences — Windows Server](#) dans la AWS documentation.

L'acquisition de licences peut avoir un impact significatif sur le coût d'exécution des charges de travail Microsoft. AWS Lorsque vous apportez vos propres licences, vous n'êtes pas obligé de payer des

frais de licence supplémentaires pour les instances exécutées dans le cloud, ce qui peut permettre de réaliser des économies considérables.

Le tableau suivant indique le coût mensuel à la demande lié à l'exécution d'une seule instance c5.xlarge 24 heures sur 24, 7 jours sur 7, sur différentes configurations.

Configuration	Coût mensuel (USD)
Édition Windows Server+ SQL Server Enterprise	1 353,00\$ (LI)
Windows Server + Édition standard de SQL Server	609,00\$ (LI)
Windows Server uniquement	259,00\$ (LI)
Calcul uniquement (Linux)	127,00\$

Vous pouvez utiliser vos licences existantes pour réduire les coûts de licence et économiser de l'argent sur votre AWS facture globale.

Pour être éligible au BYOL sur les hôtes dédiés Amazon EC2, vous devez apporter vos propres licences logicielles, par exemple pour Windows Server et SQL Server. Le BYOL vous permet d'utiliser vos licences existantes AWS et peut vous permettre de réaliser des économies. Pour apporter vos propres licences, vous devez disposer des droits de licence fournis par le fournisseur du logiciel et devez également fournir le support ou l'image d'installation du logiciel. Le support ou l'image d'installation peuvent être utilisés pour lancer des instances sur des hôtes dédiés. Pour en savoir plus sur la création d'une AMI BYOL, consultez [Comment créer un serveur Bring-Your-Own-License AMIs Windows sur site avec VM Import/Export](#) sur le blog Microsoft Workloads on. AWS

Note

Un type de licence défini sur Auto est l'équivalent d'une option [AWS incluse dans la licence](#). Cette option peut entraîner des dépenses à la demande non désirées. Vous devrez changer de [type de licence](#).

Scénarios d'optimisation des coûts

Le bon dimensionnement et l'optimisation des licences sont des éléments clés de l'optimisation des AWS coûts. Si vous mettez en œuvre les bonnes stratégies, vous pouvez réduire les coûts de licence, maintenir la conformité et tirer le meilleur parti possible de votre investissement dans les licences en utilisant les hôtes dédiés Amazon EC2 et l'option BYOL.

Cette section couvre les exemples de scénarios suivants :

- Réductions de coûts grâce aux hôtes dédiés T3
- Comparaison de la location partagée avec des hôtes dédiés avec SQL Server BYOL
- Déploiements SQL Server à haute disponibilité

Réductions de coûts grâce aux hôtes dédiés T3

Les hôtes dédiés T3 sont différents des autres hôtes dédiés Amazon EC2 qui fournissent traditionnellement des ressources CPU fixes. Les hôtes dédiés T3, en revanche, prennent en charge les instances évolutives capables de partager les ressources du processeur, de fournir des performances de base du processeur et de fonctionner en rafale en cas de besoin. Le partage des ressources du processeur, également connu sous le nom de surabonnement, permet à un seul hôte dédié T3 de prendre en charge jusqu'à quatre fois plus d'instances que des hôtes dédiés à usage général comparables.

Les hôtes dédiés T3 réduisent le coût total de possession en offrant une densité d'instances supérieure à celle de tout autre hôte dédié Amazon EC2. Les instances Burstable T3 vous permettent de consolider un plus grand nombre d'instances avec une utilisation low-to-moderate moyenne du processeur sur un nombre d'hôtes plus restreint que jamais. Les hôtes dédiés T3 proposent également des tailles d'instance plus petites dans un plus grand nombre de combinaisons de vCPU et de mémoire que les autres hôtes dédiés Amazon EC2. Des instances de plus petite taille peuvent contribuer à réduire le coût total de possession et à fournir des ratios de consolidation équivalents ou supérieurs à ceux des hôtes sur site.

Les hôtes dédiés T3 sont parfaitement adaptés à l'exécution de logiciels BYOL avec utilisation low-to-moderate du processeur et licences logicielles éligibles par socket, par cœur ou par machine virtuelle, y compris les bases de données Microsoft Windows Desktop, Windows Server, SQL Server et Oracle.

Utilisez des hôtes dédiés T3 pour réduire le nombre de licences Windows Server Datacenter (par cœur)

Dans les environnements sur site, vous profitez du fait que vous pouvez facilement surabonner votre abonnement physique CPUs sur les VMware hôtes et atteindre des niveaux de consolidation élevés.

Prenez l'exemple de code suivant. Vous utilisez actuellement des VMware hôtes 10 x 36 cœurs et 384 Go de RAM dans un environnement sur site. En outre, chaque hôte exécute 96 x 2 vCPU, 4 Go de RAM, des machines virtuelles Windows Server avec une faible utilisation moyenne du processeur.

Vous pouvez désormais atteindre des niveaux de consolidation bien supérieurs en déplaçant vos machines virtuelles vers des hôtes dédiés T3, qui disposent de deux fois plus de RAM que vos hôtes locaux actuels VMware . Vous pouvez exécuter le même nombre de serveurs sur des hôtes dédiés T3 avec un coût d'hôte réduit de 50 %. Cela peut vous aider à réduire les coûts de licence Windows Server de 33 %. Le tableau suivant met en évidence les économies réalisées grâce à l'utilisation d'hôtes dédiés T3.

	Hôtes locaux VMware	Hôtes dédiés T3	Économies
Serveurs physiques	10	5	
Nombre de cœurs physiques par hôte	36	48	
RAM par hôte (Go)	384	768	
2 vCPU, 4 Go de RAM par hôte VMs	96	192	
Nombre total de VMs	960	960	
Nombre total de licences Windows Server Datacenter (par cœur) = (Nombre de serveurs* Nombre de cœurs physiques)	10* 36 = 360	5* 48 = 240	33 %

Comparaison de la location partagée avec des hôtes dédiés avec SQL Server BYOL

Prenons un exemple pratique pour démontrer la valeur des hôtes dédiés Amazon EC2. Dans ce scénario, une organisation exécute une charge de travail SQL Server dans un environnement local de 240 cœurs et souhaite y déployer la même charge de travail de manière rentable. AWS Si cette organisation apporte ses propres licences (BYOL), elle continue à payer pour l'assistance sociale et la réduction du nombre de cœurs a une incidence directe sur ses coûts.

Le schéma suivant compare les AWS économies réalisées entre les droits Microsoft et SQL Server.

Microsoft entitlements (Enterprise Agreements)		SQL Server savings with AWS	
	Number of cores	AWS shared vCPUs	AWS BYOL/Dedicated Hosts cores
SQL Server Enterprise edition	208	120	96
SQL Server Standard edition	32	20	-
Total SA cost	\$341,000	\$197,418	\$151,355

En dimensionnant correctement les instances en location AWS partagée, vous pouvez réduire les licences SQL Server à 140 cœurs. Cela se traduit par des coûts d'assistance sociale de 197 000\$.

Les hôtes dédiés Amazon EC2 vous permettent d'octroyer une licence à SQL Server au niveau du cœur physique. Cela n'est pas possible dans le cadre d'une location partagée où les licences SQL Server sont basées sur le nombre de v CPUs alloués à l'instance. Par conséquent, en utilisant deux hôtes dédiés R5 de 48 cœurs chacun, vous n'avez besoin de couvrir que 96 cœurs au lieu des 140 V CPUs requis en location partagée. En déployant des hôtes dédiés R5 et en octroyant des licences pour les charges de travail au niveau physique, vous pouvez ramener le nombre requis de licences de l'édition Enterprise de SQL Server à 96 cœurs. Cela signifie que vous pouvez déployer jusqu'à 192 cœurs (compte tenu de l'hyperthreading) de charges de travail SQL Server, tout en respectant les exigences en matière de licences et en réalisant des économies de coûts importantes.

Dans ce cas, l'organisation paie environ 341 000\$ par an en frais d'assistance sociale. Après avoir correctement dimensionné la location partagée, ils réduisent les coûts à 197 000\$ avec 140 vCPU. Les hôtes dédiés Amazon EC2 réduisent encore les coûts à 151 000 dollars (soit une baisse d'environ 56 %).

Déploiements SQL Server à haute disponibilité

Cet exemple analyse la façon dont le coût peut influencer le déploiement de SQL Server en tenant compte de diverses considérations relatives aux licences. AWS Supposons qu'une entreprise doive déployer six serveurs SQL Server Enterprise AWS pour prendre en charge trois applications. Ces

serveurs nécessitent une haute disponibilité et disposent chacun de 16 V CPUs et de 256 Go de RAM. Consultez les détails du scénario suivants :

- Serveur — SQL Server
- Édition du système d'exploitation — Windows Server Datacenter 2019
- Édition SQL Server — SQL Server Enterprise 2019
- vCPU — 16
- Mémoire (Go) — 256
- Quantité — 6

Pour optimiser les coûts AWS sans sacrifier les performances, nous vous recommandons de dimensionner correctement les instances en fonction de l'utilisation du processeur, de la mémoire, du réseau et du disque (IOPS/BW). Après avoir correctement dimensionné les charges de travail, placez-les sur le type d'instance x2iedn.4xlarge, qui offre 16 v. CPUs Toutefois, ce type d'instance inclut également deux fois plus de mémoire nécessaire pour les charges de travail. D'autres optimisations sont toujours possibles.

Scénario 1

Une organisation déploie six serveurs SQL Server Enterprise en location AWS partagée en utilisant l'option incluse sous licence pour Windows et SQL Server. Avec cette option, le coût des licences Windows et SQL Server est intégré au prix de l'instance. Consultez les détails du scénario suivants :

- Location partagée (instance) — x2iedn.4xlarge
- Coût horaire (USD) — 10,0705 \$
- Coût mensuel par unité (USD) — 7 351,47 \$
- Nombre de serveurs : 6
- Processeur — 16
- Mémoire — 512
- Coût mensuel pour 6 serveurs : 44 108\$

Scénario 2

Une organisation dispose de SA et de BYOL pour SQL Server en location partagée. Cela signifie que l'organisation utilise l'option d'inclusion de licence pour Windows, mais fournit ses propres licences

SQL Server en fonction du nombre de v CPUs alloués à l'instance. Étant donné que l'organisation dispose de six serveurs SQL Server Enterprise de 16 V CPUs chacun, un total de 96 V CPUs est requis. Consultez les détails du scénario suivants :

- Location partagée (instance) — x2iedn.4xlarge
- Coût horaire (USD) — 4,0705 \$
- Coût mensuel par unité (USD) — 2971,47 \$
- Nombre de serveurs : 6
- Processeur — 16
- Mémoire — 512
- Noyaux BYOL — 96
- Coût mensuel pour 6 serveurs : 17 828\$

En apportant ses propres licences SQL Server avec SA, l'organisation dans ce scénario peut réaliser des économies par rapport à l'utilisation de l'option avec licence incluse pour SQL Server. Les économies de coûts précises dépendent du prix et des conditions du contrat de licence spécifique. Dans ce scénario, AWS les coûts diminuent de 26 280\$ par mois lorsque des licences SQL Server Enterprise sont transférées à AWS.

Scénario 3

Une organisation dispose d'un BYOL pour Windows et SQL Server sur des hôtes dédiés Amazon EC2. Cela signifie que l'organisation attribuera des licences au niveau du cœur physique, ce qui lui permettra de ne licencier que les cœurs physiques de l'hôte. Les licences au niveau du cœur physique vous permettent de déployer le nombre maximum d'instances sans affecter les licences requises. Ce modèle de licence est couramment utilisé avec Windows Server Datacenter et SQL Server Enterprise Edition.

Ce scénario utilise deux hôtes dédiés Amazon EC2 X2ieZN. Chaque hôte possède 24 cœurs physiques et 48 CPUs v. Cela fournit une capacité adéquate pour les six serveurs SQL Server Enterprise avec 16 V CPUs et 256 Go de RAM chacun. Consultez les détails du scénario suivants :

- Nombre d'hôtes dédiés : 2
- Famille d'instances — x2iezn
- Coût horaire (USD) — 11,009 \$

- Coût mensuel par unité (USD) — 8 036\$
- Noyau physique — 48
- vCPU disponible : 96
- Licences principales de Windows Server requises : 24
- Licences requises pour les cœurs SQL Server Enterprise : 24
- Coût mensuel — 16 073

Le coût total de deux hôtes dédiés Amazon EC2 de la famille X2ieZN est de 16 073\$ par mois. Pour plus d'informations sur la tarification, consultez l' [Calculateur de tarification AWS estimation](#) de ce scénario. Dans ce scénario, l'organisation peut économiser 1 755,65\$ par mois en apportant ses licences Windows. S'ils utilisent des hôtes dédiés Amazon EC2, ils peuvent également réduire le nombre de licences SQL Server requises. Dans le cas d'une location partagée, ils auraient besoin de 96 licences SQL Server Enterprise pour couvrir les six serveurs SQL Server Enterprise dotés chacun de 16 vCPU. Cependant, en utilisant des hôtes dédiés Amazon EC2 et en octroyant des licences au niveau du cœur physique, ils peuvent réduire le nombre de licences requises à 48 cœurs.

Les détails suivants comparent les coûts de l'exemple 3 et montrent les économies que vous pouvez réaliser en déployant des charges de travail sur des hôtes dédiés Amazon EC2 avec l'option BYOL par rapport à d'autres scénarios.

- Serveur sur site : SQL Server
- vCPU — 16
- Mémoire — 256
- Nombre de serveurs : 6
- Coût mensuel pour le scénario 1 : Windows (LI) + SQL Server Enterprise (LI) — 44 108\$
- Coût mensuel pour le scénario 2 : Windows (LI) + SQL Server Enterprise (BYOL) — 17 828\$
- Coût mensuel pour le scénario 3 : Windows (LI) + SQL Server Enterprise (BYOL) sur un hôte dédié Amazon EC2 — 16 073\$

Note

Le coût est basé sur la tarification à la demande. Vous pouvez encore réduire les coûts en utilisant des Savings Plans ou des instances réservées dédiées. Ces options offrent un modèle de tarification flexible qui permet de réaliser d'importantes économies par rapport

à la tarification à la demande. Avec ces plans, vous pouvez vous engager pour une durée d'un ou trois ans. Pour plus d'informations, consultez la section [Optimiser les dépenses pour Windows sur Amazon EC2](#) de ce guide.

Envisagez les options de paiement suivantes pour les hôtes dédiés Amazon EC2 :

- [Hôtes dédiés](#) (documentation Amazon EC2)
- [Réservations d'hôtes dédiés](#) (documentation Amazon EC2)
- [Savings Plans](#) (documentation Amazon EC2)

[Calculateur de tarification AWS](#) prend désormais en charge la tarification des hôtes dédiés. Cela peut vous aider à choisir l'hôte dédié sous-jacent approprié.

Recommandations en matière d'optimisation des coûts

Nous vous recommandons de suivre les étapes suivantes pour optimiser vos coûts en utilisant AWS Cost Explorer :

1. [Activez Cost Explorer](#).
2. Utilisez Cost Explorer pour [visualiser et analyser les coûts et l'utilisation](#) de vos déploiements d'hôtes dédiés Amazon EC2.
3. Vérifiez que vous utilisez BYOL. Vous pouvez afficher les détails de la plateforme et les valeurs des opérations d'utilisation suivants sur les instances ou les pages AMI de la console Amazon EC2, ou dans la réponse renvoyée par la commande `describe-images` `orddescribe-instances`.
 - Détails de la plateforme : Windows, opération d'utilisation : :0002 RunInstances (licence incluse)
 - Détails de la plateforme : Windows BYOL, Utilisation : 0800 RunInstances

Ressources supplémentaires

- [Types de licence éligibles pour la conversion des types de licence](#) (AWS License Manager documentation)
- [AWS License Manager et atelier d'accueil dédié](#) (AWS License Manager atelier)
- [Hôtes dédiés Amazon EC2 \(documentation FAQs AWS\)](#)

- [Comment créer un serveur Windows à Bring-Your-Own-License AMIs partir d'un serveur local avec VM Import/Export](#) (Microsoft Workloads sur le blog) AWS
- [Importation/exportation de machines virtuelles \(documentation\)](#) AWS
- [Amazon Web Services et Microsoft : questions fréquemment posées](#) (AWS documentation)
- [Conversions de types de licence dans License Manager](#) (AWS License Manager documentation)
- [Déploiement de SQL Server à haute disponibilité sur des hôtes dédiés Amazon EC2](#) AWS (blog sur les opérations et migrations dans le cloud)

Optimisez les dépenses pour Windows sur Amazon EC2

Présentation de

L'une des principales préoccupations liées à la migration des serveurs AWS concerne les coûts d'infrastructure. Il est vrai que l'un des avantages du cloud est de payer les ressources à la demande, mais certaines charges de travail de production doivent être disponibles 24 heures sur 24, 7 jours sur 7, 365 jours par an. Les [Savings Plans](#) sont conçus pour économiser de l'argent sur votre AWS utilisation permanente sur les instances EC2, et AWS Lambda. AWS Fargate

Les Savings Plans offrent un modèle de tarification flexible et peuvent vous aider à réduire le prix de l'utilisation d'Amazon EC2, Fargate, Lambda et Amazon SageMaker AI en échange d'un engagement à utiliser régulièrement (par exemple, 10 dollars de l'heure). Vous vous engagez à consacrer un montant horaire constant à vos dépenses informatiques sur un ou trois ans et, en échange, vous bénéficiez d'une réduction pour cette utilisation.

Avec Savings Plans, vous pouvez choisir entre trois options de paiement différentes :

- L'option No Upfront ne nécessite aucun paiement initial, et votre engagement est débité uniquement sur une base mensuelle.
- L'option Partial Upfront offre des prix plus bas sur les Savings Plans. Au moins la moitié de votre engagement vous est facturée à l'avance et le reste est facturé sur une base mensuelle.
- L'option All Upfront offre les prix les plus bas et l'intégralité de votre engagement est facturée en un seul paiement.

Vous pouvez suivre l'expiration de vos plans d'épargne et les prochains plans d'épargne en attente. AWS Cost Explorer Vous pouvez utiliser les alertes Savings Plans pour recevoir des alertes anticipées par e-mail 1, 7, 30 ou 60 jours avant la date d'expiration de votre plan, ou lorsqu'un

engagement est en attente d'achat. Ces notifications vous avertissent également de la date d'expiration. Vous pouvez envoyer des notifications à un maximum de 10 destinataires d'e-mails.

Comprendre les Savings Plans

Chaque type d'utilisation informatique est soumis à un tarif à la demande et à un tarif Savings Plans. Si vous vous engagez à utiliser 10 dollars par heure de calcul, vous bénéficiez des tarifs Savings Plans sur toutes les utilisations, jusqu'à 10 dollars au tarif des Savings Plans. Toute utilisation au-delà de l'engagement en matière de dépenses informatiques est facturée aux tarifs réguliers à la demande. Vous pouvez commencer à utiliser Savings Plans en utilisant Cost Explorer dans le AWS Management Console.

Vous pouvez facilement souscrire à Savings Plans en utilisant les recommandations fournies dans [Cost Explorer](#) pour réaliser les économies les plus importantes. L'engagement horaire recommandé est basé sur l'historique de votre utilisation à la demande et sur le type de forfait, la durée et l'option de paiement que vous avez choisis. Savings Plans est d'abord appliqué au compte qui a acheté le plan, puis il est partagé avec les autres comptes de la famille de facturation consolidée.

Note

L'option de partage de Savings Plans AWS Organizations est activée par défaut. Vous pouvez refuser cette option dans la AWS Billing console du compte payeur. Vous pouvez consulter votre page de [recommandations](#) pour voir les Savings Plans qui vous AWS recommandent de vous aider à économiser sur les utilisations éligibles. Ces recommandations peuvent être actualisées à tout moment pour vous permettre d'acheter facilement les meilleurs Savings Plans.

Calculer les plans d'économies

Les Compute Savings Plans offrent le maximum de flexibilité et contribuent à réduire vos coûts. Ces plans s'appliquent automatiquement à l'utilisation des instances EC2 indépendamment de la famille, de la taille, de la zone de disponibilité, de la région, du système d'exploitation ou de la location des instances. Ils s'appliquent également à l'utilisation de Fargate ou Lambda. Par exemple, avec Compute Savings Plans, vous pouvez passer d'une instance C4 à une instance M5, transférer une charge de travail de l'UE (Irlande) vers l'UE (Londres) ou déplacer une charge de travail d'EC2 vers Fargate ou Lambda à tout moment. Vous continuez automatiquement à payer le prix des Savings Plans.

EC2 Instance Savings Plans

Les plans EC2 Instance Savings Plans offrent les remises les plus importantes en échange d'un engagement à utiliser des familles d'instances individuelles dans une région (par exemple, en s'engageant à utiliser le M5 de manière constante en Virginie du Nord). Cela vous permet de bénéficier automatiquement de remises sur le prix à la demande de la famille d'instances sélectionnée dans cette région, quels que soient la zone de disponibilité, la taille, le système d'exploitation ou la location. Les plans EC2 Instance Savings Plans vous permettent de modifier votre utilisation entre les instances d'une même famille dans cette région. Par exemple, vous pouvez passer de la version c5.xlarge exécutant Windows à la version c5.2xlarge exécutant Linux et bénéficier automatiquement des prix des Savings Plans.

Les plans Compute et EC2 Instance Savings Plans s'appliquent aux instances EC2 qui font partie des clusters Amazon EMR, Amazon Elastic Kubernetes Service (Amazon EKS) et Amazon Elastic Container Service (Amazon ECS). Les frais Amazon EMR, Amazon EKS et Amazon ECS ne sont pas couverts par Savings Plans, mais les instances EC2 sous-jacentes le sont. Les Instance Savings Plans EC2 sont appliqués avant les Compute Savings Plans, car les Compute Savings Plans ont une applicabilité plus large.

Note

Vous ne pouvez pas modifier facilement un Savings Plans après avoir pris un engagement. Nous vous recommandons de bien planifier avant de vous engager dans l'une ou l'autre des options Savings Plans. Les Savings Plans proposent des prix inférieurs à ceux proposés à la demande en échange d'un engagement et ne peuvent pas être annulés pendant la durée.

Exemple d'engagement horaire

Si vous souscrivez un Savings Plans, vous vous engagez à payer une heure pour la durée du plan. Si vous vous engagez à utiliser 10 dollars par heure de calcul, la tarification de Savings Plans est automatiquement appliquée à toutes les utilisations, jusqu'à 10 dollars par heure. Toute utilisation au-delà de l'engagement est facturée aux tarifs réguliers sur demande. Vous pouvez utiliser l'outil de recommandation d'achat de Savings Plans dans Cost Explorer pour obtenir des recommandations d'engagements susceptibles de maximiser vos économies. L'engagement financier horaire pour un plan spécifique ne peut pas être modifié pour la durée du plan. Si vous souhaitez un engagement accru après avoir analysé l'utilisation, vous pouvez acheter un Savings Plans supplémentaire pour couvrir l'utilisation excédentaire.

Avantages des Savings Plans

Comparé aux instances réservées, Savings Plans propose un modèle de tarification plus flexible qui peut vous faire économiser de l'argent tout en profitant de la sélection plus large d'options de calcul proposées par Savings Plans. Les Savings Plans offrent des remises, même si vos besoins informatiques évoluent. Cela peut vous aider à suivre l'évolution constante de votre environnement dynamique sans encourir de frais de gestion supplémentaires. Voici d'autres avantages liés à l'utilisation de Savings Plans :

- Facile à utiliser : bénéficiez de remises automatiques en échange d'un engagement financier.
- Flexibilité — Un engagement unique qui s'applique à plusieurs types d'utilisation.
- Économies potentielles — Il existe différentes façons d'économiser. Considérez les exemples suivants :
 - 60 % d'économies sur les charges de travail Windows Server grâce à Compute Savings Plans ([d2.8xlarge, 3 ans, tout à l'avance, Windows, location partagée, us-east-2](#))
 - 73 % d'économies sur les charges de travail Windows Server grâce aux Instance Savings Plans EC2 ([d2.8xlarge, 3 ans, tout à l'avance, Windows, location partagée, us-east-2](#))
 - 28 à 41 % d'économies sur les types d'instances non exotiques ([famille T3, 3 ans, toutes initiales, Windows, location partagée, us-east-2](#))
- Économies moyennes de 25 à 40 % pour les serveurs Windows

Note

Les EC2 Instance Savings Plans offrent une réduction plus importante que les Compute Savings Plans en raison de leur flexibilité réduite. Vous vous engagez à l'utiliser à un prix réduit.

Chaque type d'utilisation informatique est assorti d'un tarif Savings Plans et d'un tarif à la demande. Le tableau suivant présente les Savings Plans et les tarifs à la demande pour chaque type de système d'exploitation. Les tarifs Savings Plans vous sont facturés sur la base de l'utilisation engagée et toute utilisation au-delà de l'engagement est facturée aux tarifs réguliers à la demande.

Instance name	Tarif Savings Plans	Économies à la demande	Tarif à la demande	Système d'exploitation	Région	Option de paiement	Durée du mandat
x2iedn.xlarge	0,32\$	61 %	0,83\$	Linux	USA Est (Virginie du Nord)	Sans frais initiaux	3
x2iedn.xlarge	2,01\$	50%	1,02\$	Windows	USA Est (Virginie du Nord)	Sans frais initiaux	3
x2iedn.xlarge	1,02\$	20 %	2,52\$	Licence Windows incluse + édition SQL Server Enterprise	USA Est (Virginie du Nord)	Sans frais initiaux	3
x2iedn.xlarge	0,32\$	61 %	0,83\$	BYOL	USA Est (Virginie du Nord)	Sans frais initiaux	3

Les Savings Plans incluent le système d'exploitation et proposent une réduction distincte pour le BYOL. Ils sont tous ventilés dans le [calculateur Compute Savings Plans](#).

Modèle de tarification des instances réservées

AWS dispose d'un autre modèle de tarification basé sur l'engagement connu sous le nom d'instances réservées. Ce modèle peut être problématique si votre calcul change alors que vous avez déjà pris un engagement, ce qui entraîne l'inutilisation des instances réservées. Les Savings Plans sont conçus pour offrir des réductions de coûts similaires à celles des [instances réservées standard et convertibles](#), mais avec une flexibilité bien supérieure. Les Compute Savings Plans proposent des tarifs réduits sur l'utilisation des instances EC2, indépendamment de la famille d'instance, de la taille,

du système d'exploitation, de la location ou de la région. Ils permettent également une flexibilité maximale.

Le tableau suivant peut vous aider à choisir entre Savings Plans et Reserved Instances.

	Instance réservation	EC2 Instance Savings Plans	Calculer les plans d'économies
Réduction moyenne sur 1 an	Jusqu'à 38 %	Jusqu'à 29 %	Jusqu'à 29 %
Réduction moyenne sur 3 ans	Jusqu'à 58 %	Jusqu'à 73 %	Jusqu'à 60 %
Famille d'instances	Fixe	Fixe	Flexible
Taille d'instance	Fixe (pas Linux)	Flexible	Flexible
Géographie	1 région	1 région	Flexible
Système d'exploitation	Fixe	Flexible	Flexible
Service	Amazon EC2 ou Amazon RDS	Amazon EC2	Amazon EC2, Fargate, Lambda
Options de paiement	Tout, partiel, pas d'avance	Tout, partiel, pas d'avance	Tout, partiel, pas d'avance
Limites d'instance	20 par zone de disponibilité	Aucune limite	Aucune limite

Note

Les Savings Plans fonctionnent en vous offrant un discount basé sur un engagement monétaire horaire. L'engagement financier horaire ne peut être ni annulé ni modifié pendant la durée de votre plan, mais vous pouvez acheter des Savings Plans supplémentaires pour

couvrir une utilisation supplémentaire. Cela vous permet de maintenir un engagement horaire constant à mesure que votre flotte augmente.

Vous pouvez utiliser des outils tels que [AWS Cost Explorer](#) et les [tableaux de bord AWS Cloud intelligents](#) pour suivre votre engagement. Cost Explorer fournit une limite de couverture qui peut aider votre entreprise à planifier sa stratégie de couverture Savings Plans. Si 75 % de votre charge de travail est stable, 75 % est un bon objectif. Cela laisse 25 % des dépenses à la demande/variables en fonction des charges de travail dynamiques. Si vous devez porter cette couverture à 85 %, vous pouvez souscrire un autre engagement Savings Plans pour augmenter l'engagement monétaire horaire.

Note

Nous vous recommandons d'acheter des Savings Plans plutôt que des instances réservées, mais les deux modèles d'engagement peuvent fonctionner ensemble si vous avez déjà acheté des instances réservées.

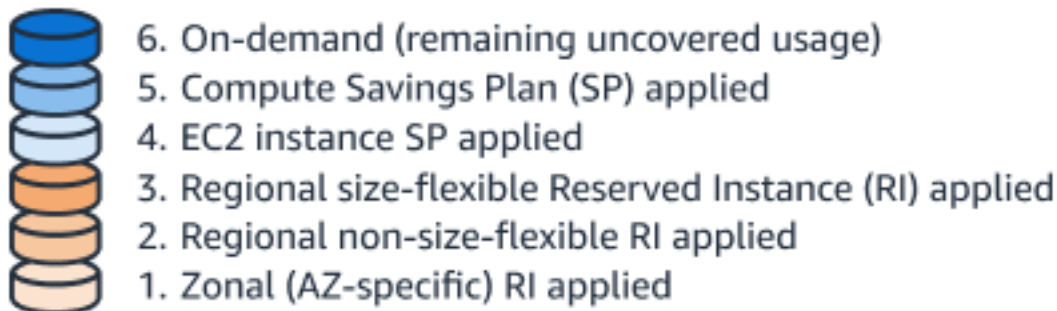
Prenons l'exemple d'un cas où vous avez acheté une instance réservée, mais que vous souhaitez commencer à essayer une option Savings Plans. Il est logique que cette combinaison s'applique à votre facturation finale. Voici une hiérarchie que vous pouvez appliquer à votre Comptes AWS :

1. L'instance réservée zonale s'applique au compte qui en est le propriétaire. S'il reste des heures à une instance réservée, cela s'applique au reste de l'organisation.
2. Les instances réservées régionales pour Windows, dont la taille n'est pas flexible, s'appliquent à l'utilisation correspondante sur le compte qui en est propriétaire. Tout ce qui reste est transféré au reste de l'organisation.
3. Les instances réservées régionales dont la taille est flexible s'appliquent au compte qui en est propriétaire (la plus petite instance de la famille d'abord, puis les plus grandes instances), puis au reste de l'organisation.
4. Les instances réservées régionales s'appliquent à toute réservation de capacité à la demande non utilisée.
5. Les EC2 Instance Savings Plans s'appliquent sur le compte qui l'a acheté.
6. Les Compute Savings Plans s'appliquent sur le compte qui l'a acheté.

Note

Les remises commencent par l'utilisation qui donne lieu à la réduction la plus élevée, puis jusqu'à la réduction la plus faible. Les instances Windows présentent généralement un potentiel de réduction inférieur à celui de Linux pour les types d'instances les plus courants (par exemple, T3, M6 et C5). Cela signifie que les instances Linux sont plus avantageuses que les instances Windows dans la plupart des cas.

Le graphique suivant montre le prix après avoir séparé les instances réservées des Savings Plans. Les plans Compute et EC2 Instance Savings Plans s'appliquent d'abord aux instances en cours d'exécution, puis aux réservations de capacité à la demande non utilisées.



Scénarios d'optimisation des coûts

Cette section couvre les scénarios d'optimisation des coûts pour les hôtes dédiés Amazon EC2 et les instances Amazon EC2 qui utilisent un modèle de facturation avec licence incluse.

Hôtes dédiés Amazon EC2

Imaginons un scénario dans lequel vous allez migrer vos charges de travail Windows locales vers. AWS Votre centre de données dispose des serveurs suivants :

- Deux serveurs avec 16 vCPU et 128 Go de RAM
- Deux serveurs avec 32 vCPU et 164 Go de RAM
- Un serveur avec 8 vCPU et 64 Go de RAM
- 16 serveurs avec vCPU et 32 Go de RAM

En outre, supposons que vous pouvez apporter votre propre licence AWS parce que vous avez suffisamment de licences pour la transférer. Le tableau suivant indique les instances de serveur que vous pouvez utiliser dans AWS.

Type d'instance	CPU	RAM	Amount
r5.4xlarge	16	128	2
r5.8xlarge	32	256	2
r5.2xlarge	8	64	1
r5.xlarge	4	32	16
			21

Une analyse montre que ces 21 machines virtuelles peuvent être réparties sur deux hôtes dédiés dotés d'un hôte de famille d'instances R5. Le tableau suivant indique le coût de ces deux hôtes dédiés.

Scénario d'hôte dédié à la demande	Paiement initial	1 mois	1 an	3 ans	Calculateur de tarification AWS
À la demande	Aucune	10 123\$	121 475\$	364 392\$	Calculateur de tarification AWS estimation
Savings Plans d'un an	Aucune	7 447\$	89 362\$	–	Calculateur de tarification AWS estimation
Savings Plans sur 3 ans	Aucune	5 476\$	65 712\$	197 128\$	Calculateur de tarification AWS estimation

Scénario d'hôte dédié à la demande	Paiement initial	1 mois	1 an	3 ans	Calculateur de tarification AWS
Savings Plans sur 3 ans avec paiement initial	84 438\$	2 755\$	117 499\$	183 618\$	Calculateur de tarification AWS estimation

Si vous avez les serveurs vers lesquels vous souhaitez migrer AWS, le prix final d'un Savings Plans d'un an est de 89 362\$ au lieu de 121 475\$ pour un prix à la demande. Cela représente une réduction de 26,5 % après un an. Si vous envisagez de rester plus AWS longtemps, vous pouvez choisir les Savings Plans sur 3 ans pour réaliser des économies encore plus importantes. Au bout de trois ans, vous payez 197 128\$ au lieu de 364 392\$. Cela se traduit par une économie de 46 % du montant total après trois ans.

Instances Amazon EC2 avec licences incluses

Imaginons un scénario dans lequel vous allez migrer une seule application à AWS trois niveaux vers laquelle vous souhaitez utiliser les licences fournies par AWS. En outre, supposons que votre application fonctionne avec les serveurs suivants :

- Deux serveurs Web avec deux V CPUs et 4 Go de RAM
- Deux serveurs d'applications avec 8 V CPUs et 16 Go de RAM
- Deux serveurs de bases de données avec 16 V CPUs et 64 Go de RAM (utilisant l'édition standard de SQL Server)

Le tableau suivant indique les instances de serveur que vous pouvez utiliser dans AWS.

Type d'instance	CPU	RAM	Amount
c5.large	2	4	2
c5.2xlarge	8	16	2
r5.2xlarge	8	64	2

Type d'instance	CPU	RAM	Amount
			6 serveurs

Le tableau suivant indique le coût de ces serveurs dans AWS.

Licence incluse par AWS	Paiement initial	1 mois	1 an	3 ans	Calculateur de tarification AWS
À la demande	Aucune	3 912\$	46 950\$	140 849\$	Calculateur de tarification AWS estimation
Savings Plans d'un an	Aucune	3 466\$	41 952\$		Calculateur de tarification AWS estimation
Savings Plans sur 3 ans sans paiement initial	Aucune	3 189\$	38 264\$	114 804\$	Calculateur de tarification AWS estimation
Savings Plans sur 3 ans avec paiement initial	112 110\$	Aucune	Aucun	Aucune	Calculateur de tarification AWS estimation

Si vous souhaitez faire fonctionner ces serveurs pour des environnements de production (24 heures sur 24, 7 jours sur 7) avec une tarification à la demande, vous devez payer un coût mensuel de 3 912\$. Le paiement de ce coût mensuel équivaut à 46 950\$ après un an et à un total de 140 849\$ après trois ans.

Si vous choisissez les Savings Plans d'un an sans paiement initial, le coût mensuel diminue à 3 466\$. À la fin de la première année, vous payez 41 952\$. Il s'agit d'un discount total de 11 %. Si vous choisissez les Savings Plans sur 3 ans sans paiement initial, le coût mensuel diminue à 3 189\$. Au bout de trois ans, vous payez 114 804\$. Cela vous permet d'économiser 18,5 %.

Recommandations en matière d'optimisation des coûts

Les deux scénarios vous permettent d'économiser de l'argent lorsque vous planifiez et prévoyez vos charges de travail. AWS Il est important de reconnaître que le discount dans le deuxième scénario est inférieur à celui du premier scénario. Dans le second scénario, le prix de la licence est inclus dans le prix du serveur cloud. AWS n'offre pas de discount sur le prix des licences, mais vous pouvez toujours apporter vos licences (selon des scénarios spécifiques) et vous AWS pouvez toujours garantir le meilleur compute/instance prix.

Nous vous recommandons de procéder comme suit pour contrôler vos AWS dépenses en ressources de calcul et d'instance :

- Recommandations d'accès
- Personnalisez les recommandations en fonction de vos besoins
- Passez en revue l'engagement horaire

Recommandations d'accès

Vous pouvez utiliser la [console Amazon EC2](#) pour accéder aux recommandations relatives à vos Savings Plans. Vous pouvez même télécharger vos recommandations pour les consulter ultérieurement au format CSV. Pour plus d'informations, consultez la section [Monitoring your Savings Plans](#) dans la documentation relative aux Savings Plans.

Personnalisez les recommandations en fonction de vos besoins

Ouvrez la [console Amazon EC2](#), développez la section Instances, puis choisissez Savings Plans. Cette page vous indique les prix des instances et des calculs avant et après avoir fait une recommandation. Vous pouvez également ajuster les facteurs suivants en fonction de votre recommandation :

- Durée — Par exemple, 1 à 3 ans
- Option de paiement : paiement initial, paiement initial partiel ou absence de paiement initial
- Historique — Par exemple, les 7, 30 ou 60 derniers jours

Passez en revue l'engagement horaire

En utilisant le même exemple, supposons que vous disposiez d'une instance qui fonctionne 24 heures sur 24, 7 jours sur 7. Il est recommandé d'utiliser un Savings Plans. Selon la taille, vous avez un prix à la demande de 120 \$/heure. Vous avez la possibilité de vous engager à hauteur de 90 \$/heure, mais cela peut varier en fonction de la région, de l'instance et de l'option d'achat. Dans cet exemple, vous pouvez économiser 25 % par rapport au coût à la demande. Vous pouvez également suivre votre utilisation et votre couverture, si elles sont inférieures au seuil que vous avez défini, et configurer une alerte lorsque le budget arrivera à expiration.

Réviser les recommandations

Nous vous recommandons de lire attentivement les recommandations de Savings Plans. AWS Je ne changerai rien sans votre permission. Ce ne sont que des recommandations et c'est à vous de les appliquer ou non.

Achetez un plan

Ouvrez la [console Amazon EC2](#), développez la section Instances, puis choisissez Savings Plans. Choisissez ensuite Purchase Savings Plans. En fonction de vos besoins, vous pouvez sélectionner les options suivantes : durée, région, famille d'instances, engagement horaire, option de paiement et même date de début. Vous pouvez choisir entre Compute Savings Plans, EC2 Instance Savings Plans SageMaker et AI Savings Plans. Pour plus d'informations, consultez [Purchasing Savings Plans](#) dans la documentation Savings Plans.

Obtenir un rapport d'utilisation

Après avoir acheté un Savings Plans, vous pouvez obtenir un rapport d'utilisation. Le rapport vous permet de vérifier votre utilisation, de voir si le plan acheté est suffisant pour couvrir et maximiser la réduction, et d'annuler ou d'ajouter de nouvelles remises. Ce rapport peut être exporté vers d'autres formats tels que CSV. Pour plus d'informations, consultez la section [Utilisation du rapport d'utilisation](#) dans la documentation Savings Plans.

Suivez les meilleures pratiques d'achat

Nous vous recommandons de suivre les bonnes pratiques suivantes avant d'acheter des Savings Plans :

- [AWS Trusted Advisor](#) À utiliser pour supprimer les ressources EC2 inactives.

- Effectuez la bonne taille avant d'acheter des Savings Plans.
- Établissez un taux horaire que vous respecterez régulièrement pendant 30 à 60 jours.
- Prenez l'engagement de couvrir la part du taux horaire constant qui convient à votre organisation. Tenez compte des fluctuations de la demande ou de la saison.
- Choisissez une révision trimestrielle du budget des Savings Plans afin de maintenir un taux constant (par exemple, un objectif de couverture de 70 % pour la couverture Savings Plans). Si le taux est inférieur à la couverture souhaitée, achetez un Savings Plans supplémentaire pour atteindre votre objectif de couverture.

Ressources supplémentaires

- [Savings Plans pour les instances réservées Amazon EC2 \(livres blancs\)](#) AWS
- [Comprendre comment les Savings Plans s'appliquent à votre AWS consommation](#) (documentation Savings Plans)
- [Annonce de la facturation par seconde pour les instances EC2 Windows Server et SQL Server](#) (AWS documentation)
- [AWS Série sur l'optimisation des coûts : vidéo sur les Savings Plans | Amazon Web Services](#) (YouTube)

Surveillez les coûts à l'aide d' AWS outils

Présentation de

La visibilité des coûts est un facteur clé pour optimiser les coûts AWS. AWS dispose d'un certain nombre d'outils que vous pouvez utiliser pour visualiser les coûts et créer des alertes en réaction à ces coûts. Il s'agit notamment d'outils qui vous aident à suivre et à déclarer vos dépenses.

AWS Budgets Cette section décrit des méthodes spécifiques pour surveiller les AWS dépenses de votre Windows, afin que vous puissiez suivre et réagir en fonction de vos exigences budgétaires. Cela inclut l'ajout des balises nécessaires à vos ressources Windows EC2. Ces balises vous permettent de surveiller correctement Windows EC2 et les autres services Microsoft en utilisant AWS Budgets.

En surveillant les dépenses et en créant des alertes à l'aide d' AWS outils, vous pouvez être mieux informé des dépenses actuelles, des dépenses prévues et des anomalies de dépenses. Si vous utilisez [Savings Plans](#) pour réduire le prix horaire de votre instance EC2, nous vous recommandons

de consulter l'utilisation globale et la couverture des Savings Plans. Cela peut vous aider à réaliser des économies en permanence. Vous pouvez l'utiliser AWS Cost Explorer pour consulter l'inventaire des Savings Plans et obtenir des recommandations pour des plans d'épargne supplémentaires en fonction de votre utilisation précédente. Vous pouvez également suivre des dépenses spécifiques en utilisant [AWS Budgets](#) et en configurant [AWS Cost Anomaly Detection](#).

Recommandations en matière d'optimisation des coûts

Nous vous recommandons de suivre les étapes suivantes pour optimiser vos coûts en utilisant AWS Budgets Cost Explorer et la détection des anomalies :

- Étiqueter les ressources Windows EC2
- Configurez des alertes en utilisant AWS Budgets
- Activer la détection des anomalies de coûts
- Obtenez une analyse des dépenses en temps réel
- Afficher les dépenses incluses dans les licences pour Windows à l'aide de Cost Explorer

Étiqueter les ressources Windows EC2

Pour contrôler efficacement vos AWS dépenses, vous devez établir une [stratégie de balisage](#) pour les charges de travail que vous souhaitez surveiller. Ceci est important pour que vous puissiez regrouper les ressources de manière catégorique et être informé des dépenses spécifiques, par opposition aux dépenses d'utilisation générales. Vous pouvez utiliser des ressources de balisage qui non seulement aident à réduire les coûts, mais peuvent également être utilisées à d'autres fins, telles que [AWS Systems Manager l'automatisation](#). En outre, nous vous recommandons de mettre en œuvre une certaine gestion des [balises requises](#).

Pour suivre vos dépenses dans AWS Budgets Cost Explorer et Cost Anomaly Detection, vous devez vous assurer que les balises appropriées sont en place. Vous pouvez utiliser des balises pour définir un budget spécifique pour les articles correspondant à ces balises afin d'être alerté lorsque les dépenses augmentent.

Par exemple, vous pouvez utiliser une simple balise telle que Key=OS Value=Windows. Toutes vos instances Windows sont ainsi regroupées dans un seul groupe pour lequel vous pouvez suivre les dépenses. Vous pouvez également utiliser des balises pour d'autres éléments, tels que Systems Manager. Après avoir créé un tag, vous devez l'activer pour le suivi des coûts. Envisagez d'ajouter une [AWS Config règle qui surveille les balises](#) associées à certaines ressources. AWS Config peut

vous avertir si certaines ressources en cours d'exécution ne contiennent pas les balises appropriées, ce qui vous fournit une représentation précise de vos dépenses Windows EC2.

Une fois vos tags en place, vous pouvez créer un budget personnalisé dans AWS Billing. Cela vous donne une visibilité sur vos dépenses liées à Windows EC2. Vous pouvez définir un budget quotidien ou mensuel.

Configurez des alertes à l'aide de AWS Budgets

Dans cet exemple de scénario, vous créez un budget quotidien pour Windows EC2. Il s'agit d'un budget récurrent qui utilise l'option d'ajustement automatique pour suivre vos dépenses et ajuster le budget en conséquence. Si votre environnement est statique, vous pouvez utiliser un budget fixe à la place. Assurez-vous de choisir une période de référence (par exemple, 30 jours).

1. Connectez-vous à la [AWS Cost Management console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation, choisissez Budgets.
3. En haut de la page, choisissez Créer un budget.
4. Sous Configuration du budget, choisissez Personnaliser (avancé).
5. Sous Types de budget, sélectionnez Budget des coûts. Ensuite, choisissez Suivant.
6. Sous Détails, dans le champ Nom du budget, entrez le nom de votre budget. Par exemple, les dépenses liées à Windows EC2.
7. Sous Définir le montant du budget, dans Période, sélectionnez Quotidien.
8. Pour le type de renouvellement du budget, choisissez Budget récurrent pour un budget qui est réinitialisé après la période budgétaire.
9. Pour Date de début, choisissez la date ou la période de début pour commencer le suivi par rapport au montant budgétisé.
10. Pour la méthode de budgétisation, choisissez Ajustement automatique (Nouveau).
11. Pour la plage de temps de référence, choisissez Plage personnalisée, puis entrez 30 jours.
12. Choisissez Suivant.
13. Dans la section Étendue du budget, sélectionnez Filtrer les dimensions de AWS coût spécifiques. C'est ici que les balises sont utilisées pour créer les bonnes dimensions. AWS Budgets ne prend pas en charge le type de plateforme en tant qu'option dans ses filtres. Pour cette raison, vous devez appliquer des balises de système d'exploitation.
14. Choisissez Ajouter un filtre, puis sélectionnez l'option Tag dans Dimensions.

15. Choisissez la balise OS, puis choisissez la valeur Windows correspondante afin de créer un budget pour la balise.
16. Choisissez Suivant.
17. Sur la page Configurer les alertes, choisissez Ajouter un seuil d'alerte. Ici, vous configurez deux alertes : l'une pour un seuil de 50 % et l'autre pour un seuil de 100 %. Si le seuil d'alerte de 50 % est dépassé avant le milieu du mois, un avertissement sera émis. Ainsi, vous pouvez vérifier si vos dépenses sont supérieures aux prévisions et réagir avant la fin du mois.
18. Pour Seuil, entrez 50 et sélectionnez % du montant budgétisé.
19. Pour Trigger, choisissez Actual.
20. Pour les destinataires des e-mails, entrez une adresse e-mail. Ajoutez une autre alerte pour un seuil de 100.

Note

Cet exemple utilise une notification par e-mail pour l'alerte, mais vous pouvez également utiliser d'autres approches, telles que [Slack](#).

Activer la détection des anomalies de coûts

Vous pouvez utiliser vos étiquettes de coût pour configurer des alertes de dépenses en cas d'anomalie. Par exemple, vous pouvez l'utiliser [AWS Cost Anomaly Detection](#) pour créer des moniteurs de vos dépenses et être alerté lorsque le système détecte des dépenses anormales sur votre compte.

Pour configurer un moniteur et des alertes pour les balises Key=OS et Value=Windows que vous avez créées précédemment, procédez comme suit :

1. Connectez-vous à la [AWS Cost Management console AWS Management Console et ouvrez-la](#).
2. Dans le panneau de navigation, choisissez Cost Anomaly Detection (Détection des anomalies de coûts).
3. Choisissez l'onglet Moniteurs de coûts, puis choisissez Créer un moniteur.
4. À l'étape 1, choisissez le tag de répartition des coûts comme type de moniteur.
5. Pour la clé Cost Allocation Tag, choisissez Windows EC2 spend.
6. Pour la valeur de l'étiquette de répartition des coûts, sélectionnez Windows.
7. Dans Nom de votre moniteur, saisissez les dépenses liées à Windows EC2.

8. Choisissez Suivant.

9. Pour créer un abonnement aux alertes, sélectionnez Créer un nouvel abonnement. Si vous disposez d'abonnements existants, sélectionnez Choose an existing subscription (Choisir un abonnement existant).

10 Dans le champ Nom de l'abonnement, entrez une anomalie de dépenses Windows EC2.

11 Pour Fréquence des alertes, choisissez Résumés quotidiens.

12 Pour les destinataires des alertes, entrez votre adresse e-mail.

13. Choisissez Ajouter un seuil. Pour Seuil, entrez 10, puis sélectionnez le pourcentage au-dessus de la vitesse attendue.

14. Choisissez Create monitor (Créer un contrôle).

Obtenez une vue en temps réel des dépenses

Une alerte est un outil utile pour surveiller vos dépenses Windows EC2, mais vous devez utiliser Cost Explorer si vous souhaitez avoir une vue en temps réel des dépenses. Regardez cette vidéo pour découvrir comment Cost Explorer vous permet d'analyser et de réduire vos coûts EC2. Pour plus d'informations, regardez la vidéo [AWS Supports You | Understanding and Reducing Your EC2 Costs](#) sur le YouTube sujet.

Afficher les dépenses incluses dans les licences pour Windows

Vous pouvez consulter les dépenses Windows EC2 enregistrées sur votre compte à l'aide de Cost Explorer. Pour voir les dépenses incluses dans les licences pour Windows, vous devez définir les [filtres](#) appropriés suivants dans Cost Explorer :

- Pour Platform, choisissez Windows (Amazon VPC). Pour le fonctionnement de l'API, choisissez:0002RunInstance. Il s'agit du AWS Billing code pour les instances Windows EC2 incluses dans une licence.
- Si vous souhaitez consulter les dépenses relatives à votre instance BYOL, remplacez :0002 par :0800RunInstance. RunInstance Il s'agit du code de facturation pour Windows EC2 BYOL.

Grâce à cette visibilité dans Cost Explorer, vous pouvez rapidement filtrer vos coûts en fonction de ce que vous dépensez exactement pour Windows EC2. Si vous souhaitez approfondir vos AWS dépenses, vous pouvez AWS Cost and Usage Report les filtrer au niveau de chaque instance. Vous pouvez également générer des rapports qui peuvent être visualisés dans Amazon Quick et créer des tableaux de bord personnalisés.

Pour plus d'informations, regardez la [vidéo intitulée AWS Supports You - Visualizing Your Cost and Usage Reports](#) sur YouTube.

Ressources supplémentaires

- [Configuration des balises requises avec AWS Config](#) (AWS Config documentation)
- [AWS Budgets Tutoriel - Configuration des alertes pour AWS Billing | Amazon Web Services](#) (YouTube)
- [AWS Cost and Usage Report Bibliothèque de requêtes](#) (AWS Well-Architected Labs)

SQL Server

Les clients exécutent des charges de travail Microsoft AWS depuis plus de 15 ans, soit plus longtemps que tout autre fournisseur de cloud. Cela s'explique en grande partie par le fait qu'il AWS possède le plus d'expérience avec les applications Microsoft dans le cloud et qu'il propose la meilleure plateforme pour Windows Server et Microsoft SQL Server dans les domaines suivants :

- Performances et fiabilité accrues
- Services de sécurité et d'identité améliorés
- Plus d'assistance à la migration
- Les capacités les plus étendues et les plus approfondies
- Réduction du coût total de possession (TCO)
- Options de licence flexibles

AWS prend en charge tout ce qui est nécessaire pour créer et exécuter des applications Windows basées sur SQL Server, notamment Active Directory, .NET, SQL Server, Windows desktop as a service et toutes les versions prises en charge de Windows Server. Grâce à son expertise éprouvée, AWS elle peut vous aider à lever, déplacer, refactoriser ou même moderniser facilement vos charges de travail Windows.

Cette section du guide couvre les sujets suivants :

- [Choisissez une solution de haute disponibilité et de reprise après sinistre](#)
- [Comprendre les licences SQL Server](#)
- [Sélectionnez l'instance EC2 adaptée aux charges de travail SQL Server](#)
- [Consolider les instances](#)
- [Comparez les éditions de SQL Server](#)
- [Évaluer l'édition SQL Server Developer](#)
- [Évaluer SQL Server sous Linux](#)
- [Optimisation des stratégies de sauvegarde de SQL Server](#)
- [Moderniser les bases de données SQL Server](#)
- [Optimisation du stockage pour SQL Server](#)
- [Optimisez les licences SQL Server à l'aide de Compute Optimizer](#)
- [Optimisez le dimensionnement de SQL Server à l'aide de Compute Optimizer](#)

- [Consultez Trusted Advisor les recommandations relatives aux charges de travail SQL Server](#)

Choisissez une solution de haute disponibilité et de reprise après sinistre

Présentation de

Nous vous recommandons de concevoir une architecture pour votre déploiement de SQL Server AWS qui réponde aux besoins de votre entreprise tout en respectant vos [objectifs de reprise après sinistre \(DR\)](#), notamment votre objectif de temps de restauration (RTO) et votre objectif de point de reprise (RPO). Les solutions suivantes peuvent vous aider à concevoir l'architecture appropriée pour SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2) tout en optimisant les coûts de vos charges de travail SQL Server.

- Groupes de disponibilité SQL Server Always On : les groupes de disponibilité SQL Server Always On fournissent une haute disponibilité et une reprise après sinistre (HA/DR) solutions for SQL Server databases. An availability group consists of a set of user databases that fail over together. Always On availability groups also provide redundancy at the database level, but don't require shared storage—each replica has its own local storage. You can deploy this feature as an HA/DR solution). Pour plus d'informations, voir [Qu'est-ce qu'un groupe de disponibilité permanent ?](#) dans la documentation Microsoft.
- Instances de cluster SQL Server Always On Failover (FCI) : SQL Server Always On FCIs utilise le cluster de basculement Windows Server (WSFC) pour fournir une haute disponibilité au niveau de l'instance SQL Server. FCIs nécessitent un stockage partagé pour héberger les bases de données. Vous pouvez utiliser le stockage par blocs partagé ou le stockage de fichiers partagé. Par exemple, vous pouvez utiliser Amazon FSx pour Windows File Server ou Amazon FSx for NetApp ONTAP comme solution de stockage partagé avec plusieurs zones de disponibilité. Pour plus d'informations, consultez la section [Instances de cluster Always On Failover \(SQL Server\)](#) dans la documentation Microsoft.
- SIOS DataKeeper — Le SIOS DataKeeper peut vous aider à répondre aux exigences de haute disponibilité et de reprise après sinistre en activant un SQL Server FCI qui couvre à la fois les zones de disponibilité et. Régions AWS SIOS DataKeeper crée un SAN virtuel en cluster en utilisant les volumes Amazon Elastic Block Store (Amazon EBS) locaux et utilise la réplication synchrone entre les zones de disponibilité pour la haute disponibilité, tout en utilisant la réplication asynchrone entre les régions et pour la reprise après sinistre. Pour plus d'informations, consultez la

section [Protection de haute disponibilité pour les applications Windows](#) dans la documentation du SIOS.

- **Groupes de disponibilité distribués** — Les groupes de disponibilité distribués sont un type spécial de groupe de disponibilité qui s'étend sur deux groupes de disponibilité Always On distincts. Un groupe de disponibilité peut résider dans deux régions distinctes (par exemple, us-east-1 et us-west-1). Vous pouvez considérer un groupe de disponibilité distribué comme un groupe de disponibilité composé de groupes de disponibilité, car les groupes de disponibilité Always On sous-jacents sont configurés sur deux clusters WSFC différents. L'édition Enterprise de SQL Server est requise pour déployer des groupes de disponibilité distribués. Pour plus d'informations, consultez la section [Groupes de disponibilité distribués](#) dans la documentation Microsoft.
- **Expédition des journaux** : vous pouvez implémenter l'expédition des journaux pour protéger vos bases de données dans plusieurs régions, dans les rares cas où une région est affectée et devient indisponible. En fonction de la fréquence d'expédition des transactions et des journaux, vous pouvez atteindre le RPO et le RTO en quelques minutes. Pour plus d'informations, consultez [À propos de l'expédition des journaux \(SQL Server\)](#) dans la documentation Microsoft.
- **Reprise après sinistre AWS Elastic**— Elastic Disaster Recovery est une application logicielle en tant que service (SaaS) qui gère la réplication des serveurs depuis n'importe quelle infrastructure à AWS des fins de reprise après sinistre. Vous pouvez également utiliser Elastic Disaster Recovery pour répliquer SQL Server dans différentes régions. Elastic Disaster Recovery est une solution basée sur un agent qui réplique des machines virtuelles complètes, y compris le système d'exploitation, toutes les applications installées et toutes les bases de données dans une zone intermédiaire. Pour plus d'informations, consultez [Qu'est-ce qu'Elastic Disaster Recovery ?](#) dans la documentation d'Elastic Disaster Recovery.
- **AWS Database Migration Service (AWS DMS)** — AWS DMS prend en charge la migration en direct des données vers et depuis AWS, y compris une autre région. Vous pouvez utiliser cette fonctionnalité pour configurer une instance SQL Server distincte dans une région différente afin de servir de base de données de reprise après sinistre. Pour plus d'informations, voir [Qu'est-ce que c'est AWS Database Migration Service ?](#) dans la AWS DMS documentation.

Groupes de disponibilité SQL Server Always On

Si vous utilisez l'édition Enterprise de SQL Server uniquement pour un [groupe de disponibilité Always On à haute disponibilité](#), vous pouvez passer à l'édition SQL Server Standard en tirant parti des groupes de disponibilité de base. Vous pouvez réduire les coûts de 65 à 75 % en utilisant des groupes de disponibilité de base au lieu de groupes de disponibilité Always On.

Note

Pour plus d'informations sur les différences de coûts entre les différentes éditions de SQL Server, consultez la section [Comparer les éditions de SQL Server](#) de ce guide.

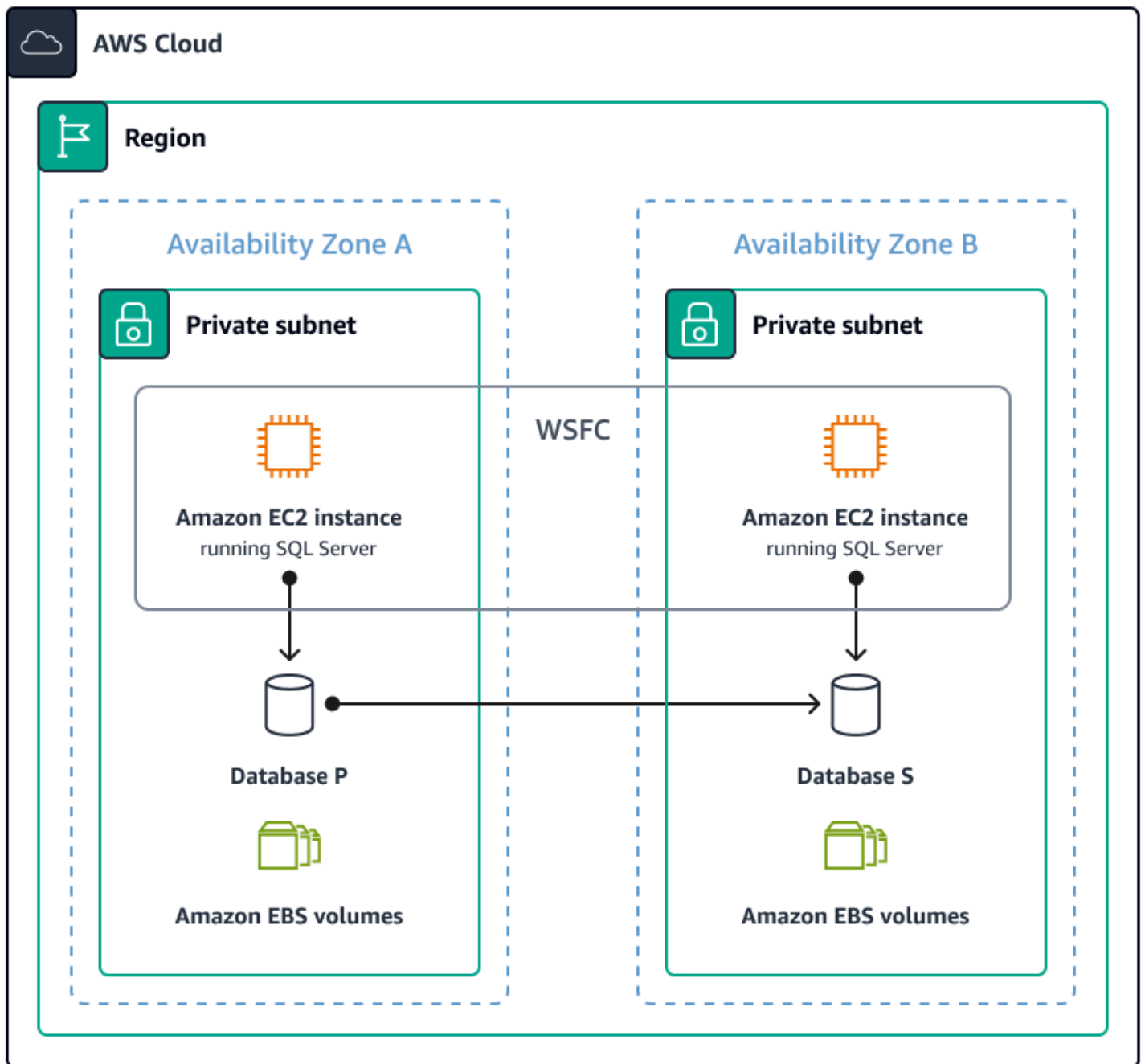
Fonctions

- Disponible dans l'édition standard de SQL Server
- Limite de deux répliques (principale et secondaire)
- Aucun accès en lecture sur la réplique secondaire
- Aucun contrôle d'intégrité sur les répliques secondaires

Limites

- Support pour une seule base de données de disponibilité par groupe de disponibilité
- Les groupes de disponibilité de base ne peuvent pas faire partie d'un groupe de disponibilité distribué


Le schéma suivant montre un exemple d'architecture pour une solution Windows Server Failover Cluster.



Instances de cluster Always On Failover de SQL Server

Vous pouvez utiliser des instances de cluster failover (FCIs) pour garantir la continuité des opérations de base de données tout en minimisant les temps d'arrêt et en réduisant le risque de perte de données. FCIs offrent une solution fiable si vous recherchez une haute disponibilité pour votre base de données SQL Server sans configuration de réplication en lecture.

Contrairement aux groupes de disponibilité, les FCIs peuvent fournir une solution de basculement fiable sans nécessiter l'édition Enterprise de SQL Server. Exigez plutôt uniquement les licences de l'édition standard de SQL Server. Vous pouvez utiliser les FCIs pour réduire les coûts de licence SQL Server de 65 à 75 %.

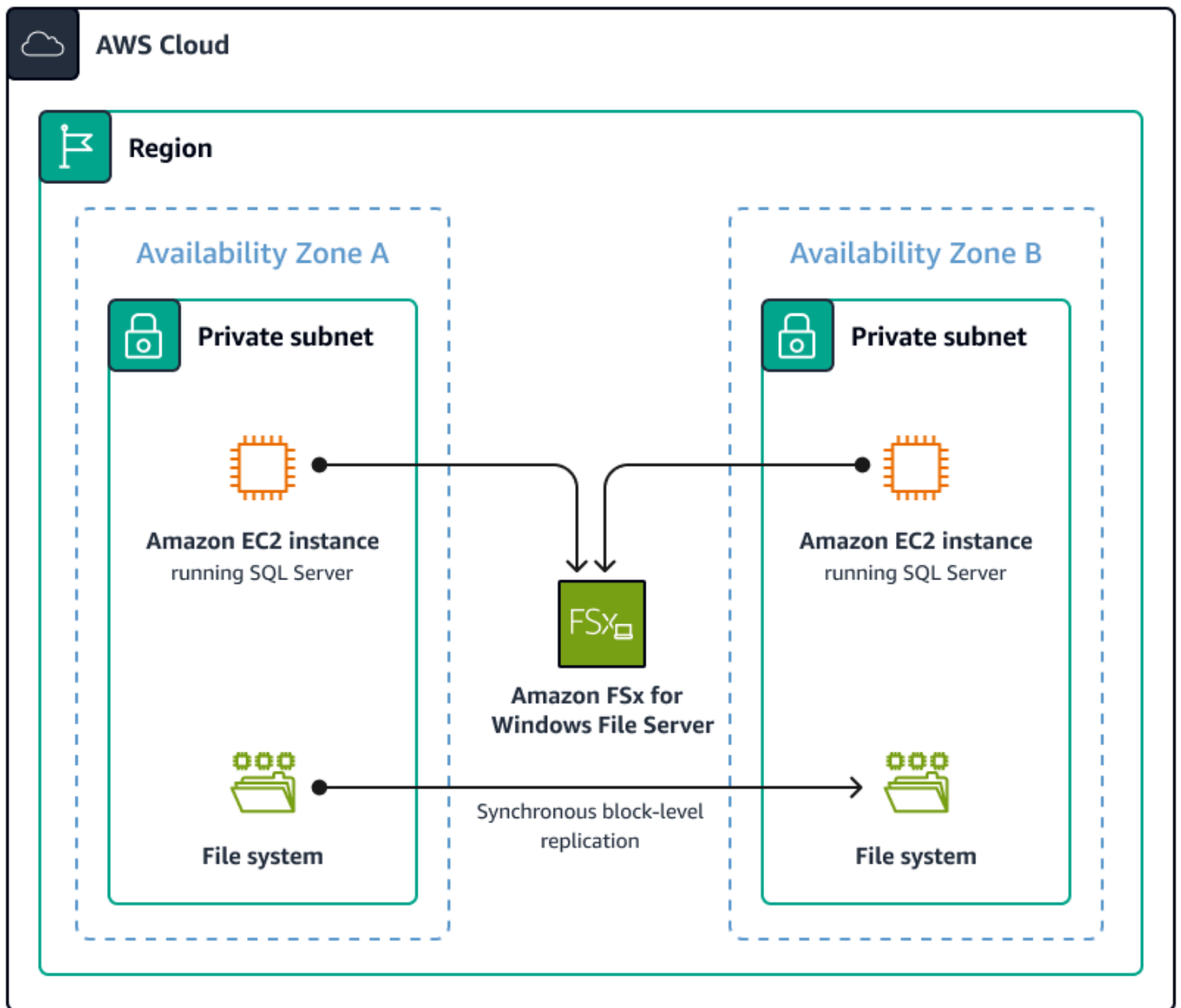
 Note

Pour plus d'informations sur les différences de coûts entre les éditions de SQL Server, consultez la section [Comparer les éditions de SQL Server](#) de ce guide.

Éléments à prendre en compte :

- Amazon FSx pour Windows File Server propose une solution puissante pour répondre à vos exigences de stockage partagé SQL Server FCI. Vous pouvez utiliser Windows File Server FSx pour éviter d'avoir à acheter une licence pour une solution de réplication du stockage et pour gérer vous-même le stockage partagé. Cela peut se traduire par des économies de coûts importantes de 30 à 40 %. Pour plus d'informations, consultez le billet [Simplifiez vos déploiements de haute disponibilité Microsoft SQL Server à l'aide d'Amazon FSx pour Windows File Server](#) sur le blog AWS de stockage.
- Grâce au [résumé des avantages de la Software Assurance](#) (PDF téléchargeable) et au modèle Bring Your Own License (BYOL), vous pouvez bénéficier des avantages du basculement passif, à condition que le serveur secondaire soit passif. Cela permet de réaliser des économies sur les licences SQL, car vous n'avez pas à fournir de licences au nœud passif du cluster.

Le schéma suivant montre un exemple d'architecture pour un SQL Server FCI à l'aide d'un serveur FSx de fichiers Windows.




SIOS DataKeeper

Nous vous recommandons de prendre en compte les exigences de stockage partagé si vous prévoyez de déployer SQL Server FCIs sur AWS. Les installations sur site traditionnelles utilisent généralement un réseau de stockage (SAN) pour répondre aux exigences de stockage partagé, mais cette option n'est pas viable. AWS Amazon FSx pour Windows File Server est la solution de stockage recommandée pour SQL Server FCI on AWS, mais elle présente des limites qui empêchent d'ajouter des serveurs de cluster dans différents Régions AWS serveurs.

Vous pouvez utiliser le [SIOS DataKeeper](#) pour créer un SQL Server FCI qui couvre à la fois les zones de disponibilité et les régions tout en réduisant les coûts de 58 à 71 %. Le SIOS DataKeeper peut vous aider à tirer parti des avantages de la haute disponibilité de FCI. Cela fait du SIOS DataKeeper une solution rentable et fiable pour les entreprises.

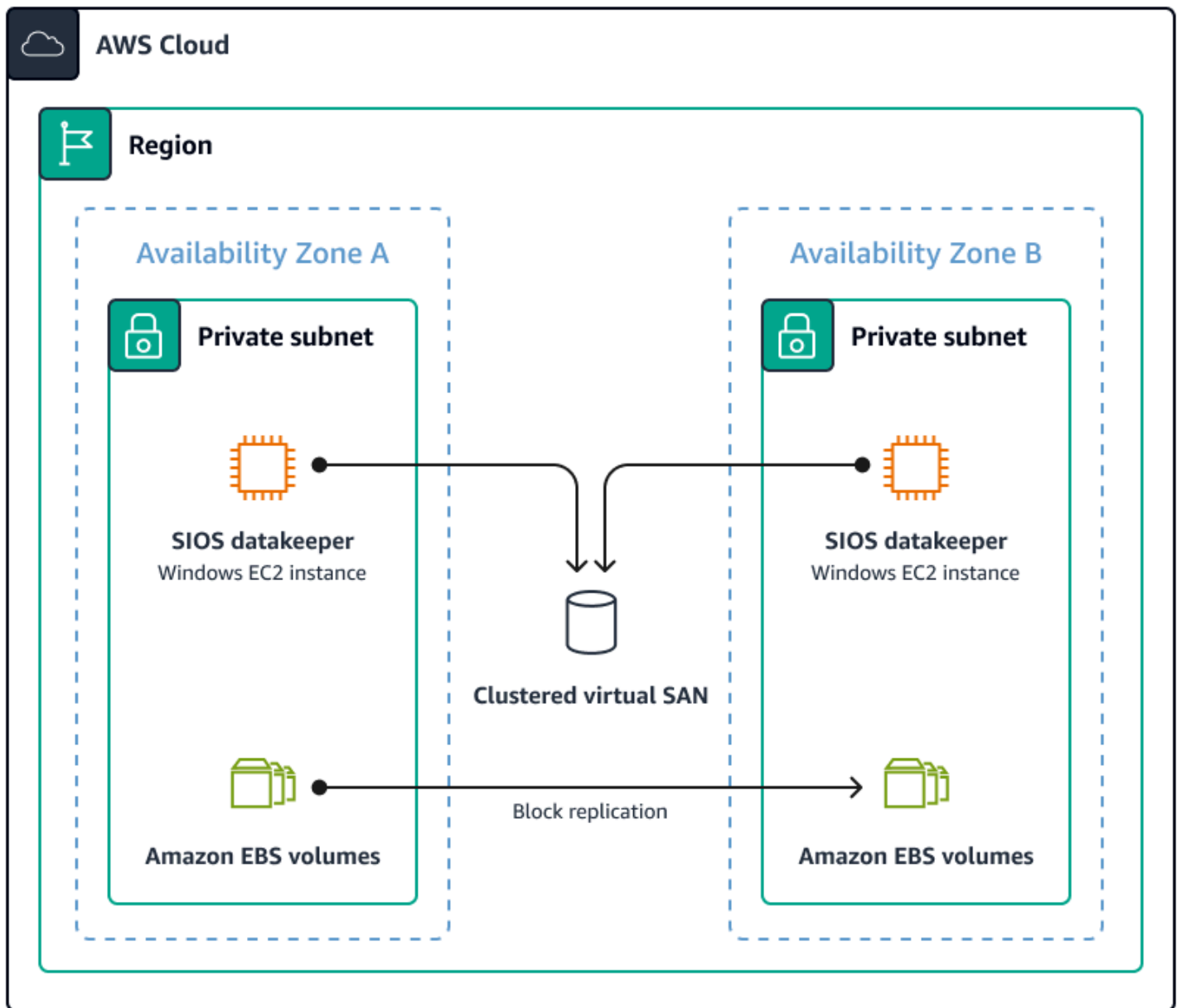
Tenez compte des avantages supplémentaires suivants liés à l'utilisation du SIOS DataKeeper :

- SIOS DataKeeper crée un SAN virtuel en cluster à l'aide de volumes EBS locaux et utilise la réplication synchrone entre les zones de disponibilité pour une haute disponibilité. Pour la reprise après sinistre, le SIOS DataKeeper utilise la réplication asynchrone entre les régions.
- SIOS DataKeeper fournit des fonctionnalités de clustering de niveau professionnel en utilisant l'édition standard de SQL Server. Cela permet de réduire les coûts de licence SQL Server de 65 à 75 % par rapport à la mise en œuvre de la haute disponibilité avec les groupes de disponibilité SQL Server Always On qui utilisent l'édition Enterprise de SQL Server. Avec SIOS DataKeeper, vous pouvez créer un environnement SQL Server hautement disponible, flexible et rentable qui répond aux besoins de votre entreprise.

 Note

Pour plus d'informations sur les différences de coûts entre les éditions de SQL Server, consultez la section [Comparer les éditions de SQL Server](#) de ce guide.

Le schéma suivant montre un exemple d'architecture pour un SQL Server FCI utilisant une solution Virtual SAN en cluster.

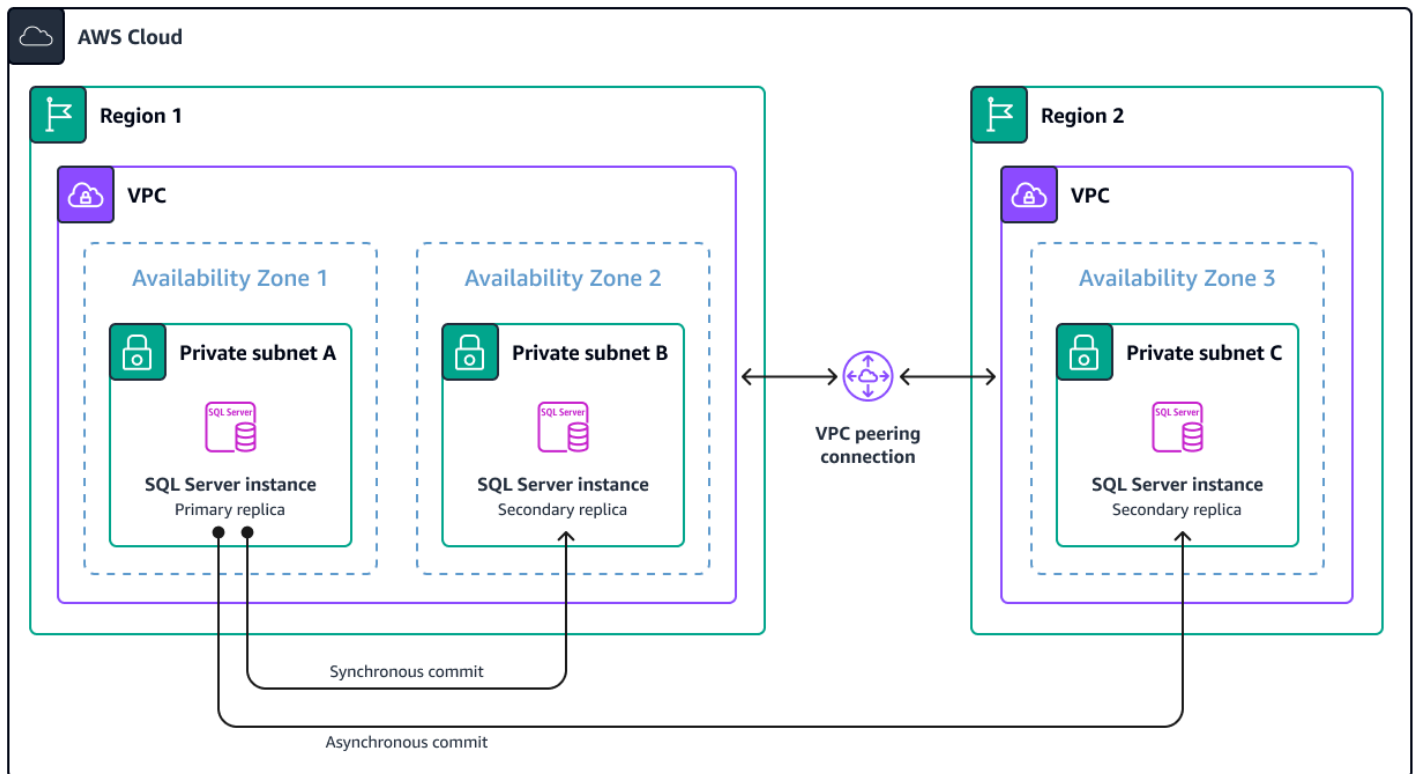


Groupes de disponibilité Always On

Vous pouvez utiliser les groupes de disponibilité Always On à des fins de haute disponibilité et de reprise après sinistre. Vous pouvez atteindre une haute disponibilité en déployant SQL Server sur deux zones de disponibilité au sein d'une même région. Vous pouvez réaliser une reprise après sinistre en étendant les groupes de disponibilité à toutes les régions.

Le schéma suivant montre un exemple d'architecture pour une solution basée sur les groupes de disponibilité Always On. Les répliques de la région 1 du diagramme utilisent un commit synchrone,

qui permet un basculement automatique du groupe de disponibilité. La réplique de la région 2 utilise un commit asynchrone, qui nécessitera un basculement manuel du groupe de disponibilité.



Groupes de disponibilité distribués

Pour les déploiements critiques de SQL Server où vous ne pouvez pas faire de compromis sur la fiabilité ou la reprise après sinistre, nous recommandons une approche multirégionale. Répartir vos groupes de disponibilité sur plusieurs régions est la solution la plus résiliente pour assurer la continuité des activités et minimiser les temps d'arrêt.

Cette architecture tire pleinement parti des fonctionnalités d'Amazon FSx pour Windows File Server, notamment le stockage partagé, la réplification synchrone au niveau des blocs et SQL Server. FCIs Ces fonctionnalités vous permettent de créer un environnement SQL Server hautement disponible qui couvre plusieurs zones de disponibilité. En reproduisant cette configuration dans une autre région, vous obtenez un système entièrement redondant capable de gérer les perturbations les plus graves. Ce qui distingue cette solution, c'est le niveau de flexibilité et de sécurité qu'elle fournit. L'architecture indépendante du domaine des groupes de disponibilité distribués permet aux serveurs de clusters Windows sous-jacents de rejoindre différents domaines Active Directory, tandis que l'authentification basée sur des certificats garantit une protection maximale de vos environnements SQL Server et répond à des exigences RTO et RPO élevées pour une stratégie de reprise après

sinistre multirégionale. Pour plus d'informations sur la création d'une architecture multirégionale, voir [Notes de terrain : création d'une architecture multirégionale pour SQL Server à l'aide de FCI et de groupes de disponibilité distribués](#) dans le blog sur l' AWS architecture.

Le schéma suivant montre un exemple d'architecture pour une solution multirégionale utilisant des groupes de disponibilité distribués.



Expédition de journaux

L'expédition des journaux est une méthode éprouvée, fiable et rentable pour protéger vos bases de données dans toutes les régions en cas de panne imprévue. Organisations utilisent le transport de logs pour protéger leurs données depuis des décennies.


Si vous implémentez l'expédition des journaux AWS, vous pouvez atteindre le RPO et le RTO en quelques minutes, en fonction de la fréquence des transactions et des tâches d'expédition des journaux. Dans le cas peu probable où une région deviendrait inaccessible, l'expédition des logs garantit la sécurité et la restauration de vos données.

Tenez compte des avantages supplémentaires suivants liés à l'utilisation de l'expédition de grumes :

- Réduisez les coûts et répondez aux exigences de votre entreprise en utilisant l'expédition de journaux pour renforcer la résilience en cas de sinistre dans toutes les régions. L'expédition des

journaux réduit votre coût total de possession car vous n'avez besoin que de licences SQL Server Standard Edition ou SQL Server Web Edition.

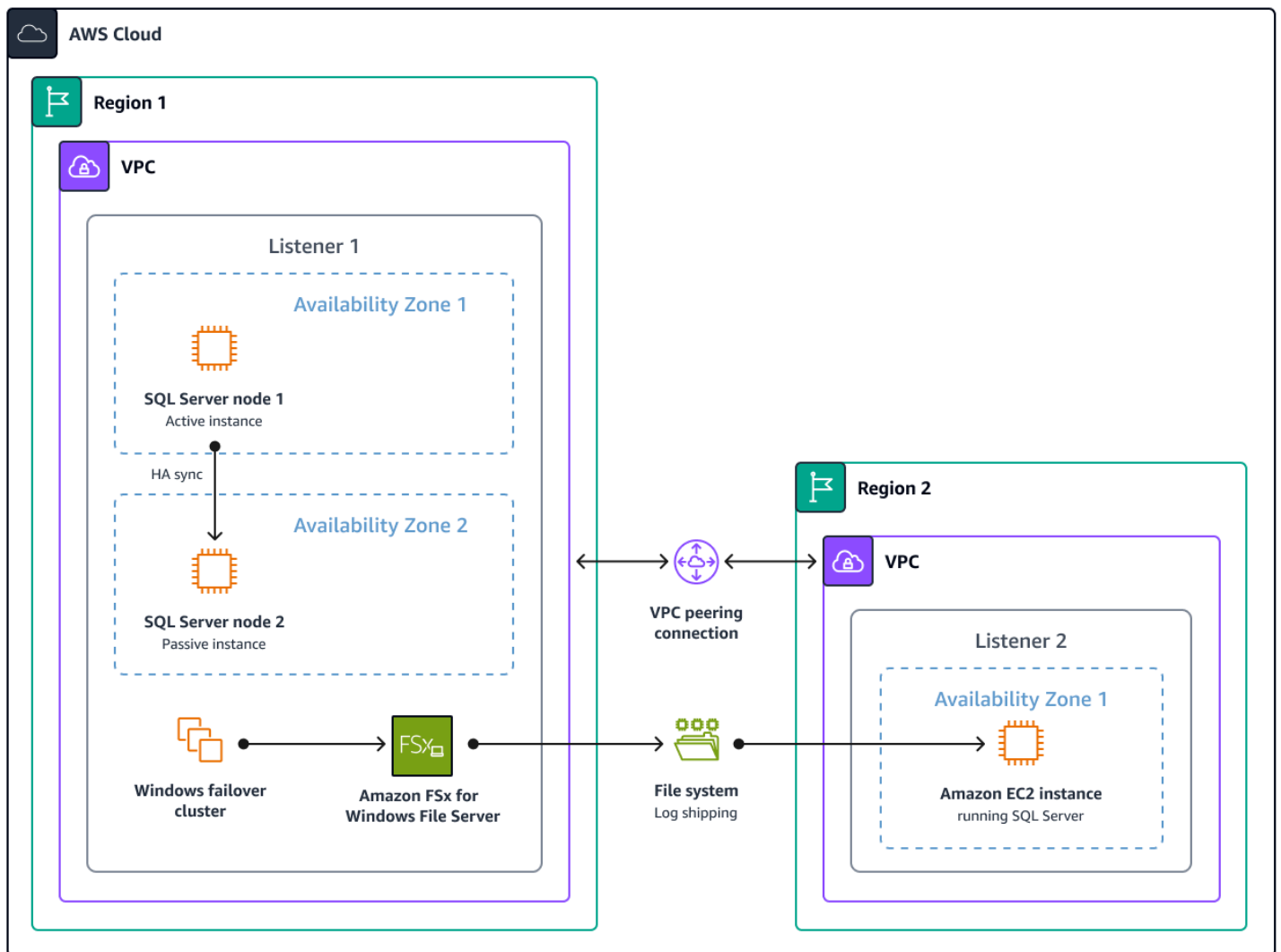
- Supprimez les coûts de licence liés à un recovery/passive serveur en cas de sinistre en utilisant l'expédition des journaux avec Active [Software Assurance](#). Seul le serveur primary/active SQL doit être titulaire d'une licence lorsque vous utilisez l'expédition de journaux avec Software Assurance.
- Réduisez les coûts de licence SQL Server de 65 à 75 % en supprimant la nécessité de configurer des groupes de disponibilité distribués entre les régions dans l'édition SQL Server Enterprise. Pour ce faire, vous pouvez utiliser l'édition Standard de SQL Server et SQL Server FCIs combinés à l'expédition des journaux pour répondre à vos exigences en matière de reprise après sinistre.

 Note

Pour plus d'informations sur les différences de coûts entre les éditions de SQL Server, consultez la section [Comparer les éditions de SQL Server](#) de ce guide.

Pour plus d'informations, consultez la section [Extension de SQL Server DR à l'aide de l'expédition de journaux pour la configuration de SQL Server FCI avec Amazon FSx pour Windows](#) dans le blog AWS d'architecture.

Le schéma suivant montre un exemple d'architecture pour une solution d'expédition de journaux.

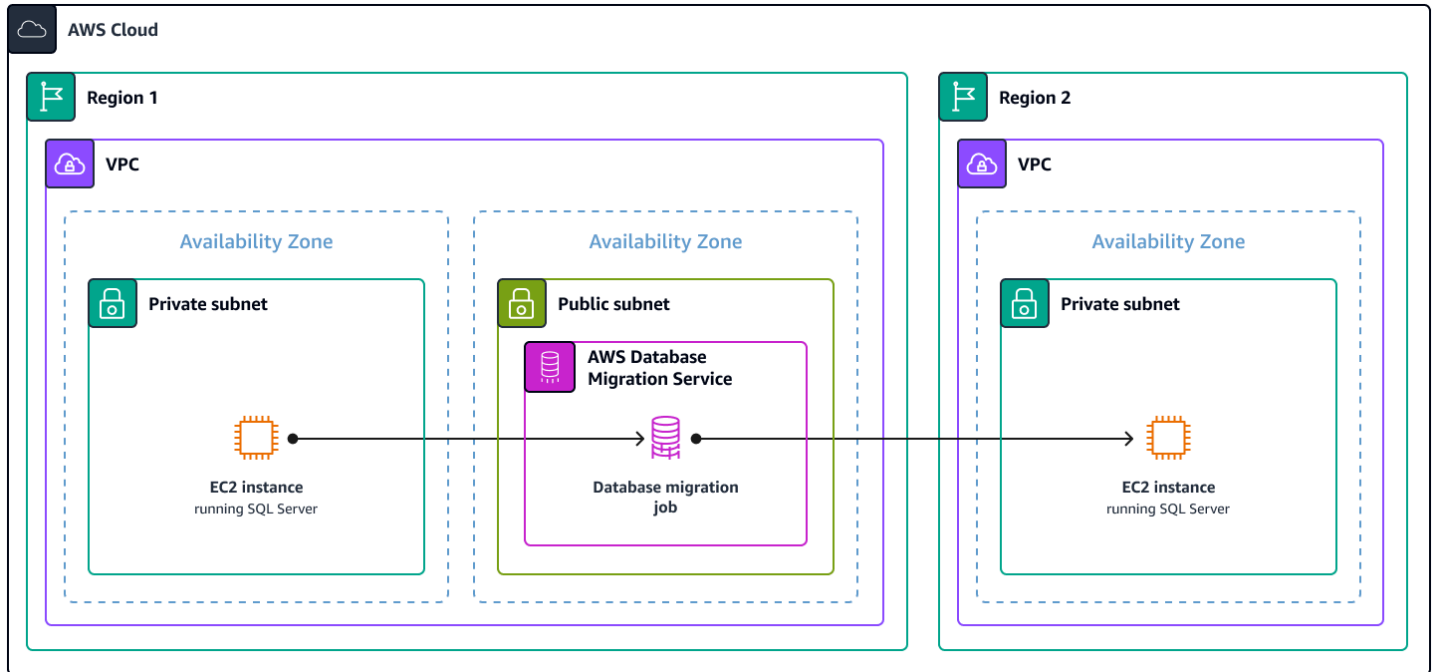


AWS Database Migration Service

Vous pouvez utiliser AWS Database Migration Service (AWS DMS) pour concevoir une HA/DR solution en fonction des besoins de votre application. AWS DMS vous permet de copier facilement des données vers une base de données SQL Server secondaire dans la même région (HA) ou entre régions (DR). Cette approche est techniquement solide et vous permet de maximiser votre investissement dans l' AWS infrastructure tout en optimisant l'utilisation de vos ressources.

AWS DMS est un service rentable. Vous êtes facturé uniquement pour les ressources du processeur utilisées pendant le processus de transfert et pour tout stockage de journal supplémentaire. Cela signifie que vous pouvez bénéficier de cette solution sans encourir de coûts supplémentaires importants. Vous pouvez vous AWS DMS en servir pour garantir la disponibilité et l'accessibilité de vos données, tout en minimisant les coûts associés aux licences et à l'utilisation des ressources.

Le schéma suivant montre un exemple d'architecture pour une solution basée sur AWS DMS.



Reprise après sinistre AWS Elastic

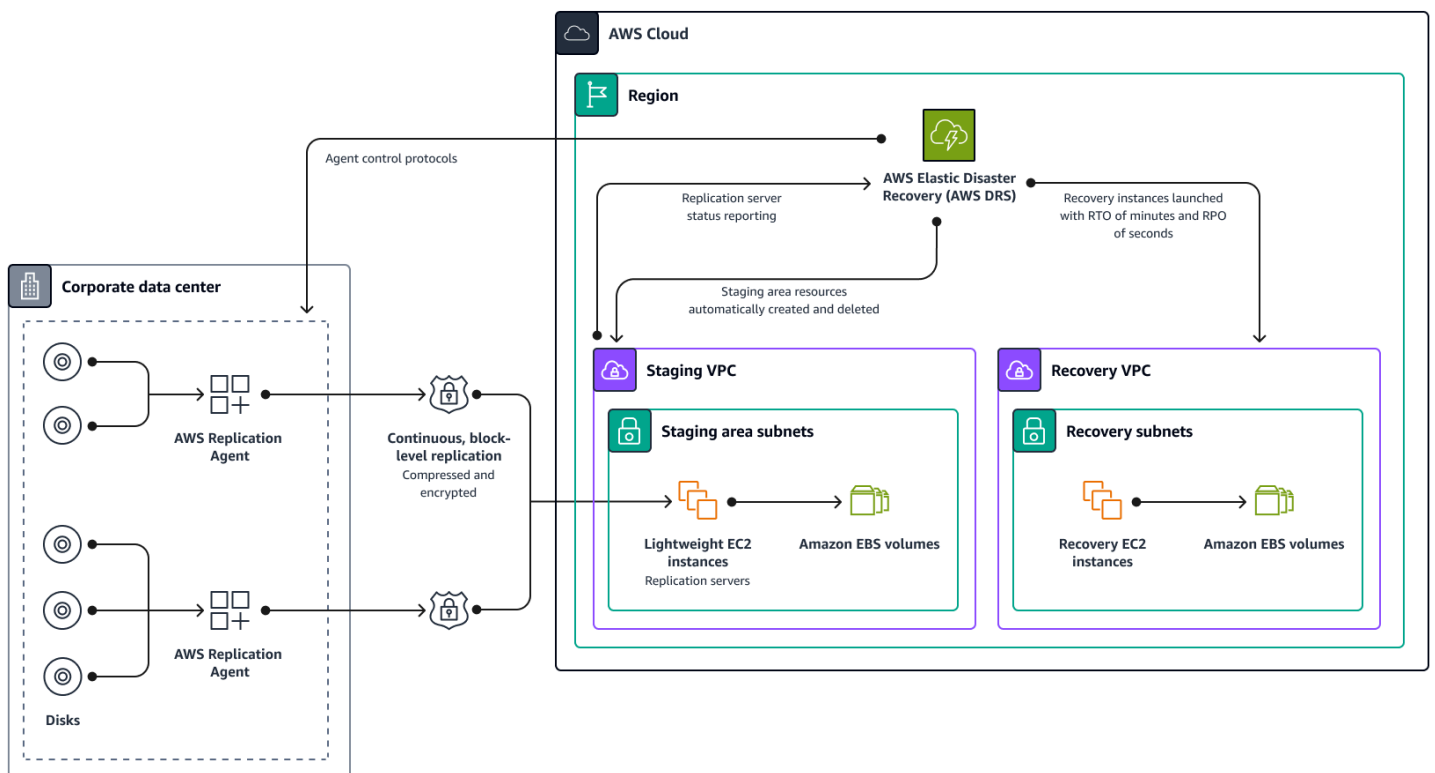
Certaines entreprises doivent s'assurer que toutes les applications métier critiques disposent d'un plan de reprise après sinistre. Par le passé, bon nombre de ces entreprises ont investi massivement dans les solutions traditionnelles de reprise après sinistre, qui nécessitent la préconstruction et la maintenance d'une infrastructure dupliquée complète. Cette approche est coûteuse, prend du temps et est difficile à mettre en œuvre à grande échelle.

Vous pouvez désormais l'utiliser Reprise après sinistre AWS Elastic pour éliminer le besoin de pré-construire une infrastructure de reprise après sinistre. Les machines de reprise après sinistre ne sont démarrées dans Elastic Disaster Recovery que lorsque cela est nécessaire. Vous ne payez donc que ce que vous utilisez lorsque vous en avez besoin. Cela signifie que vous pouvez réduire de manière significative vos licences logicielles et vos coûts de calcul de haute performance.

En outre, la zone de transit de la solution de reprise après sinistre contient des volumes Amazon Elastic Block Store (Amazon EBS) à faible coût. Les volumes EBS réduisent encore le coût de provisionnement des ressources dupliquées. Cela vous permet de réduire vos coûts globaux de reprise après sinistre tout en conservant une solution de reprise après sinistre robuste et fiable qui répond aux exigences de votre entreprise. Vous pouvez utiliser Elastic Disaster Recovery pour vous concentrer sur vos activités principales, tout en AWS prenant en charge l'infrastructure sous-jacente de votre solution de reprise après sinistre.

Pour SQL Server, vous pouvez utiliser Elastic Disaster Recovery comme option de reprise après sinistre rentable. La licence pour le nœud passif dans une architecture SQL Server hautement disponible et tolérante aux pannes est couverte si vous utilisez Active Software Assurance. Cependant, vous devez toujours payer des frais de calcul pour que le serveur passif soit en ligne. Avec Elastic Disaster Recovery, le serveur principal peut effectuer une réplication vers l'environnement de reprise après sinistre sans avoir à maintenir une assurance logicielle active et sans avoir à payer les coûts de calcul liés à la reprise après sinistre. Cette combinaison d'économies peut réduire les coûts de reprise après sinistre de SQL Server de 50 % ou plus.

Le schéma suivant montre un exemple d'architecture pour une solution basée sur Elastic Disaster Recovery.



Pour plus d'informations, consultez [Comment configurer la haute disponibilité pour SQL Server Reprise après sinistre AWS Elastic sur le site DR restauré à l'aide du AWS blog Microsoft Workloads on.](#)

Comparaison des coûts

Le tableau suivant compare les coûts des HA/DR solutions abordées dans cette section. Les hypothèses suivantes sont formulées aux fins de cette comparaison :

- Type d'instance : r5d.xlarge
- Type de licence : licence incluse pour Windows et SQL Server
- Région — us-east-1

Solution	Haute disponibilité	Reprise après sinistre	Édition Enterprise	Édition Standard	Cost
Expédition de journaux	Non	Oui	Oui	Oui	Édition SQL Server Enterprise : 32 674,8\$ (2 nœuds) Édition standard de SQL Server : 14 804,4\$ (2 nœuds)
Groupes de disponibilité Always On	Oui	Oui	Oui	Oui, mais les groupes de disponibilité de base (2 nœuds)	Édition SQL Server Enterprise : 32 674,8\$ (2 nœuds) Édition standard de SQL Server : 14 804,4\$ (2 nœuds)
Toujours allumé FCIs	Oui	Non	Oui	Oui (2 nœuds)	Édition standard de SQL Server : 14 804,4\$

Solution	Haute disponibilité	Reprise après sinistre	Édition Enterprise	Édition Standard	Cost
Groupes de disponibilité distribués	Oui	Oui	Oui	Non	Édition SQL Server Enterprise : 65 349,6\$ (4 nœuds)

Solution	Haute disponibilité	Reprise après sinistre	Édition Enterprise	Édition Standard	Cost
Reprise après sinistre Elastic	Non	Oui	Oui	Oui	<p>Environ 107,48 \$/mois pour la répliation d'une instance et de 1 To de stockage</p> <p>Remarque : Elastic Disaster Recovery est facturé à l'heure, par serveur de répliation. Le coût est le même, quels que soient le nombre de disques, la taille du stockage, le nombre de lancements d'exploration ou de restauration, ou la région que vous répliquez.</p>

Solution	Haute disponibilité	Reprise après sinistre	Édition Enterprise	Édition Standard	Cost
Gardien de données SIOS	Oui	Oui	Oui	Oui	<p>Groupes de disponibilité Always On avec Software Assurance (2 nœuds, 24 cœurs) : 213 480\$</p> <p>Cluster SQL Server à 2 nœuds exécuté sur l'édition SQL Server Standard avec SIOS DataKeeper et Software Assurance : 61 530\$ (2 nœuds)</p>
AWS DMS	Non	Oui	Oui	Oui	745,38 \$/ mois pour une instance r5.xlarge et 1 To de stockage

Recommandations en matière d'optimisation des coûts

Nous vous recommandons de suivre les étapes suivantes pour choisir une HA/DR solution répondant aux exigences de votre organisation :

- Consultez la section [Sélectionnez l'instance EC2 adaptée aux charges de travail SQL Server](#) de ce guide.
- Déterminez les besoins en IOPS et en débit de vos charges de travail en exécutant des compteurs de performance pendant les pics de charge de travail :
 - IOPS = disque reads/sec + écritures sur disque par seconde
 - Débit = lecture sur bytes/sec disque+octets d'écriture sur disque par seconde
- Utilisez les types de volumes de stockage suivants pour améliorer les performances et réaliser des économies :
 - NVMe stockage d'instance pour extension tempdb du pool de mémoire tampon
 - volumes io2 pour fichiers de base de données
- [AWS Trusted Advisor](#) À utiliser pour des recommandations sur l'optimisation des coûts pour SQL Server sur Amazon EC2. Il n'est pas nécessaire d'installer un agent pour Trusted Advisor effectuer des vérifications d'optimisation de SQL Server. Trusted Advisor inspecte les configurations d'instance incluses dans la licence Amazon EC2 SQL Server, telles que CPUs virtual (CPUsv), version et édition. Ensuite, Trusted Advisor formule des recommandations basées sur les meilleures pratiques.
- Optimiseur de calcul AWS À utiliser à la fois pour les recommandations relatives à l'instance Amazon EC2 et à la bonne taille d'Amazon EBS.
- [Calculateur de tarification AWS](#) À utiliser pour concevoir votre HA/DR stratégie d'estimation des coûts.
- Pour déterminer s'il est possible de passer de l'édition SQL Server Enterprise à l'édition SQL Server Standard, utilisez la vue de gestion dynamique [sys dm_db_persisted_sku_features pour identifier les fonctionnalités](#) spécifiques à l'édition qui sont actives dans la base de données actuelle.

Note

Side-by-side les migrations sont nécessaires pour modifier l'édition de SQL Server lorsque vous utilisez des instances EC2 incluses dans une licence.

- Effectuez des exercices de reprise après sinistre semestriels ou annuels afin de mieux concevoir une conception capable de restaurer la base de données avec un RTO et un RPO définis. Cela peut également vous aider à identifier les faiblesses de l'architecture.

Ressources supplémentaires

- [Simplifiez vos déploiements de haute disponibilité de Microsoft SQL Server à l'aide d'Amazon FSx pour Windows File Server](#) (AWS Storage Blog)
- [Notes de terrain : Création d'une architecture multirégionale pour SQL Server à l'aide de FCI et de groupes de disponibilité distribués](#) (blog d'AWS architecture)
- [Concevez une reprise après sinistre pour SQL Server sur AWS : Partie 1](#) (blog AWS de base de données)
- [Haute disponibilité de Microsoft SQL avec Amazon FSx pour Windows](#) (YouTube)
- [Optimisation des performances de Microsoft SQL Server avec Amazon EBS](#) (AWS Storage Blog)
- [Comparaison de vos modèles de stockage sur site avec les services AWS de AWS stockage](#) (Storage Blog)
- [Planification du remplacement d'un NAS de centre de données par Amazon FSx File Gateway](#) (AWS Storage Blog)
- [Optimisation des coûts pour vos déploiements SQL Server à haute disponibilité sur AWS](#) (AWS Storage Blog)
- [Comment configurer la reprise après sinistre pour les groupes de disponibilité SQL Server Always On à l'aide de Reprise après sinistre AWS Elastic](#)(Microsoft Workloads activé AWS)
- [Comment configurer la haute disponibilité pour SQL Server sur un site de reprise après sinistre restauré à l'aide de Reprise après sinistre AWS Elastic](#)(Microsoft Workloads activé AWS)

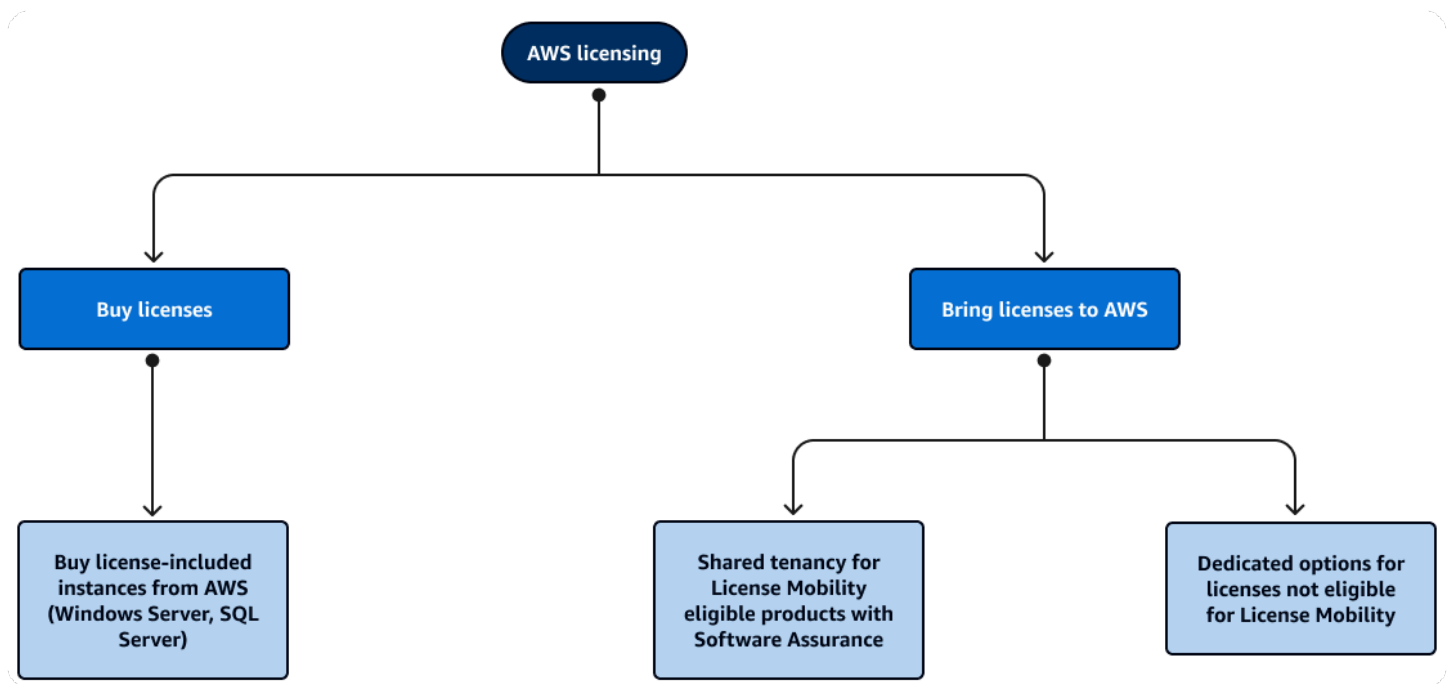
Comprendre les licences SQL Server

Présentation de

Alors que de plus en plus d'entreprises transfèrent leurs charges de travail vers le cloud, l'optimisation des coûts sur les plateformes cloud est devenue une priorité absolue. Les licences constituent l'un des coûts les plus importants associés à l'exécution des charges de travail Microsoft. AWS Cette section explique comment optimiser les coûts AWS en optimisant les licences Microsoft pour SQL Server.

AWS options de licence

AWS propose une gamme de choix flexibles d'optimisation des coûts pour les licences. Ces options de licence sont conçues pour vous aider à réduire les coûts, à maintenir la conformité et à répondre aux besoins de votre entreprise.



AWS classe les licences en trois catégories principales :

1. Licence incluse — Cette option de licence vous permet d'acheter et d'utiliser des licences à la demande, en ne payant que pour ce que vous utilisez. L'option avec licence incluse est idéale pour les scénarios dans lesquels vous avez besoin de flexibilité dans l'utilisation des licences et souhaitez éviter les coûts initiaux. Vous pouvez choisir parmi une gamme de produits Windows Server, SQL Server et d'autres produits Microsoft.
2. Produits BYOL (Bring Your Own License) avec mobilité des licences : cette option de licence est conçue pour les scénarios dans lesquels vous possédez déjà des licences existantes et souhaitez les utiliser dans le cloud. AWS permet aux clients de transférer leurs propres licences dans le cloud par le biais du programme [License Mobility](#) de Microsoft. Vous pouvez proposer des produits dotés de la mobilité des licences, tels que SQL Server with Software Assurance (SA), dans le cadre d'une location partagée ou dédiée afin de réduire les coûts de vos AWS instances.
3. Produits BYOL sans mobilité de licence : pour les produits Microsoft ne disposant pas de License Mobility, tels que Windows Server, AWS des options dédiées permettant d'utiliser ces produits dans le cloud sont proposées. En outre, les hébergeurs dédiés offrent la possibilité d'octroyer

des licences au niveau du noyau physique. Cela peut vous faire économiser 50 % ou plus sur les licences requises pour exécuter vos charges de travail. Les hébergeurs dédiés constituent une excellente option pour des charges de travail stables et prévisibles exécutées la plupart du temps.

Impact financier de l'acquisition de licences

L'acquisition de licences peut avoir un impact significatif sur le coût d'exécution des charges de travail Microsoft. AWS Si vous apportez vos propres licences, vous n'êtes pas tenu de payer de frais de licence supplémentaires pour les instances exécutées dans le cloud. Cela peut permettre de réaliser d'importantes économies.

La comparaison suivante montre le coût mensuel à la demande lié à l'exécution d'une seule instance c5.xlarge 24 heures sur 24, 7 jours sur 7 :

- Édition Windows Server + SQL Server Enterprise : 1353 \$/mois (licence incluse)
- Édition Windows Server + SQL Server Standard : 609 \$/mois (licence incluse)
- Windows Server uniquement : 259 \$/mois (licence incluse)
- Calcul uniquement (Linux) : 127 \$/mois

En fin de compte, le fait d'apporter vos propres licences peut avoir un impact significatif sur le coût d'exécution des charges de travail Microsoft. AWS Si vous utilisez vos licences existantes, vous pouvez réduire les coûts de licence et économiser de l'argent sur votre AWS facture globale.

Optimisation des licences

Une évaluation de AWS l'optimisation et des licences (AWS OLA) peut vous aider à optimiser vos licences en réduisant les coûts de calcul et de licence. AWS OLA est conçu pour évaluer vos exigences en matière de licences pour les charges de travail en cours d'exécution AWS ou pour les charges de travail dont la migration est prévue. AWS L'OLA fournit des recommandations sur l'optimisation de l'utilisation des licences.

L'une des stratégies clés pour optimiser l'utilisation des licences consiste à [dimensionner correctement les instances](#). Le bon dimensionnement implique de sélectionner le type d'instance adapté à votre charge de travail en fonction de ses besoins en termes de processeur, de mémoire et de stockage. En choisissant la taille d'instance appropriée, vous pouvez vous assurer que vous utilisez les ressources de manière rentable. Cela peut permettre de réaliser d'importantes économies.

Dans le cas des licences logicielles Microsoft, le nombre de cœurs sur lesquels le logiciel s'exécute est un facteur essentiel pour déterminer les coûts de licence. Par exemple, les licences Windows Server et SQL Server sont généralement accordées en fonction du nombre de cœurs. En dimensionnant correctement les instances, vous pouvez réduire le nombre de cœurs sur lesquels s'exécute le logiciel Microsoft et, par conséquent, réduire à la fois le coût de l'instance et le nombre de licences requises.

Recommandations en matière d'optimisation des coûts

L'optimisation des licences est un élément clé de l'optimisation des coûts sur AWS. En mettant en œuvre les bonnes stratégies, vous pouvez réduire les coûts de licence, maintenir la conformité et tirer le meilleur parti possible de votre investissement dans les licences. Cette section décrit plusieurs stratégies d'optimisation des licences.

Apportez vos licences Windows Server éligibles

Apporter vos propres licences Windows Server est l'une des stratégies les plus efficaces pour optimiser les licences. Cette stratégie vous permet de tirer parti de vos investissements existants pour réduire vos AWS dépenses.

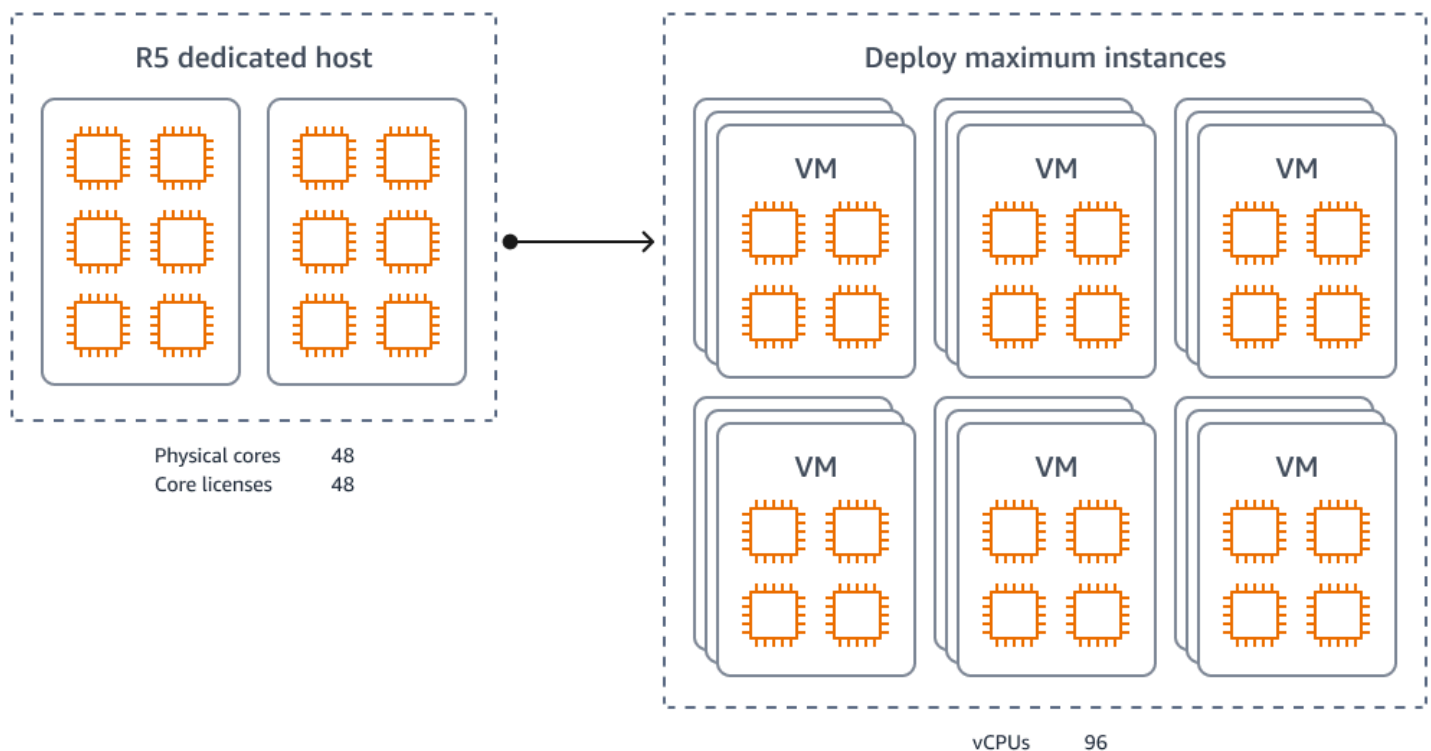
Par exemple, vous pouvez déployer Windows Server 2019 et les versions antérieures sur des [hôtes dédiés Amazon EC2](#) si vous avez acheté les licences avant le 1er octobre 2019 ou si vous les avez achetées dans le cadre de contrats d'entreprise actifs signés avant cette date. Cette règle est basée sur une modification apportée par Microsoft en 2019 à ses conditions générales de licence pour les produits sans licence Mobility, tels que Windows Server, lorsqu'ils sont déployés sur des [fournisseurs répertoriés](#) (par exemple AWS, Alibaba ou Google Cloud). Selon les nouvelles conditions, vous ne pouvez pas apporter vos propres licences Windows Server, AWS mais vous devez utiliser des instances incluses dans des licences. Toutefois, si vous avez acheté des licences perpétuelles avant cette date, vous pouvez toujours déployer ces licences Windows Server sur des hôtes dédiés Amazon EC2.

Licences au niveau physique

Les licences au niveau du cœur physique vous permettent de ne licencier que les cœurs physiques d'un hôte, afin de pouvoir ensuite déployer un nombre maximum d'instances sans affecter le nombre de licences requises. Cela se fait généralement à l'aide de Windows Server Datacenter et de SQL Server Enterprise Edition.

Prenons l'exemple d'un hôte dédié R5 avec 48 cœurs, ce qui correspond à 96 v. CPUs. Si vous utilisez l'édition Windows Server Datacenter, vous n'avez besoin que de 48 licences. Cela vous

permet de déployer une combinaison d'instances jusqu'à 96 VCPUs, comme le montre le schéma suivant.



Cette approche peut être particulièrement rentable si vous disposez de suffisamment de charges de travail pour maximiser le nombre d'instances que vous pouvez exécuter sur un hôte. En octroyant des licences au niveau du cœur physique, vous pouvez éviter des coûts de licence supplémentaires pour chaque instance et obtenir le meilleur rapport qualité-prix possible pour votre investissement en licences.

Licence au niveau physique de base de SQL Server

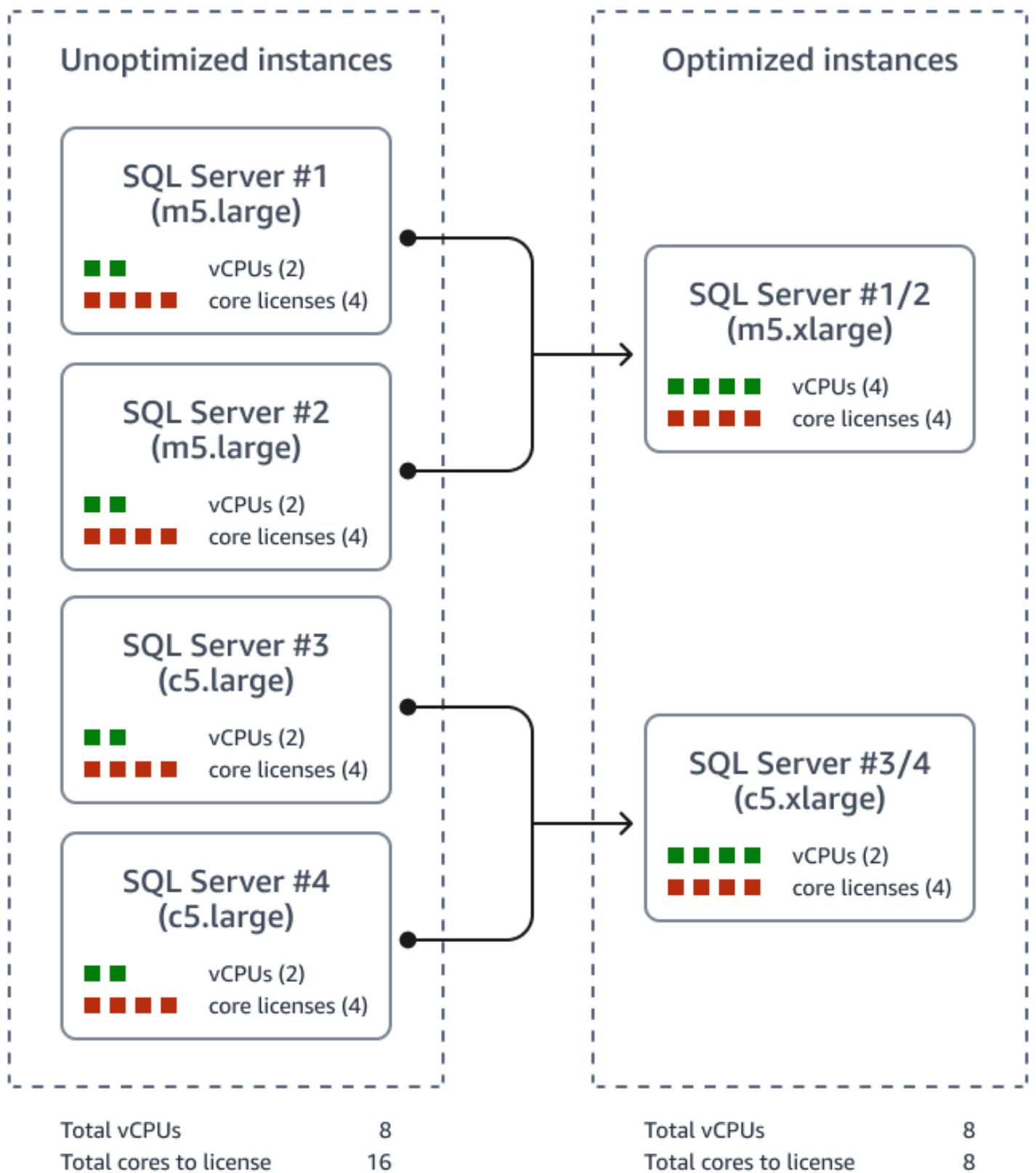
Dans le cadre de la location partagée, les licences SQL Server sont basées sur le nombre de v CPUs alloués à l'instance. En revanche, avec les hôtes dédiés, vous pouvez obtenir une licence pour SQL Server Enterprise Edition au niveau du cœur physique ou au niveau du vCPU.

Comme dans l'exemple précédent de l'hôte dédié R5, si vous octroyez une licence pour l'édition SQL Server Enterprise au niveau du cœur physique, vous n'avez besoin que de 48 licences de l'édition SQL Server Enterprise pour octroyer des licences aux hôtes. En revanche, dans le cadre de la location partagée, où la seule option est d'octroyer une licence par vCPU, vous devez disposer de 96 licences de l'édition Enterprise de SQL Server pour la même charge de travail. Par conséquent, les hôtes dédiés peuvent vous faire économiser jusqu'à 50 % sur les coûts de licence SQL Server

par rapport à la location partagée. Cela permet également de réaliser des économies sur les coûts d'instance en proposant des licences Windows éligibles.

Consolider les instances de SQL Server

La [consolidation de SQL Server](#) est le processus qui consiste à combiner plusieurs instances de SQL Server sur un seul serveur. SQL Server nécessite un minimum de quatre licences principales par instance, même si l'instance ne possède que deux versions CPUs. Cela signifie que l'exécution de SQL Server sur des serveurs dotés de moins de quatre cœurs peut vous amener à surlicencier ces instances et à utiliser plus de licences que nécessaire.



Par exemple, la consolidation de deux instances de deux v CPUs chacune en une seule instance de quatre v CPUs peut réduire les exigences de licence de 50 %. Cela est dû au fait que seules quatre licences principales sont requises au lieu de huit.

Pour plus d'informations sur la consolidation, consultez la section de ce guide consacrée à la [consolidation de SQL Server](#).

Rétrograder les éditions de SQL Server

La [modification des éditions de SQL Server](#) peut constituer une stratégie clé pour optimiser l'utilisation des licences et réduire les coûts. L'édition Enterprise de SQL Server étant considérablement plus chère que l'édition Standard, la rétrogradation peut entraîner des économies importantes.

Le chiffrement transparent des données (TDE) et les groupes de disponibilité Always On sont deux fonctionnalités populaires de l'édition Enterprise de SQL Server. Il existe toutefois des alternatives économiques à ces fonctionnalités que vous pouvez envisager si vous n'avez pas besoin de l'ensemble complet des fonctionnalités de l'édition Enterprise de SQL Server. Par exemple, vous pouvez obtenir le TDE dans l'édition standard de SQL Server à partir de SQL Server 2019. À la place des groupes de disponibilité Always On, vous pouvez utiliser le clustering sur incident avec stockage partagé activé FSx pour Windows File Server afin d'obtenir une haute disponibilité avec l'édition standard de SQL Server.

En rétrogradant de l'édition SQL Server Enterprise à l'édition SQL Server Standard, vous pouvez réduire considérablement vos coûts de licence. Pour plus d'informations, consultez l'article [Optimisation des coûts pour vos déploiements SQL Server à haute disponibilité AWS publié sur le blog sur le AWS stockage](#).

En plus de réduire les coûts de licence, la rétrogradation des éditions de SQL Server peut vous aider à réduire vos dépenses d'assurance logicielle et à éviter de futures erreurs. Si vous remettez les licences non utilisées en rayon, vous pouvez éviter des coûts de licence supplémentaires et tirer le meilleur parti de votre investissement en licences.

Il est important d'évaluer soigneusement vos charges de travail SQL Server et de déterminer quelles fonctionnalités sont essentielles pour répondre aux besoins de votre entreprise. Pour plus d'informations, consultez la section [Évaluation de votre environnement](#) dans AWS Prescriptive Guidance et déterminez si votre base de données Microsoft SQL Server utilise des fonctionnalités spécifiques à l'édition Enterprise de SQL Server.

Si vous choisissez la bonne édition de SQL Server et que vous utilisez des alternatives aux fonctionnalités de l'édition Enterprise de SQL Server, vous pouvez réaliser d'importantes économies tout en garantissant la conformité et en répondant aux besoins de votre entreprise. Pour plus d'informations sur les options de rétrogradation, consultez la section [Comparer les éditions de SQL Server](#) de ce guide.

Utiliser l'édition SQL Server Developer dans des environnements hors production

Dans les environnements hors production, vous pouvez déployer des éditions sous licence de SQL Server, telles que l'édition Enterprise ou Standard, en utilisant des abonnements MSDN dans des environnements sur site. Cependant, l'abonnement MSDN ne dispose pas de License Mobility. Donc, si vous migrez vers AWS, vous ne pouvez pas transférer ces licences. Vous devez plutôt utiliser l'édition SQL Server Developer.

L'édition SQL Server Developer est une édition complète de SQL Server disponible gratuitement. Cette édition est disponible pour les versions 2016 et ultérieures de SQL Server. Vous pouvez le télécharger sur le site Web de Microsoft. L'édition SQL Server Developer est destinée à être utilisée dans tous les environnements non liés à la production, tels que le développement, les tests et les tests, tant qu'elle ne se connecte pas aux données de production en direct.

Si vous utilisez l'édition SQL Server Developer dans des environnements hors production, vous pouvez éviter des coûts de licence supplémentaires. Pour plus d'informations, consultez la section [Evaluate SQL Server Developer edition](#) de ce guide.

Optimisation du processeur pour les charges de travail SQL Server

Dans certains cas, vous pouvez être amené à choisir un type d'instance dont le volume de travail est CPUs supérieur à ce qui est requis pour votre charge de travail en raison d'autres facteurs tels que la RAM ou les limites du réseau. Cependant, AWS fournit une solution pour vous aider à optimiser vos coûts de licence dans ces situations.

Comme la plupart des clients qui proposent des licences SQL Server Core, vous pouvez désactiver l'hyperthreading ou désactiver les processeurs sur l'instance EC2 afin de limiter le nombre de licences disponibles CPUs pour l'hôte. Cette option vous permet de tirer parti d'autres fonctionnalités de l'instance, telles que la RAM, tout en économisant sur le coût d'achat de licences supplémentaires.

Par exemple, si vous déployez une instance r5.4xlarge parce que votre charge de travail nécessite 128 Go de mémoire mais que vous n'avez besoin que de huit cœurs de SQL Server, vous pouvez désactiver l'hyperthreading d'une instance dont seulement huit sont actifs. CPUs Vous pouvez ainsi

économiser 50 % sur les licences SQL Server requises, car il vous suffit de licencier les huit cœurs activement utilisés.

Type d'instance	Total v CPUs	vCPU actif avec fonction d'optimisation CPUs	Économies de licence SQL Server
r5.4xlarge	16	8	50%
r5.12xlarge	48	8	83 %

La fonctionnalité Optimize CPU peut être configurée lors de la configuration de lancement d'Amazon EC2 ou en modifiant une instance existante. Il peut également être appliqué à la fois aux instances BYOL et aux instances Amazon EC2 incluses dans une licence. Cette flexibilité vous aide à adapter votre processeur aux besoins de votre charge de travail, tout en réduisant le nombre Windows Server de SQL Server licences. Pour les instances Amazon EC2 incluses dans une licence, la CPUs réduction permet de réaliser des économies instantanées sur les coûts de licence.

Si vous dimensionnez correctement vos instances, vous pouvez vous assurer d'utiliser les types d'instances les plus économiques pour vos charges de travail. Au fur AWS et à mesure de l'introduction de nouveaux types d'instances, il est important d'évaluer si ces nouvelles instances peuvent répondre aux exigences de charge de travail avec moins de cœurs.

Ressources supplémentaires

- [Amazon Web Services et Microsoft : questions fréquemment posées](#) (AWS documentation)

Sélectionnez l'instance EC2 adaptée aux charges de travail SQL Server

Important

Avant de lire cette section, nous vous recommandons de lire d'abord les sections [Comprendre les licences SQL Server](#) et [Sélectionner le type d'instance adapté aux charges de travail Windows](#) de ce guide.

Présentation de

Microsoft SQL Server fonctionne sur des instances Amazon Elastic Compute Cloud (Amazon EC2) depuis plus de 15 ans. AWS a tiré parti de cette expérience pour développer des instances Amazon EC2 adaptées aux charges de travail SQL Server, qu'il s'agisse de spécifications minimales ou de clusters multirégionaux à hautes performances.

Le choix de l'instance EC2 appropriée pour SQL Server dépend largement de votre charge de travail. Comprendre comment SQL Server est licencié, comment il utilise la mémoire et comment les fonctionnalités de SQL Server s'alignent sur les offres Amazon EC2 peut vous aider à trouver l'instance EC2 la mieux adaptée à votre application.

Cette section aborde diverses charges de travail SQL Server et explique comment elles peuvent être associées à certaines instances EC2 afin de minimiser vos coûts de licence et de calcul.

Comparaison des coûts

Amazon EC2 vous permet d'apporter votre propre licence (BYOL) ou de payer au fur et à mesure avec les licences Windows Server et SQL Server. Pour les pay-as-you-go licences, les coûts de licence pour les licences Windows Server et SQL Server sont intégrés au coût horaire de l'instance EC2. Par exemple, vous pouvez avoir AMIs des prix différents. Le prix de l'AMI dépend de l'édition SQL Server sur laquelle l'AMI est exécutée.

Les tarifs de Windows Server et SQL Server ne sont pas détaillés. Vous ne trouverez pas de prix détaillés sur des outils tels que le [Calculateur de tarification AWS](#). Si vous sélectionnez différentes combinaisons d'offres avec licence, les coûts de licence peuvent être déduits, comme le montre le tableau suivant.

instance EC2	AMI	Calculez le prix	Prix de la licence Windows	Prix de la licence SQL	Prix total
r5.xlarge	Linux (tarification du calcul)	183,96\$	-	-	183,96\$
r5.xlarge	Développeur Linux + SQL	183,96\$	\$0	\$0	183,96\$

instance EC2	AMI	Calculez le prix	Prix de la licence Windows	Prix de la licence SQL	Prix total
r5.xlarge	Windows Server (LI)	183,96\$	134,32\$	-	318,28\$
r5.xlarge	Développeur SQL pour Windows +	183,96\$	134,32\$	\$0	318,28\$
r5.xlarge	Windows + SQL Web (LI)	183,96\$	134,32\$	49,64\$	367,92\$
r5.xlarge	Windows + SQL Standard (LI)	183,96\$	134,32\$	350,4\$	668,68\$
r5.xlarge	Windows + SQL Entreprise (LI)	183,96\$	134,32\$	1095\$	1413,28\$

Note

Les prix indiqués dans le tableau précédent sont basés sur les tarifs à la demande dans la us-east-1 région.

La méthode la plus rentable pour exécuter SQL Server consiste à utiliser une édition de niveau inférieur jusqu'à ce que vous ayez besoin d'une fonctionnalité d'une édition de niveau supérieur. Pour plus d'informations, consultez la section [Comparaison des éditions de SQL Server](#) de ce guide. La mise à niveau de SQL Server Web Edition vers SQL Server Standard coûte plus de sept fois le coût de licence SQL Server et plus de trois fois le coût du passage de l'édition Standard à l'édition Enterprise. La disparité des coûts de licence est un facteur important à prendre en compte et est explorée dans le reste de cette section.

Scénario d'optimisation des coûts

Prenons l'exemple d'un scénario dans lequel une société d'analyse effectuant le suivi des véhicules de livraison cherche à améliorer les performances de SQL Server. Après qu'un expert MACO a examiné les problèmes de performance de l'entreprise, celle-ci passe des instances x1e.2xlarge aux instances x2iedn.xlarge. Bien que la taille de l'instance soit plus petite, les améliorations apportées aux instances x2 améliorent les performances et l'optimisation de SQL Server en utilisant des extensions de pool de mémoire tampon. Cela a permis à l'entreprise de passer de l'édition SQL Server Enterprise à l'édition SQL Server Standard et de réduire ses licences SQL Server de 8 CPUs à 4 versions. CPUs

Avant l'optimisation :

Serveur	instance EC2	Edition SQL Server	Coût mensuel
Prod DB1	x1e.2xlarge	Enterprise	3 918,64\$
Prod DB2	x1e.2xlarge	Enterprise	3 918,64\$
Total			7 837,28\$

Après optimisation :

Serveur	instance EC2	Edition SQL Server	Coût mensuel
Prod DB1	x2iedn.xlarge	Standard	1 215,00\$
Prod DB2	x2iedn.xlarge	Standard	1 215,00\$
Total			2 430,00\$

Les modifications combinées des instances x1e.2xlarge aux instances x2iedn.xlarge ont permis au client exemple d'économiser 5 407\$ par mois sur ses serveurs de base de données de production. Cela a permis de réduire le coût total de la charge de travail de 69 %.

Note

Les prix indiqués dans le tableau précédent sont basés sur les tarifs à la demande dans la us-east-1 région.

Recommandations en matière d'optimisation des coûts

instances de mémoire optimisée

L'un des aspects les plus importants de SQL Server est de comprendre sa dépendance à la mémoire. SQL Server essaie d'utiliser toute la RAM disponible qui n'est pas utilisée par le système d'exploitation (jusqu'à 2 To pour une installation par défaut). Il le fait pour des raisons de performance. Travailler avec des données en mémoire est bien plus performant que de devoir constamment extraire des données du disque, apporter des modifications, puis les réécrire sur le disque. SQL Server essaie plutôt de charger autant de données que possible à partir des bases de données jointes et conserve ces données dans la RAM. Les modifications apportées aux données sont enregistrées en mémoire et sont ensuite enregistrées sur disque.

Note

Pour une explication détaillée de la façon dont SQL Server écrit les modifications, consultez la section [Écrire des pages](#) dans la documentation Microsoft.

Étant donné que SQL Server fonctionne mieux avec de plus grandes quantités de RAM, nous recommandons généralement de commencer par les types d'instances [optimisés pour la mémoire Amazon EC2](#). Les instances optimisées pour la mémoire sont polyvalentes et offrent de nombreuses options différentes. La famille R a un vCPU-to-RAM ratio de 1 pour 8 et propose des options pour les processeurs Intel, les processeurs AMD, une mise en réseau améliorée, des performances EBS améliorées, le stockage d'instance et une vitesse de processeur améliorée. Pour les charges de travail gourmandes en mémoire, il existe également une gamme X qui combine bon nombre des mêmes options et étend le vCPU-to-RAM ratio de 1 à 32. Grâce à la polyvalence des instances optimisées pour la mémoire, vous pouvez les appliquer à des charges de travail SQL Server de toutes formes et de toutes tailles.

Charges de travail inférieures au minimum de ressources (moins de 4 VCPUs)

Bien que certains cas d'utilisation fonctionnent bien avec les instances burstables (T3), nous vous recommandons généralement d'éviter d'utiliser des instances burstables pour les charges de travail SQL Server. La licence pour SQL Server est basée sur le nombre de v CPUs attribués à une instance. Si SQL Server est inactif la majeure partie de la journée et qu'il obtient des crédits en rafale, vous payez pour des licences SQL que vous n'utilisez pas pleinement. En outre, SQL Server requiert une licence minimale de 4 cœurs par serveur. Cela signifie que si votre charge de travail SQL Server ne nécessite pas une puissance CPUs de calcul de 4 V, vous payez une licence SQL Server que vous n'utilisez pas. Dans ces scénarios, il serait préférable de [consolider plusieurs instances de SQL Server](#) sur un serveur plus grand.

Charges de travail utilisant un minimum de ressources (moins de 64 Go de RAM)

De nombreuses charges de travail SQL Server inférieures à 64 Go de RAM ne privilégient ni les performances ni la haute disponibilité. Pour ces types de charges de travail, l'édition Web de SQL Server peut être adaptée si l'application est couverte par les restrictions de licence de Microsoft.

Important

L'édition Web de SQL Server comporte un cas d'utilisation restreint basé sur les conditions de licence de Microsoft. L'édition SQL Server Web peut prendre en charge uniquement les pages Web, les sites Web, les applications Web et les services Web bénéficiant d'un accès public et Internet. Il ne peut pas être utilisé pour prendre en charge line-of-business des applications (par exemple, la gestion de la relation client, la gestion des ressources d'entreprise et d'autres applications similaires).

L'édition Web de SQL Server peut atteindre 32 V CPUs et 64 Go de RAM et est 86 % moins chère que l'édition Standard de SQL Server. Pour les charges de travail nécessitant peu de ressources, l'utilisation d'une instance AMD optimisée pour la mémoire telle que le modèle r6a, dont le prix de calcul est inférieur de 10 % à celui de son homologue Intel, est également un bon moyen de réduire au minimum les coûts de calcul et de licence SQL.

Charges de travail avec des ressources moyennes (moins de 128 Go de RAM)

L'édition Standard de SQL Server est utilisée pour la majorité des charges de travail SQL Server jusqu'à 128 Go de RAM. L'édition Standard de SQL Server est 65 à 75 % moins chère que l'édition

Enterprise de SQL Server et peut évoluer jusqu'à 48 vCPU avec 128 Go de RAM. Étant donné que la limite de 128 Go de RAM est généralement atteinte avant la limite de 48 vCPU, c'est l'objectif de la plupart des clients qui souhaitent éviter de passer à l'édition Enterprise de SQL Server.

SQL Server possède une fonctionnalité appelée [extension du pool de mémoire tampon](#). Cette fonctionnalité permet à SQL Server d'utiliser une partie d'un disque pour agir comme une extension de la RAM. L'extension de pool de mémoire tampon fonctionne bien lorsqu'elle est associée à un stockage ultrarapide, comme celui NVMe SSDs utilisé dans le stockage [d'instances Amazon EC2](#). Les instances Amazon EC2 contenant un stockage d'instance sont signalées par un « d » dans le nom de l'instance (par exemple, r5d, r6id et x2iedn).

Les extensions de pool de mémoire tampon ne remplacent pas la RAM normale. Toutefois, si vous avez besoin de plus de 128 Go de RAM, vous pouvez utiliser des extensions de pool de mémoire tampon avec des instances EC2 telles que r6id.4xlarge et x2iedn.xlarge pour retarder la mise à niveau vers les licences Enterprise Edition.

Charges de travail à hautes performances (plus de 128 Go de RAM)

Les charges de travail SQL Server nécessitant des performances élevées sont difficiles à optimiser en termes d'optimisation des coûts en raison de leur dépendance à de nombreuses ressources. Cependant, comprendre les différences entre les instances EC2 peut vous éviter de faire le mauvais choix.

Le tableau suivant présente diverses instances EC2 optimisées pour la mémoire et leurs limites de performances.

	r5b	r6idn	r7iz	x2iden	x2iezn
Processeur	3.1 GHz	3,5 GHz	3.9 GHz	3,5 GHz	4,5 GHz
	Processeur Intel Xeon de 2e génération	Processeur Intel Xeon de 3e génération	Processeur évolutif Intel Xeon de 4e génération	Processeur Intel Xeon de 3e génération	Processeur Intel Xeon de 2e génération
Rapport CPU:RAM	1:8	1:8	1:8	1:32	1:32

	r5b	r6idn	r7iz	x2iden	x2iezn
Nombre maximal de vCPU	96	128	128	128	48
RAM maximale	768 GO	1 024 GO	1 024 GO	4 096 GO	1 536 GO
Stockage d'instances	–	NVMe SSD (4 x 1900 Go)	–	NVMe SSD (2 x 1900 Go)	–
io2 Block Express	Pris en charge	Pris en charge	Pris en charge	Pris en charge	–
Nombre maximal d'IOPS EBS	260 000	350 000	160 000	260 000	80 000
Débit EBS maximal	60 Gbit/s	80 Gbit/s	40 Gbit/s	80 Gbit/s	19 Gbit/s
Bande passante réseau maximale	25 Gb/s	200 Gb/s	50 Gb/s	100 Gbit/s	100 Gbit/s

Chaque instance est utilisée dans un but différent. Comprendre votre charge de travail SQL Server peut vous aider à choisir le type d'instance qui vous convient le mieux.

Détails sur les attributs :

- r5b — L'attribut « b » dans r5b signifie que ce type d'instance est axé sur les performances EBS élevées. Dans la cinquième génération d'instances optimisées pour la mémoire, le r5b était le choix préféré. Il s'agit du premier type d'instance à utiliser les volumes io2 Block Express et à atteindre un maximum d'IOPS de stockage de 260 000. Le type d'instance r5b reste une alternative rentable pour les besoins de performances EBS élevés.

- **r6idn** — La sixième génération d'instances optimisées pour la mémoire offre des améliorations considérables par rapport à la génération précédente. Les améliorations de performances de l'EBS par rapport au r5b vont encore plus loin avec le r6idn, portant le maximum d'IOPS à 350 000. Le r6idn dispose également d'un volume de stockage d'instance pour les extensions tempdb et de pool de mémoire tampon afin d'améliorer encore les performances de SQL Server.
- **x2iedn** — Le x2iedn est similaire au r6idn. Il offre des niveaux similaires d'EBS amélioré, de mise en réseau améliorée et de stockage d'instance NVMe SSD, mais avec un vCPU-to-RAM ratio de 1:32 pour les charges de travail élevées en mémoire et les faibles quantités de processeur (coûts de licence SQL Server inférieurs).
- **x2iezn** — L'attribut « z » dans x2iezn indique que ce type d'instance est axé sur les performances élevées du processeur. Le processeur Cascade Lake possède une fréquence turbo tout-cœur allant jusqu'à 4,5 GHz. Nous vous recommandons d'utiliser cette instance EC2, associée à un vCPU-to-RAM ratio de 1:32, dans un scénario où vous souhaitez limiter la quantité de vCPU. Cela permet à son tour de maintenir les coûts de licence SQL Server à un faible niveau.
- **r7iz** — L'attribut « z » dans r7iz indique que ce type d'instance est axé sur les performances élevées du processeur. Le processeur Sapphire Rapids a une fréquence turbo tout-cœur allant jusqu'à 3,9 GHz. Comme les instances x2iezn, le r7iz donne la priorité aux performances des processeurs haute fréquence, mais avec un ratio de 1:8. vCPU-to-RAM

Ressources supplémentaires

- [Instances Amazon EC2 à usage général \(documentation\)](#) AWS
- [Outil de comparaison](#) (Vantage)
- [Licences — SQL Server](#) (AWS documentation)

Consolider les instances

Cette section se concentre sur la technique d'optimisation des coûts qui consiste à combiner plusieurs instances de SQL Server sur le même serveur afin de minimiser les coûts de licence et de maximiser l'utilisation des ressources.

Présentation de

La création d'une instance fait partie du processus d'installation du moteur de base de données SQL Server. L'instance SQL Server est une installation complète contenant ses propres fichiers

serveur, ses identifiants de sécurité et ses propres bases de données système (master, model, msdb et tempdb). Comme une instance possède ses propres fichiers et services, vous pouvez installer plusieurs instances de SQL Server sur le même système d'exploitation sans que les instances n'interfèrent les unes avec les autres. Cependant, étant donné que les instances sont toutes installées sur le même serveur, elles partagent toutes les mêmes ressources matérielles, telles que le calcul, la mémoire et le réseau.

Il est courant de n'utiliser qu'une seule instance SQL Server par serveur dans les environnements de production afin qu'une instance « occupée » n'abuse pas des ressources matérielles partagées. Il est préférable de doter chaque instance de SQL Server de son propre système d'exploitation, avec ses propres ressources, plutôt que de se fier à la gouvernance des ressources. Cela est particulièrement vrai pour les charges de travail SQL Server à hautes performances qui nécessitent de grandes quantités de RAM et de ressources CPU.

Cependant, les charges de travail SQL Server n'utilisent pas toutes une grande quantité de ressources. Par exemple, certaines organisations attribuent à chacun de leurs clients sa propre instance SQL Server dédiée à des fins de conformité ou de sécurité. Pour les petits clients ou les clients qui ne sont généralement pas actifs, cela signifie qu'il faut exécuter les instances de SQL Server avec un minimum de ressources.

Comme indiqué dans le [guide Microsoft SQL Server 2019 : Licensing](#), chaque serveur exécutant SQL Server doit disposer d'un minimum de quatre licences de processeur. Cela signifie que même si vous exécutez un serveur avec seulement deux vCPUs, vous devez toujours obtenir une licence SQL Server pour quatre CPUs v. Sur la base [de la tarification publique de Microsoft pour SQL Server](#), cela représente une différence de 3 945\$ si vous utilisez l'édition standard de SQL Server. Pour les entreprises qui exploitent plusieurs serveurs avec des instances uniques de SQL Server utilisant un minimum de ressources, le coût combiné de la licence des ressources non utilisées peut être substantiel.

Scénario d'optimisation des coûts

Cette section explore un exemple de scénario qui compare la différence entre l'exécution de quatre serveurs Windows Server, chacun doté d'une seule instance SQL Server, et l'exécution simultanée de plusieurs instances de SQL Server par un seul serveur Windows Server de plus grande taille.

Si chaque instance SQL Server n'a besoin que de deux v CPUs et de 8 Go de RAM, le coût total par serveur est de 7 890\$ pour la licence SQL Server, en plus d'un coût de calcul horaire de 0,096\$.

instance EC2	v CPUs	RAM	Price	v CPUs vers la licence	Coût total des licences SQL Server
m6i.large	2	8	0,096	4	7 890\$

En l'étendant à quatre serveurs, le coût total de la licence SQL Server est de 31 560 dollars, avec un coût de calcul horaire de 0,384 dollars.

instance EC2	v CPUs	RAM	Price	v CPUs vers la licence	Coût total des licences SQL Server
4 x 6 mm de large	2	32	0,384	16	31 560\$

Si vous combinez les quatre instances SQL Server en une seule instance EC2, la quantité totale de ressources de calcul et de calcul reste la même. Toutefois, en supprimant les coûts de licence inutiles de SQL Server, vous pouvez réduire le coût total d'exécution de la charge de travail de 15 780\$.

instance EC2	v CPUs	RAM	Price	v CPUs vers la licence	Coût total des licences SQL Server
m6i.2xlarge	8	32	0,384	8	15 780\$

Note

Dans les tableaux précédents, les coûts de calcul indiquent la tarification horaire à la demande pour les serveurs Amazon EC2 exécutant Windows Server dans la us-east-1 région. Les coûts de licence de SQL Server Standard Edition se réfèrent à la [tarification publique de Microsoft pour SQL Server](#).

Recommandations en matière d'optimisation des coûts

Si vous envisagez de consolider des instances SQL Server, la principale préoccupation est la consommation de ressources pour chacune des instances que vous souhaitez consolider. Il est important d'obtenir des indicateurs de performance sur de longues périodes afin de mieux comprendre les modèles de charge de travail sur chaque serveur. [Amazon CloudWatch](#), [Windows Performance Monitor](#) (perfmon) et les outils de surveillance [natifs de SQL Server sont parmi les outils courants de surveillance](#) de la consommation des ressources.

Nous vous recommandons de vous poser les questions suivantes lorsque vous déterminez si vos charges de travail SQL Server peuvent être combinées pour utiliser les mêmes ressources du serveur sans qu'elles interfèrent les unes avec les autres :

- Quelles ressources (processeur, mémoire et bande passante réseau) sont consommées pendant votre période d'équilibre ?
- Quelles ressources (processeur, mémoire et bande passante réseau) sont consommées pendant les pics ?
- À quelle fréquence les pics se produisent-ils ? Les pics sont-ils constants ?
- Les pics de ressources d'un serveur coïncident-ils avec les pics de ressources d'un autre serveur ?
- Quels sont les IOPS et le débit de stockage utilisés par SQL Server ?

Si vous souhaitez aller de l'avant avec un plan de combinaison d'instances SQL Server, consultez l'article [Exécuter plusieurs instances de SQL Server sur une instance Amazon EC2](#) sur le blog AWS Cloud Operations & Migrations. Ce billet fournit des instructions sur la façon d'apporter les modifications de configuration dans SQL Server pour ajouter des instances supplémentaires. Avant de commencer, considérez les différences mineures lorsque plusieurs instances sont installées sur le même serveur :

- L'instance de base de données SQL Server par défaut est nommée MSSQLSERVER et utilise le port 1433.
- Chaque instance supplémentaire installée sur le même serveur est une instance de base de données « nommée ».
- Chaque instance nommée possède un nom d'instance et un port uniques.
- Le [navigateur SQL Server](#) doit s'exécuter pour coordonner le trafic vers les instances nommées.
- Chaque instance peut utiliser des emplacements distincts pour les fichiers de données de base de données et des connexions distinctes.

- Les [paramètres de mémoire maximale du serveur](#) SQL Server doivent être configurés en fonction des besoins de performance de chaque instance, leur total combiné laissant également suffisamment de mémoire pour le système d'exploitation sous-jacent.
- Vous pouvez utiliser les fonctionnalités [natives de sauvegarde et de restauration](#) de SQL Server ou [AWS DMS](#) pour la migration ou la consolidation.

Ressources supplémentaires

- [Fiche technique des licences SQL Server](#) (blog AWS sur les opérations et les migrations dans le cloud)
- Article de blog sur la [configuration d'instances multiples de SQL Server](#) (blog AWS sur les opérations et les migrations dans le cloud)

Comparez les éditions de SQL Server

Présentation de

Les licences Microsoft SQL Server constituent l'une des dépenses les plus importantes pour un environnement de charge de travail Windows. Les coûts de licence pour SQL Server peuvent facilement dépasser les coûts de calcul liés à l'exécution de la charge de travail. Si vous choisissez la mauvaise édition, vous pourriez payer pour des fonctionnalités que vous n'utilisez pas ou dont vous n'avez même pas besoin. Cette section compare les éditions de SQL Server suivantes, notamment leurs fonctionnalités et leurs coûts relatifs :

- **Enterprise** — L'édition SQL Server Enterprise fournit des fonctionnalités de centre de données avec des performances élevées, une virtualisation illimitée et plusieurs outils de business intelligence (BI).
- **Standard** — L'édition standard de SQL Server fournit une gestion des données de base et des informations commerciales aux petites entreprises et aux services.
- **Web** — L'édition Web de SQL Server convient aux entreprises hébergeant des sites Web ou fournisseurs de valeur ajoutée sur le Web (VAPs). Cette édition offre un faible coût total de possession et fournit des fonctionnalités d'évolutivité et de gérabilité pour les propriétés Web de petite à grande échelle.

⚠ Important

Vous pouvez utiliser l'édition Web de SQL Server pour prendre en charge uniquement les pages Web, les sites Web, les applications Web et les services Web publics et accessibles à Internet. Vous ne pouvez pas utiliser l'édition Web de SQL Server pour prendre en charge line-of-business des applications (telles que les applications de gestion de la relation client ou de gestion des ressources d'entreprise).

- **Développeur** : l'édition SQL Server Developer inclut toutes les fonctionnalités de l'édition Enterprise, mais elle est uniquement destinée au développement.
- **Express** — L'édition SQL Server Express est une base de données gratuite qui peut être utilisée pour apprendre ou pour créer des applications de bureau. Vous pouvez mettre à jour l'édition Express vers d'autres éditions.

📘 Note

L'édition d'évaluation de SQL Server est disponible pour une période d'essai de 180 jours.


Impact sur les coûts

Vous pouvez acheter des licences SQL Server auprès d'un revendeur Microsoft et les transférer AWS à Software Assurance. Vous pouvez également utiliser des licences SQL Server avec un pay-as-you-go modèle incluant une licence Amazon EC2. AMLs

Si vous achetez des licences SQL Server auprès de revendeurs Microsoft, les licences principales sont vendues par lots de deux et vous devez obtenir une licence d'au moins quatre cœurs par serveur. Le tableau suivant présente une comparaison des coûts entre les éditions Enterprise et Standard.

Version	Édition SQL Server Enterprise (pack de 2 cœurs)	Édition standard de SQL Server (pack de 2 cœurs)	Économies
2022	15 123\$	3 945\$	74 %

Version	Édition SQL Server Enterprise (pack de 2 cœurs)	Édition standard de SQL Server (pack de 2 cœurs)	Économies
2019	13 748\$	3 586\$	74 %

 Note

Les prix indiqués dans le tableau précédent sont basés sur les tarifs publics de Microsoft pour [SQL Server 2022](#) et [SQL Server 2019](#).

La comparaison des coûts suivante montre l'hébergement de différentes éditions de SQL Server avec une licence Amazon EC2 incluse. AMIs Dans cette comparaison, SQL Server est hébergé sur r6i.xlarge (4 vCPU) dans la région. us-east-1

Instance	Coût de calcul	Coût de la licence Windows	Coût de licence SQL Server	Total
R6i.xLarge (Linux)	183,96\$	–	–	183,96\$
R6i.xLarge+ Windows	183,96\$	134,32\$	–	318,28\$
R6i.xLarge + Édition Web SQL Server	183,96\$	134,32\$	49,35\$	367,63\$
R6i.xLarge+ Édition standard de SQL Server	183,96\$	134,32\$	350,4\$	668,68\$
Édition R6i.xLarge+ SQL Enterprise	183,96\$	134,32\$	1 095\$	1 413,28\$

Vous pouvez économiser jusqu'à 95 % sur les coûts de licence SQL Server en sélectionnant l'édition de SQL Server adaptée à votre charge de travail. Le tableau suivant compare le coût des licences SQL Server sur les instances r6i.xlarge.

Edition	% d'économies
Standard par rapport à Enterprise	68 %
Comparaison entre le Web et le standard	86 %
Comparaison entre le Web et l'entreprise	95 %

Dans la plupart des scénarios, les entreprises passent de l'édition Enterprise à l'édition Standard, mais dans certains cas, il est possible de passer de l'édition Standard ou Enterprise à l'édition Web.

Recommandations en matière d'optimisation des coûts

Vous pouvez choisir l'édition la mieux adaptée à votre charge de travail en fonction des limites d'évolutivité, de la haute disponibilité, des performances et de la sécurité. Le tableau suivant présente les fonctionnalités prises en charge dans toutes les éditions de SQL Server. Cela peut vous aider à choisir l'édition à utiliser. Cette comparaison s'applique à [SQL Server 2016 SP1 et aux versions ultérieures](#).

Limites d'échelle

Le tableau suivant compare les limites de dimensionnement des différentes éditions de SQL Server.

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
Capacité de calcul maximale utilisée par une instance unique du moteur de base de données SQL Server, de SQL	Système d'exploitation maximal	Limité à la moins élevée des valeurs suivantes : 4 sockets ou 24 cœurs	Limité à la moins élevée des valeurs suivantes : 4 sockets ou 16 cœurs	Limité à un nombre inférieur de 4 sockets ou 4 cœurs

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
Server Analysis Services (SSAS) ou de SQL Server Reporting Services (SSRS)				
Mémoire maximale pour le pool de tampons par instance du moteur de base de données SQL Server	Système d'exploitation maximal	128 Go	64 Go	1410 MB
Capacité maximale d'extension du pool de mémoire tampon par instance du moteur de base de données SQL Server	32 fois la mémoire maximale configurée	4 fois la mémoire maximale configurée	N/A	N/A
Taille maximale de la base de données relationnelle	524 POINTS DE BASE	524 POINTS DE BASE	524 POINTS DE BASE	10 Go
Mémoire maximale pour les caches de Columnstore ou pour les données optimisées	Système d'exploitation maximal	32 GO	16 Go	352 MO

Si votre application nécessite moins de 16 cœurs (32 vCPUs) et 64 Go de RAM, vous pouvez commencer l'évaluation à partir de l'édition Web de SQL Server. Si votre charge de travail nécessite plus de 64 Go de mémoire ou d'autres options de haute disponibilité, vous devez passer à l'édition Standard de SQL Server.

Vous pouvez utiliser l'édition Web de SQL Server pour prendre en charge les pages Web, les sites Web, les applications Web et les services Web publics et accessibles à Internet, mais vous ne pouvez pas utiliser l'édition Web de SQL Server pour prendre en charge les applications métier. Pour plus d'informations sur les cas d'utilisation de l'édition Web de SQL Server, contactez le [Support des licences Microsoft](#) ou votre revendeur Microsoft.

Vous pouvez utiliser l'édition Standard de SQL Server pour des charges de travail allant jusqu'à 24 cœurs (48 vCPUs) et 128 Go de mémoire. Cependant, vous pouvez utiliser des [extensions de pool de mémoire tampon](#) pour permettre à l'édition standard de SQL Server d'utiliser le [stockage d'instance local](#), comme ceux présents dans les instances EC2 r6id. Cela permet d'étendre la mémoire jusqu'à quatre fois la configuration maximale de la mémoire. Cette combinaison de fonctionnalités peut empêcher un serveur de passer à l'édition Enterprise lorsque les besoins en mémoire commencent à augmenter.

Vous pouvez identifier l'utilisation de la mémoire en recherchant les pages des bases de données dans le pool de mémoire tampon et les compteurs d'[espérance de vie des pages](#). La durée de vie de la page indique la durée pendant laquelle la page est en mémoire avant d'être renvoyée sur le disque. La valeur par défaut de ce compteur est 300. Si les pages restent en mémoire pendant des heures ou des jours, il est possible que la mémoire allouée soit réduite.

Haute disponibilité

Le tableau suivant compare les fonctionnalités de haute disponibilité des différentes éditions de SQL Server.

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
Support principal du serveur 1	Oui	Oui	Oui	Oui
Expédition de journaux	Oui	Oui	Oui	Non

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
Mise en miroir de bases de données	Oui	Mode de sécurité COMPLET	Uniquement en tant que témoin	Uniquement en tant que témoin
Compression de sauvegarde	Oui	Oui	Non	Non
Instances de cluster Always On Failover	16 nœuds	2 nœuds	Non	Non
Groupes de disponibilité Always On	Jusqu'à 8 répliques secondaires, dont 2 répliques secondaires synchrones	Non	Non	Non
Groupes de disponibilité de base	Non	2 nœuds	Non	Non
Restauration de pages et de fichiers en ligne	Oui	Non	Non	Non
Indexation en ligne	Oui	Non	Non	Non
Modification du schéma en ligne	Oui	Non	Non	Non
Récupération rapide	Oui	Non	Non	Non

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
Sauvegardes en miroir	Oui	Non	Non	Non
Ajouter à chaud de la mémoire et du processeur	Oui	Non	Non	Non
Sauvegarde cryptée	Oui	Oui	Non	Non
Sauvegarde hybride vers Microsoft Azure (sauvegarde vers une URL)	Oui	Oui	Non	Non
Serveur failover pour la reprise après sinistre	Oui	Oui	Non	Non
Serveurs failover pour une haute disponibilité	Oui	Oui	Non	Non

Autres caractéristiques communes

Le tableau suivant compare les fonctionnalités les plus courantes des différentes éditions de SQL Server. Pour une liste complète des fonctionnalités, consultez la section [Éditions et fonctionnalités prises en charge de SQL Server 2019](#) dans la documentation Microsoft.

Fonctionnalité	Édition Enterprise	Édition Standard	Édition Web	Édition Express
(Performance) Gouverneur des ressources	Oui	Non	Non	Non
(Sécurité) Chiffrement transparent des bases de données (TDE)	Oui	Oui	Non	Non
(Sécurité) Gestion extensible des clés (EKM)	Oui	Non	Non	Non
(Réplication) Publication Oracle	Oui	Non	Non	Non
(Réplication) Réplication transactionnelle poste à poste	Oui	Non	Non	Non
Capture de données modifiées	Oui	Oui	Non	Non

Édition SQL Server Developer

Toutes les charges de travail non liées à la production, telles que les environnements de développement, d'assurance qualité, de test, de préparation et d'UAT, peuvent utiliser l'édition SQL Server Developer pour économiser 100 % sur les coûts de licence SQL Server. Après avoir [téléchargé SQL Server](#), vous pouvez installer l'édition SQL Server Developer sur les instances EC2 à l'aide de la location partagée. Aucune infrastructure dédiée n'est requise pour l'édition SQL Server

Developer. Pour plus d'informations, consultez les recommandations de ce guide pour l'[édition SQL Server Developer](#).

Changer d'édition

Pour les charges de travail existantes, le passage d'une édition à une autre nécessite des tests approfondis. Il est recommandé de vérifier les charges de travail exécutées sur les éditions Enterprise ou Standard pour voir si des fonctionnalités spécifiques à l'édition sont utilisées et s'il existe des solutions alternatives pour ces fonctionnalités. Par exemple, si vous souhaitez savoir si vos bases de données utilisent des fonctionnalités de niveau entreprise, vous pouvez exécuter des [vues de gestion dynamique \(DMV\)](#) sur toutes les bases de données, comme le montre l'exemple de commande suivant.

```
SELECT feature_name FROM sys.dm_db_persisted_sku_features; GO
```

Certaines fonctionnalités de l'édition Enterprise ne peuvent pas être capturées dans T-SQL, telles que la réindexation en ligne dans le cadre de tâches de maintenance SQL. Ils doivent être vérifiés manuellement.

Considérations concernant la migration

La façon dont vous octroyez une licence à SQL Server déterminera les options qui s'offrent à vous pour changer d'édition. Le coût de licence des AMI, y compris les AMI SQL Server, est inclus dans le prix de l'instance EC2 ; le coût de licence est lié à l'AMI. Vous pouvez utiliser [des codes AWS de facturation](#) pour vérifier la version de SQL Server incluse dans votre AMI. Pour les instances AWS incluses dans une licence, la modification de l'édition de SQL Server dans le système d'exploitation ne modifiera pas la facturation associée à l'AMI. Vous devez migrer vos bases de données vers une nouvelle instance EC2 dotée d'une AMI exécutant la nouvelle édition de SQL Server.

Si vous apportez votre propre licence, vous bénéficiez d'une plus grande flexibilité. Il est généralement toujours recommandé de migrer vers une autre instance EC2 qui exécute la nouvelle version. Cela permet un retour en arrière facile si quelque chose ne se passe pas comme prévu. Toutefois, si vous devez utiliser le serveur existant, vous pouvez toujours side-by-side installer SQL Server et migrer les bases de données entre les instances. Pour des étapes plus détaillées sur les rétrogradations d' side-by-side édition, consultez la section [Mise à niveau d'édition et rétrogradation dans SQL Server](#) sur le MSSQLTips site Web.

Ressources supplémentaires

- [Éditions et fonctionnalités prises en charge de SQL Server 2022](#) (Microsoft Learn)

- [sys.dm_db_persisted_sku_features \(Transact-SQL\)](#) (Microsoft Learn)
- [Quelle version de SQL Server devriez-vous utiliser ?](#) (Brent Ozar Illimité)
- [Calculateur de tarification AWS](#) (AWS)

Évaluer l'édition SQL Server Developer

Présentation de

L'[édition SQL Server Developer](#) est une édition gratuite de SQL Server qui contient toutes les fonctionnalités de l'édition Enterprise et peut être utilisée dans n'importe quel environnement hors production. Dans le cloud, où les licences Microsoft Developer Network (MSDN) ne peuvent pas être utilisées, l'édition SQL Server Developer constitue un bon moyen de réduire les coûts sans avoir à fournir de licences pour les charges de travail de développement et de test. Cela est particulièrement vrai pour les équipes qui gèrent de grands environnements de développement et de test et cherchent à réduire les coûts inutiles.

Un environnement de production est défini comme un environnement auquel accèdent les utilisateurs finaux d'une application (tel qu'un site Web Internet) et qui n'est pas uniquement utilisé pour recueillir des commentaires ou tester l'acceptation de cette application. Les autres scénarios qui constituent des environnements de production incluent :


- Environnements qui se connectent à une base de données de production
- Environnements prenant en charge la reprise après sinistre ou la sauvegarde pour un environnement de production
- Environnements utilisés pour la production au moins de temps en temps, tels qu'un serveur qui passe en production pendant les périodes de pointe

Pour plus d'informations sur les licences, consultez [Amazon Web Services et Microsoft : Frequently Asked Questions](#) dans la AWS documentation.

Impact sur les coûts

Si vous utilisez l'édition SQL Server Developer pour les charges de travail hors production, vous pouvez économiser 100 % de vos coûts de licence SQL Server actuels pour les environnements de développement et de test.

Version SQL Server Version	Édition SQL Server Enterprise (pack de 2 cœurs)	Édition standard de SQL Server (pack de 2 cœurs)	Édition SQL Server Developer
2022	15 123\$	3 945\$	Free
2019	13 748\$	3 586\$	Free

 Note

Les prix indiqués dans le tableau précédent sont basés sur les tarifs publics de Microsoft pour [SQL Server 2022](#) et [SQL Server 2019](#).

Le tableau suivant compare le coût des différentes éditions de SQL Server fonctionnant en version 4 V CPUs et utilisant la tarification à la demande dans la us-east-2 région. Cela s'applique aux scénarios qui reposent sur des instances incluses dans une licence provenant de AWS

instance EC2	AMI	Calculez le prix	Prix de la licence Windows	Prix de la licence SQL Server	Prix total
r5.xlarge	Linux (tarifica tion du calcul)	183,96\$	–	–	183,96\$
r5.xlarge	Édition Linux + SQL Server Developer	183,96\$	\$0	\$0	183,96\$
r5.xlarge	Windows Server (LI)	183,96\$	134,32\$	–	318,28\$
r5.xlarge	Édition Windows + SQL Server Developer	183,96\$	134,32\$	\$0	318,28\$

instance EC2	AMI	Calculez le prix	Prix de la licence Windows	Prix de la licence SQL Server	Prix total
r5.xlarge	Édition Web Windows + SQL Server (LI)	183,96\$	134,32\$	49,64\$	367,92\$
r5.xlarge	Windows + Édition standard de SQL Server (LI)	183,96\$	134,32\$	350,4\$	668,68\$
r5.xlarge	Windows + Édition SQL Server Enterprise (LI)	183,96\$	134,32\$	1095\$	1413,28\$

Scénario d'optimisation des coûts

Après avoir effectué une nouvelle acquisition, une société d'intégrité des données a souhaité migrer la charge de travail nouvellement acquise depuis son emplacement actuel sur un fournisseur d'hébergement géré afin de la consolider avec ses autres charges de travail dans le AWS Cloud. Le prix initial indiquait que la charge de travail de SQL Server de l'entreprise coûterait 60 % AWS de plus qu'avec le fournisseur de services gérés actuel. Une PME de MACO a évalué l'estimation et a constaté que le client payait réellement les licences SQL Server auprès du fournisseur d'hébergement géré pour ses environnements de développement et de test. En transférant les charges de travail hors production vers l'édition SQL Server Developer pendant la migration, l'entreprise a réduit ses licences SQL Server de 40 %.

Licence SQL Server incluse sur Amazon EC2

Si vous avez des instances SQL Server sur EC2 qui utilisent une [licence incluse AMIs](#), il n'est pas possible d'effectuer une conversion directe de l'édition Enterprise à l'édition Developer. Les coûts

de licence pour les instances incluses dans la licence sont liés à l'AMI. Même si SQL Server est désinstallé du système d'exploitation, les frais de licence restent facturés à l'instance EC2.

Pour passer à l'édition Developer, vous devez [télécharger l'édition SQL Server Developer](#), l'installer sur une nouvelle instance EC2, puis migrer vos bases de données. Vous pouvez migrer des bases de données SQL Server entre des instances EC2 à l'aide de diverses méthodes. Pour plus d'informations, consultez la section [Méthodes de migration de bases de données SQL Server](#) dans le AWS Cloud guide intitulé Migration des bases de données Microsoft SQL Server. Vous pouvez également utiliser la [solution Automated SQL Server Developer](#) pour préparer la nouvelle instance vers laquelle vous prévoyez de migrer.

SQL Server BYOL sur Amazon EC2

Si vous avez des instances SQL Server qui utilisent un BYOL, vous pouvez choisir parmi les options de conversion ou de side-by-side rétrogradation sur place suivantes :

- Téléchargez l'[édition SQL Server Developer](#) sur le site Web de Microsoft. Pour obtenir des instructions d'installation manuelle ou automatique, consultez le billet [Automating SQL Server Developer Developer](#) sur le AWS blog.
- Utilisez la [sauvegarde et la restauration natives de SQL Server](#) pour migrer des bases de données ou des detach/attach bases de données d'une instance SQL à une autre.
- Utilisez un [outil d'automatisation](#) pour les déploiements en masse.

Note

L'édition SQL Server Developer est réservée aux environnements hors production.

Ressources supplémentaires

- [Automatisation des déploiements de SQL Server Developer pour le déploiement de SQL Server Developer Edition sur EC2 \(Blog\)AWS](#)
- [Tarification de SQL 2022](#) (Microsoft)
- [Tarification de SQL 2019](#) (Microsoft)
- [Options de licence](#) (SQL Server sur Amazon EC2)
- [Calculateur de tarification AWS](#)(documentation SQL Server sur Amazon EC2)

- [Guide des licences Microsoft SQL Server 2019](#) (à télécharger sur le site de Microsoft)
- [Édition SQL Server 2022 Developer](#) (téléchargement depuis Microsoft)

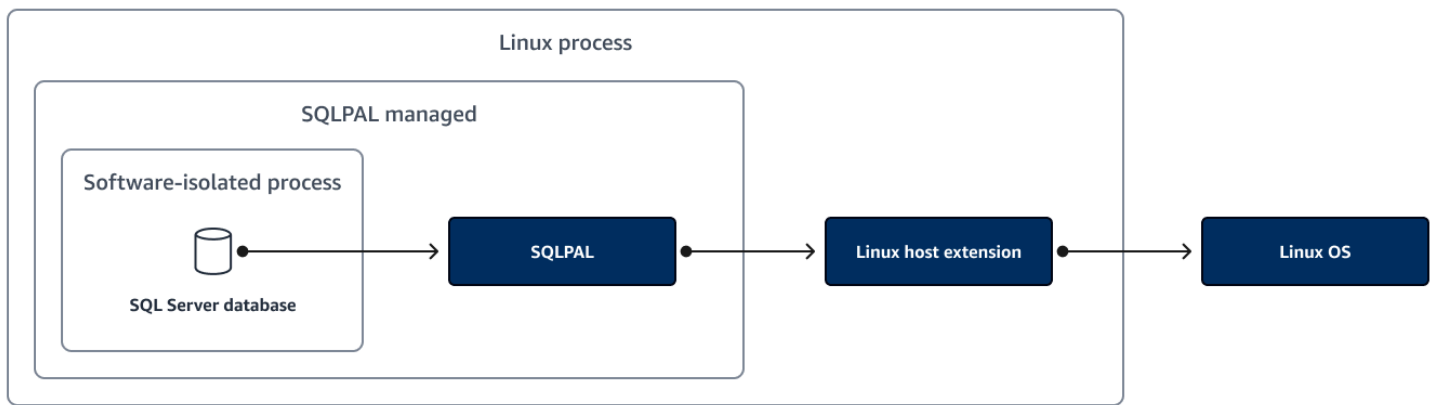
Évaluer SQL Server sous Linux

Présentation de

Depuis SQL Server 2017, il est possible d'installer SQL Server sur les systèmes d'exploitation Linux. SQL Server sous Linux est adapté aux entreprises et offre flexibilité, hautes performances, fonctionnalités de sécurité, un coût total de possession réduit, HA/DR des fonctionnalités et une expérience utilisateur exceptionnelle. Vous pouvez passer de SQL Server sur Windows Server à SQL Server sur Linux pour économiser sur les coûts de licence Windows Server.

Pour Linux, SQL Server peut être déployé sur Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), Ubuntu et Amazon Linux 2. Le moteur de base de données SQL Server fonctionne de la même manière sous Windows Server et Linux, mais certaines tâches sont fondamentalement modifiées sous Linux. L'une des principales différences entre l'exécution d'applications SQL Server Always On sous Linux et Windows est liée au clustering par basculement. Si vous déployez des groupes de disponibilité Always On sur un hôte Windows Server, vous pouvez tirer parti de [Windows Server Failover Clustering \(WSFC\)](#) et d'Active Directory en tant que fonctionnalités intégrées qui prennent en charge le clustering de basculement. Cependant, ni WSFC ni Active Directory ne sont disponibles pour prendre en charge le clustering par basculement sous Linux. [Si vous souhaitez lancer le clustering par basculement pour SQL Server sous Linux, vous pouvez simplifier la configuration du cluster et l'installation de SQL sur des instances Linux AWS Launch Wizard à l'aide ClusterLabs de Pacemaker.](#)

SQL Server sous Windows et Linux partagent une base de code commune. C'est-à-dire que le moteur principal de SQL Server n'a pas du tout été modifié pour fonctionner sous Linux. SQL Server a introduit une couche d'abstraction de plate-forme (SQLPAL), comme le montre le schéma suivant.



SQLPAL est responsable de l'abstraction des appels et de la communication entre SQL Server et le système d'exploitation sous-jacent. L'extension hôte est simplement une application Linux native. Les fonctions du système d'exploitation de bas niveau sont des appels natifs destinés à optimiser les E/S, la mémoire et l'utilisation du processeur. Lorsque l'extension hôte démarre, elle charge et initialise SQLPAL, qui ouvre ensuite SQL Server. SQLPAL lance des processus logiciels isolés qui fournissent la traduction requise pour le reste du code. L'ajout de cette nouvelle couche à l'architecture SQL Server signifie que les mêmes fonctionnalités et avantages essentiels au niveau de l'entreprise qui ont rendu SQL Server si puissant sous Windows sont disponibles quel que soit le système d'exploitation.

Impact sur les coûts

Pour les instances r5.2xlarge, la réduction des coûts de licence Windows Server est d'environ 268\$ dans chaque scénario. La réduction représente un pourcentage plus élevé du coût total du serveur par rapport à l'utilisation d'éditions moins coûteuses de SQL Server. Le tableau suivant montre les économies de coûts.

Instance	Edition	Coût mensuel de SQL Server sous Windows	Coût mensuel de SQL Server sous Linux	Économies
r5.2xlarge	Web	735\$	466\$	37%
r5.2xlarge	Standard	1 337\$	1 068\$	20 %
r5.2xlarge	Enterprise	2 826\$	2 558\$	10 %

Note

L'estimation des prix dans le tableau précédent est basée sur la tarification à la demande dans la us-east-1 Région et peut être consultée directement dans le [Calculateur de tarification AWS](#).

Prenons un exemple de scénario dans lequel un client ISV du segment des PME cherche à réduire les coûts de son environnement de développement. Ils utilisent déjà l'édition SQL Server Developer sur un ensemble de serveurs Windows. En passant de l'édition Windows avec SQL Server Developer à Linux avec l'édition SQL Server Developer, le client ISV peut économiser 33 % sur sa charge de travail de développement. Le tableau suivant indique les coûts estimés suivants pour ce scénario.

Estimate	Coût mensuel
Windows + SQL Server	9 307,72\$
Linux + SQL Server	6 218,36\$
Économies de coûts estimées	3 089,36\$ (33 %)

Dans un autre exemple de scénario, une entreprise migre des instances SQL Server EC2 incluses dans une licence de Windows vers Linux. L'entreprise économise au total 300 000 dollars par an sur les coûts de licence Windows Server, soit environ 20 % de sa facture totale. AWS

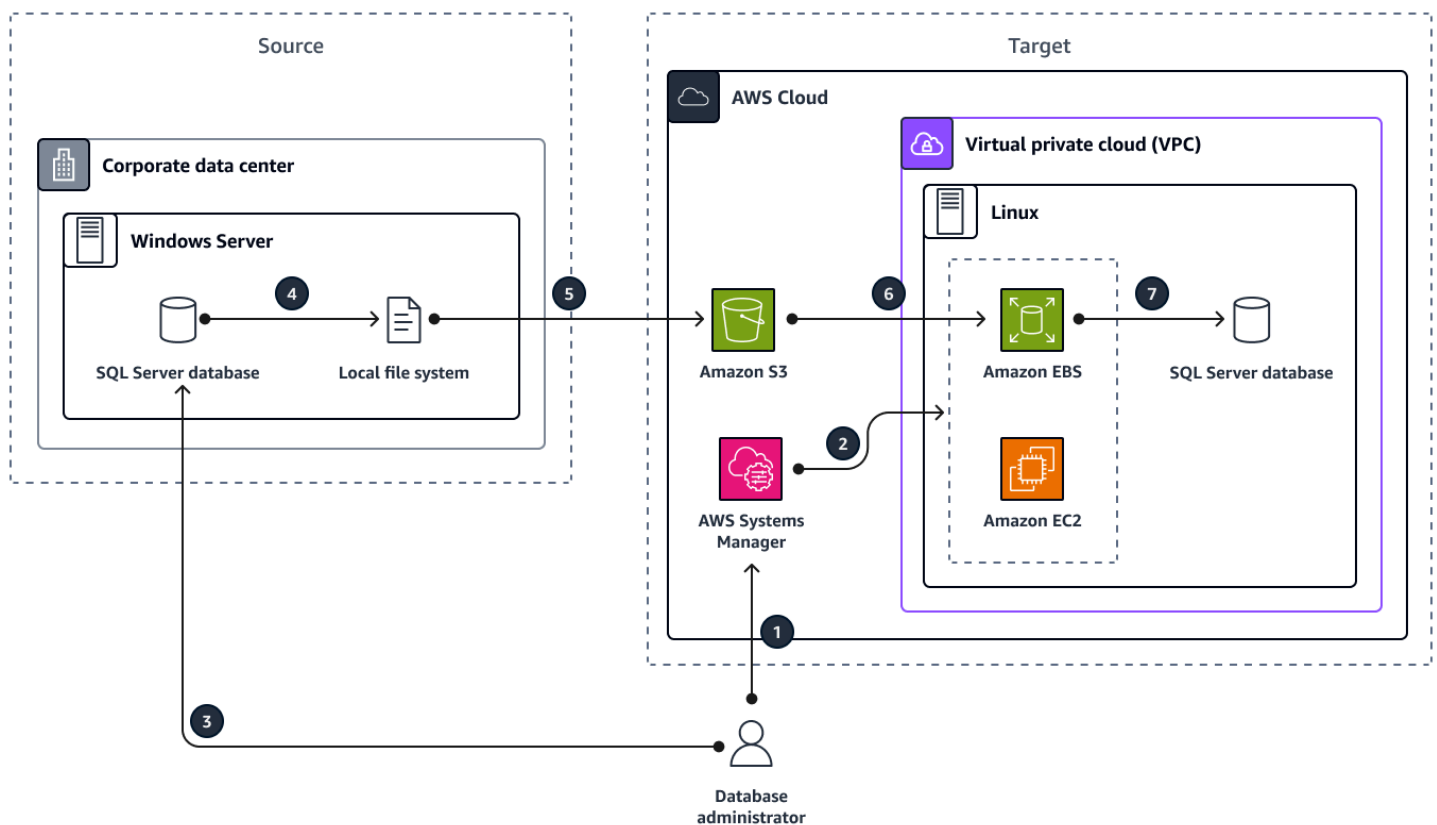
Recommandations en matière d'optimisation des coûts

Nous vous recommandons de prendre en compte les points suivants :

- SQL Server sous Linux est pris en charge à partir de SQL Server 2017.
- Pour faciliter la transition, vous pouvez utiliser l'[assistant de replateforme Windows vers Linux pour les bases de données Microsoft SQL Server](#). L'assistant de replateforme est un outil de script qui peut vous aider à déplacer les charges de travail SQL Server existantes des systèmes d'exploitation Windows vers les systèmes d'exploitation Linux en vérifiant les incompatibilités courantes, en exportant les bases de données depuis l'hôte Windows, puis en les important dans une instance EC2 exécutant Microsoft SQL Server 2017 sur Ubuntu 16.04.

- Vous pouvez également utiliser les fonctionnalités de [sauvegarde et de restauration](#) de SQL Server pour passer de SQL Server sous Windows à Linux.
- Vous pouvez facilement et rapidement déployer sur SQL Server sous Linux ou Ubuntu à l'aide du [AWS Launch Wizard](#). Le Launch Wizard peut déployer SQL Server sous Linux ou Ubuntu dans des scénarios autonomes et de haute disponibilité en fonction des besoins de votre application. Pour plus d'informations, consultez l'article [Déploiement sur SQL Server Always sous Linux avec AWS Launch Wizard](#) publié sur le AWS blog Microsoft Workloads on.

Le schéma suivant montre l'architecture d'une solution qui utilise l'assistant de replatforme Windows vers Linux pour les bases de données Microsoft SQL Server.



Ressources supplémentaires

- [Présentation de SQL Server sous Linux](#) (Microsoft Learn)
- [Guide d'installation pour SQL Server sous Linux](#) (Microsoft Learn)
- [Déploiement sur SQL Server Always on Linux avec AWS Launch Wizard](#) (Microsoft Workloads on AWS Blog)
- [SQL Server hautement disponible sous Linux](#) (blog AWS Open Source)

Optimisation des stratégies de sauvegarde de SQL Server

Présentation de

La plupart des entreprises recherchent la bonne solution pour protéger leurs données sur SQL Server sur [Amazon EC2](#) afin de répondre à leurs exigences actuelles en matière d'objectif de point de restauration (RPO), le délai maximum acceptable depuis la dernière sauvegarde, et d'objectif de temps de restauration (RTO), le délai maximum acceptable entre l'interruption de service et le rétablissement du service. Si vous exécutez SQL Server sur des instances EC2, plusieurs options s'offrent à vous pour créer des sauvegardes de vos données et les restaurer. Les politiques de sauvegarde destinées à protéger les données de SQL Server sur Amazon EC2 sont les suivantes :

- Sauvegarde au niveau du serveur à l'aide d'instantanés [Amazon Elastic Block Store \(Amazon EBS\)](#) compatibles avec Windows Volume Shadow Copy Service (VSS) ou [AWS Backup](#)
- Sauvegarde au niveau de la base de données à l'aide de la [sauvegarde et de la restauration natives](#) dans SQL Server

Vous disposez des options de stockage suivantes pour la sauvegarde native au [niveau de la base de données](#) :

- Une sauvegarde locale avec un volume [Amazon EBS](#)
- Une sauvegarde du système de fichiers réseau avec [Amazon FSx pour Windows File Server](#) ou Amazon FSx pour NetApp ONTAP
- Une sauvegarde réseau vers Amazon Simple Storage Service (Amazon S3) à l'aide de [AWS Storage Gateway](#)
- Sauvegarde directe sur Amazon S3 pour SQL Server 2022

Cette section effectue les opérations suivantes :

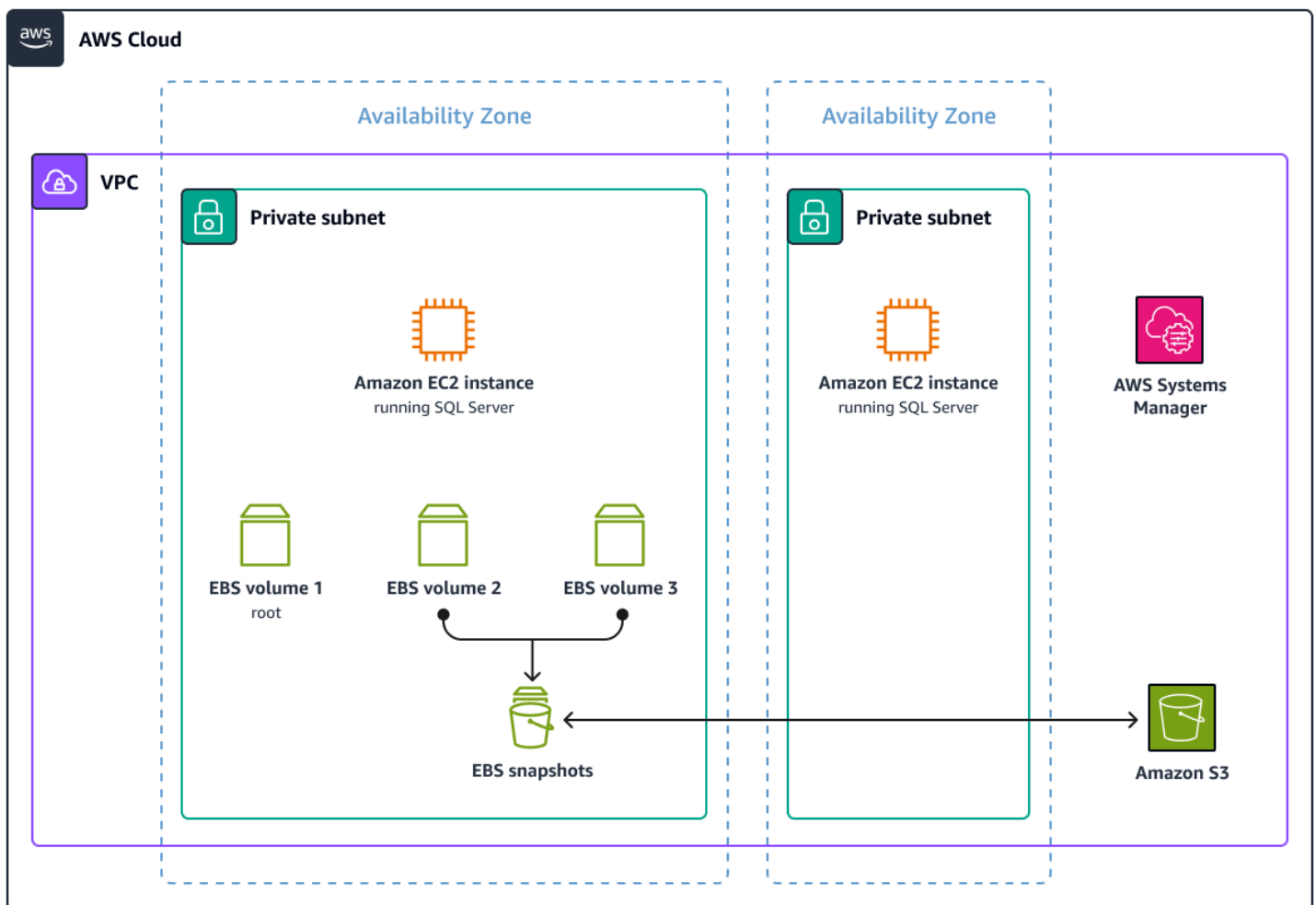
- Met en évidence les fonctionnalités qui vous aident à économiser de l'espace de stockage
- Compare les coûts entre les différentes options de stockage principal
- Fournit des liens vers une documentation détaillée pour aider à mettre en œuvre ces recommandations

Sauvegarde au niveau du serveur à l'aide de snapshots compatibles VSS

Une architecture de snapshots compatible VSS utilise la [commande AWS Systems Manager Run](#) pour installer l'agent VSS sur vos instances SQL Server. Vous pouvez également utiliser la commande Run pour appeler l'ensemble du flux de travail consistant à vider les tampons du système d'exploitation et des applications sur le disque, à suspendre les I/O opérations, à prendre un point-in-time instantané des volumes EBS, puis à reprendre les E/S.

Cette commande d'exécution crée des instantanés automatisés de tous les volumes EBS attachés à une instance cible. Vous avez également la possibilité d'exclure le volume racine, car les fichiers de base de données utilisateur sont généralement stockés sur d'autres volumes. Si vous découpez plusieurs volumes EBS pour créer un système de fichiers unique pour les fichiers SQL Server, Amazon EBS prend également en charge les instantanés à volumes multiples en cas de panne à l'aide d'une seule commande d'API. Pour plus d'informations sur les instantanés [EBS compatibles VSS cohérents avec les applications](#), consultez la section [Créer un instantané cohérent avec les applications VSS dans la documentation Amazon EC2](#).

Le schéma suivant montre une architecture de sauvegarde au niveau du serveur à l'aide de snapshots compatibles VSS.



Tenez compte des avantages suivants liés à l'utilisation de snapshots compatibles VSS :

- Le premier instantané d'une instance de base de données contient les données de l'instance de base de données complète. Les instantanés suivants de la même instance de base de données sont [incrémentiels](#), ce qui signifie que seules les données qui ont changé depuis l'instantané le plus récent sont enregistrées.
- Les instantanés EBS assurent point-in-time la restauration.
- Vous pouvez [effectuer une restauration vers une nouvelle instance SQL Server EC2 à partir d'un instantané](#).
- Si une instance est chiffrée à l'aide d'Amazon EBS ou si une base de données est chiffrée dans l'instance à l'aide de TDE, cette instance ou base de données est automatiquement restaurée avec le même chiffrement.
- Vous pouvez copier vos [sauvegardes automatisées entre régions](#).

- Lorsque vous restaurez un volume EBS à partir d'un instantané, les applications peuvent immédiatement y accéder. Cela signifie que vous pouvez immédiatement mettre SQL Server en ligne après avoir restauré un ou plusieurs de ses volumes EBS sous-jacents à partir d'instantanés.
- Par défaut, les volumes restaurés récupèrent les blocs sous-jacents depuis Amazon S3 la première fois qu'une application essaie de les lire. Cela signifie qu'il peut y avoir un décalage dans les performances après la restauration d'un volume EBS à partir d'un instantané. Le volume finit par rattraper les performances nominales. Cependant, vous pouvez éviter ce décalage en utilisant des instantanés de [restauration rapide des instantanés \(FSR\)](#).
- Vous pouvez utiliser la [gestion du cycle de vie pour les instantanés EBS](#).

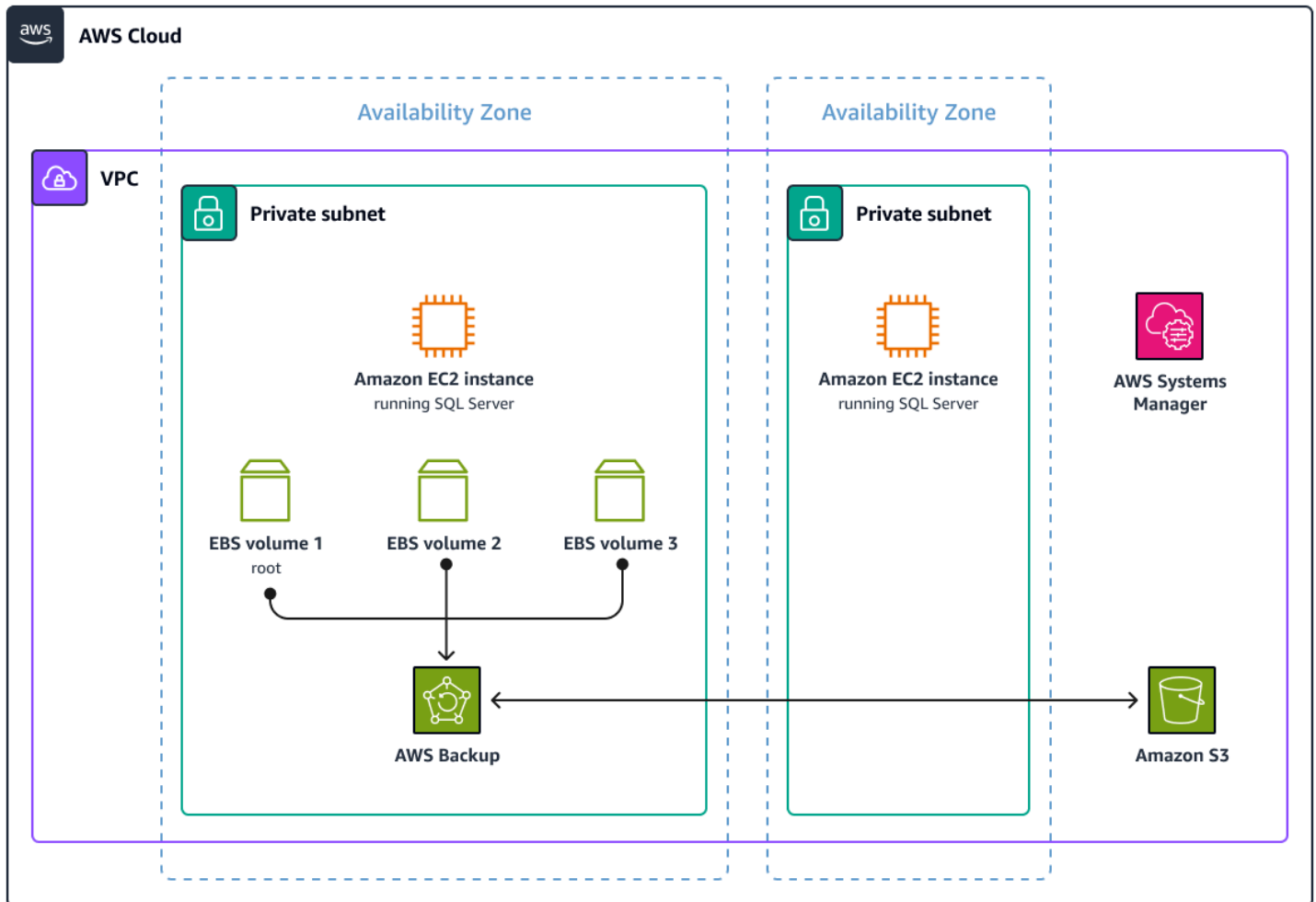
Tenez compte des limites suivantes liées à l'utilisation de snapshots compatibles VSS :

- Vous ne pouvez pas effectuer de point-in-time restauration entre régions avec un instantané chiffré pour une instance de SQL Server.
- Vous ne pouvez pas créer un instantané chiffré d'une instance non chiffrée.
- Vous ne pouvez pas restaurer une base de données individuelle car le cliché est pris au niveau du volume EBS.
- Vous ne pouvez pas restaurer l'instance elle-même.
- Un instantané de l'instance de base de données doit être chiffré à l'aide de la même clé AWS Key Management Service (AWS KMS) que l'instance de base de données.
- I/O Le stockage est suspendu pendant une fraction de seconde (environ 10 millisecondes) pendant le processus de sauvegarde des instantanés.

Sauvegarde de SQL Server à l'aide de AWS Backup

Vous pouvez l'utiliser [AWS Backup](#) pour centraliser et automatiser la protection des données dans l'ensemble Services AWS. AWS Backup propose une solution rentable, entièrement gérée et basée sur des règles qui simplifie la protection des données à grande échelle. AWS Backup vous aide également à respecter vos obligations de conformité réglementaire et à atteindre vos objectifs de continuité d'activité. Associé à AWS Organizations, vous AWS Backup permet de déployer de manière centralisée des politiques de protection des données (sauvegarde) pour configurer, gérer et gouverner vos activités de sauvegarde au sein de votre organisation Comptes AWS et de ses ressources.

Le schéma suivant montre l'architecture d'une solution de sauvegarde et de restauration pour SQL Server sur EC2 à l'aide AWS Backup de.



Tenez compte des avantages suivants liés à la sauvegarde de SQL Server en utilisant AWS Backup :

- Vous pouvez automatiser la planification des sauvegardes, la gestion de la conservation et la gestion du cycle de vie.
 - Vous pouvez centraliser votre stratégie de sauvegarde au sein de votre entreprise, en couvrant plusieurs comptes et Régions AWS.
 - Vous pouvez centraliser la surveillance de votre activité de sauvegarde et l'envoi d'alertes.
- Services AWS
- Vous pouvez implémenter des sauvegardes entre régions pour la planification de la reprise après sinistre.
 - La solution prend en charge les sauvegardes entre comptes.

- Vous pouvez effectuer des sauvegardes sécurisées à l'aide du chiffrement de sauvegarde secondaire.
- Toutes les sauvegardes prennent en charge le chiffrement à l'aide AWS KMS de clés de chiffrement.
- La solution fonctionne avec TDE.
- Vous pouvez effectuer une restauration vers un point de récupération spécifique à partir de la console AWS Backup .
- Vous pouvez sauvegarder une instance SQL Server complète, qui inclut toutes les bases de données SQL Server.

Sauvegarde au niveau de la base de données

Ces approches utilisent la fonctionnalité de sauvegarde native de Microsoft SQL Server. Vous pouvez effectuer des sauvegardes de bases de données individuelles sur l'instance SQL Server et restaurer une base de données individuelle.

Chacune de ces options de sauvegarde et de restauration natives de SQL Server prend également en charge les éléments suivants :

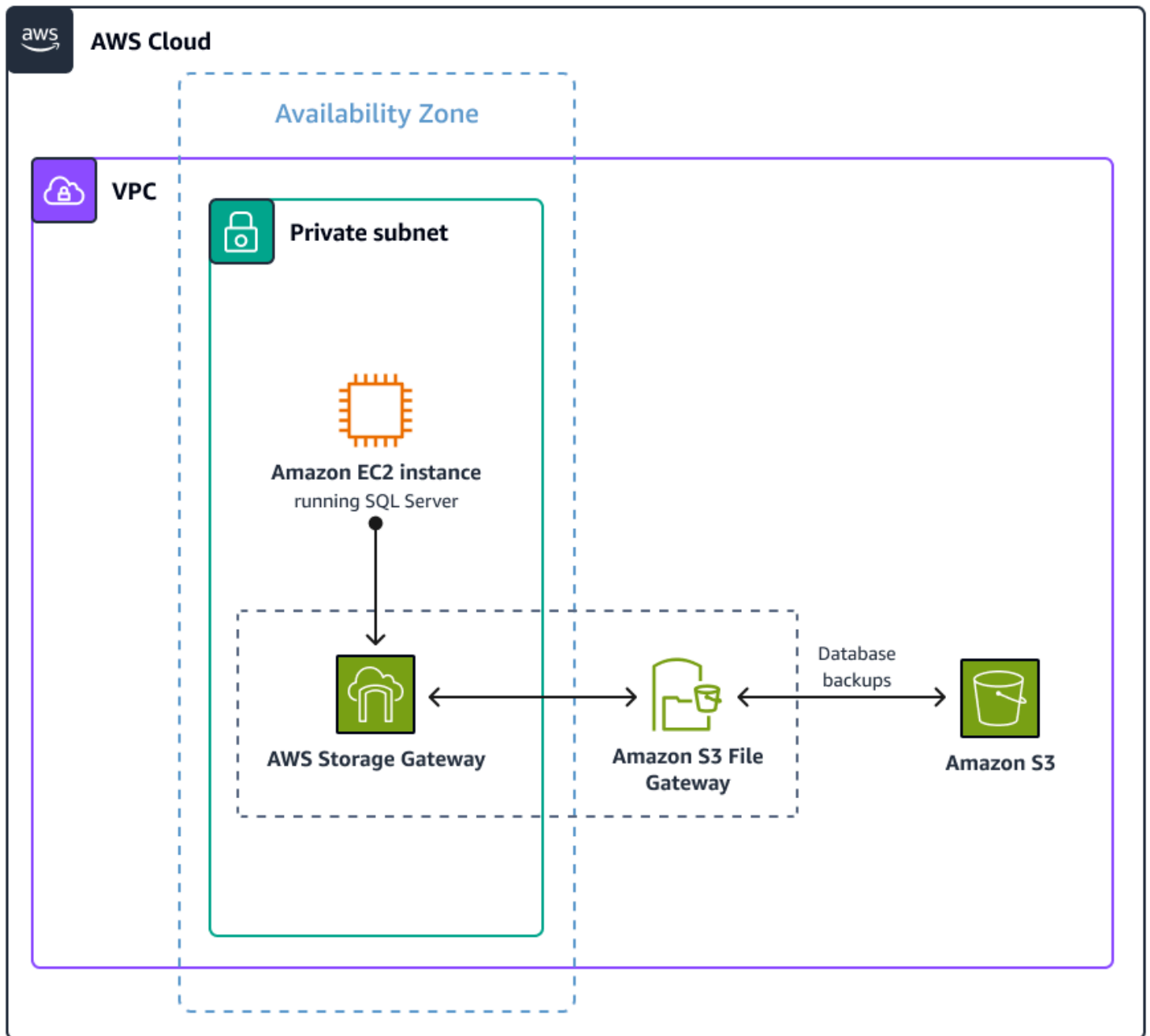
- Compression et sauvegarde de fichiers multiples
- Sauvegardes complètes, différentielles et T-log
- Bases de données cryptées TDE

Sauvegarde et restauration natives de SQL Server sur Amazon S3

SQL Server sur Amazon EC2 prend en charge la sauvegarde et la restauration natives des bases de données SQL Server. Vous pouvez effectuer une sauvegarde de votre base de données SQL Server, puis restaurer le fichier de sauvegarde sur une base de données existante ou sur une nouvelle instance SQL Server EC2, Amazon RDS for SQL Server ou un serveur sur site.

Storage Gateway est un service de stockage cloud hybride qui permet aux applications sur site d'accéder à un stockage cloud pratiquement illimité. Vous pouvez utiliser Storage Gateway pour sauvegarder vos bases de données Microsoft SQL Server directement sur Amazon S3, afin de réduire votre encombrement de stockage sur site et d'utiliser Amazon S3 pour un stockage durable, évolutif et rentable.

Le schéma suivant montre l'architecture d'une solution native de sauvegarde et de restauration qui utilise Storage Gateway et Amazon S3.



Tenez compte des avantages suivants liés à l'utilisation de la sauvegarde native de SQL Server avec Storage Gateway :

- Vous pouvez mapper une passerelle de stockage en tant que partage de fichiers SMB (Server Message Block) sur l'instance EC2 et envoyer la sauvegarde vers Amazon S3.

- La sauvegarde est directement envoyée dans le compartiment S3 ou via le cache de fichiers Storage Gateway.
- Les sauvegardes de fichiers multiples sont prises en charge.

Tenez compte des limites suivantes de la sauvegarde native à l'aide de Storage Gateway :

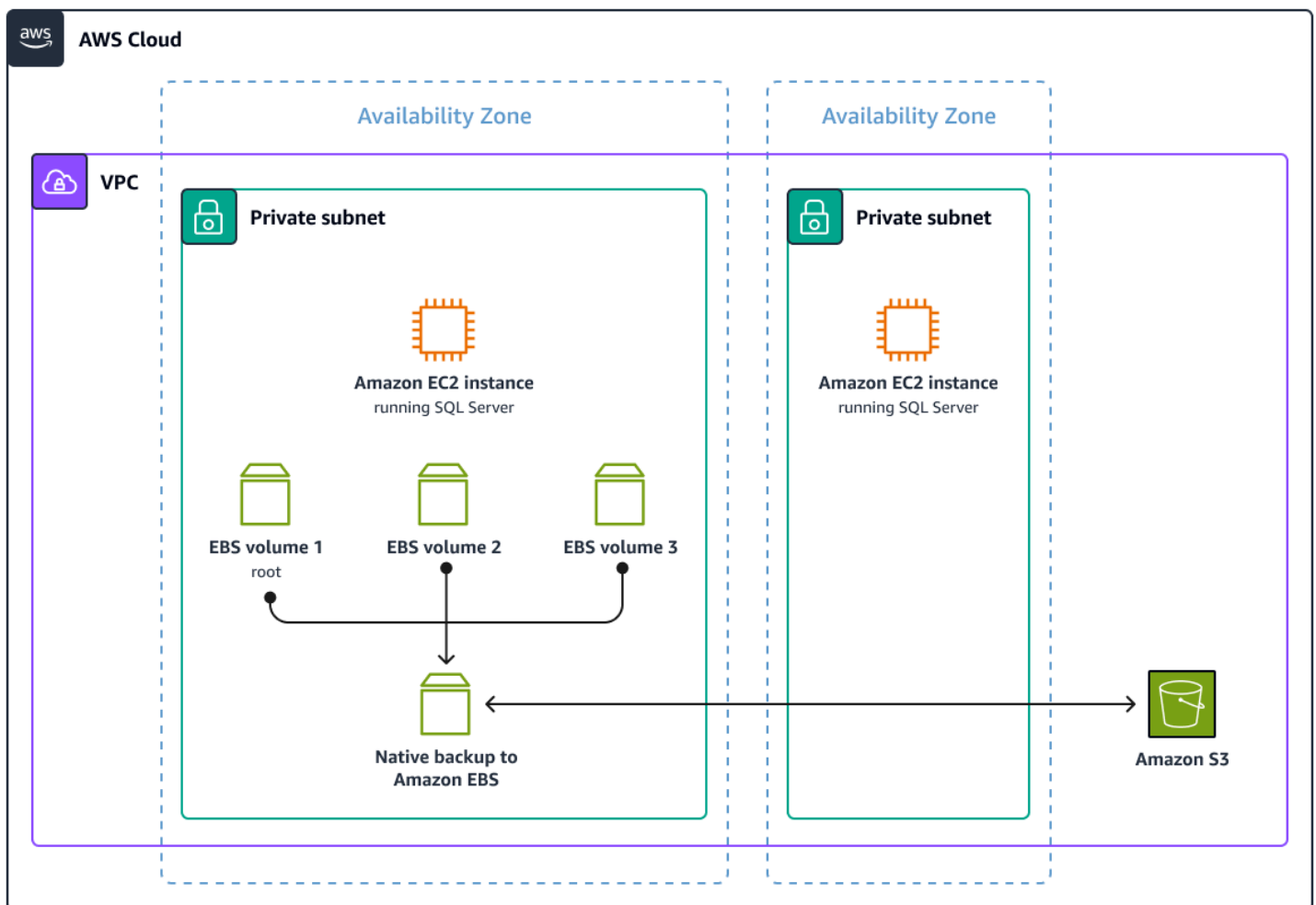
- Vous devez configurer la sauvegarde et la restauration pour chaque base de données individuelle.
- Vous devez gérer la [politique de cycle de vie Amazon S3](#) pour les fichiers de sauvegarde.

Pour plus d'informations sur la configuration de Storage Gateway, consultez l' AWS Storage Gateway article sur le AWS blog [consacré aux sauvegardes de Store SQL Server dans Amazon S3](#).

Sauvegarde native de SQL Server sur des volumes EBS

Vous pouvez effectuer une sauvegarde native de votre base de données SQL Server et stocker le fichier dans un volume Amazon EBS. Amazon EBS est un service de stockage par blocs très performant. Les volumes EBS sont élastiques, ce qui permet le chiffrement. Ils peuvent être détachés et attachés à une instance EC2. Vous pouvez sauvegarder SQL Server sur une instance EC2 du même type de volume EBS ou sur un autre type de volume EBS. L'un des avantages de la sauvegarde sur un autre volume EBS est la réduction des coûts.

Le schéma suivant montre l'architecture d'une sauvegarde native sur un volume EBS.



Tenez compte des avantages suivants liés à l'utilisation de la sauvegarde native de SQL Server sur des volumes EBS :

- Vous pouvez effectuer des sauvegardes de bases de données individuelles sur une instance SQL Server EC2 et restaurer une base de données individuelle au lieu de devoir restaurer l'instance complète.
- Les sauvegardes de fichiers multiples sont prises en charge.
- Vous pouvez planifier des tâches de sauvegarde à l'aide de l'agent SQL Server et du moteur de tâches SQL Server.
- Vous pouvez obtenir des avantages en termes de performances grâce à vos choix de matériel. Par exemple, vous pouvez utiliser les volumes de stockage st1 pour obtenir un débit plus élevé.

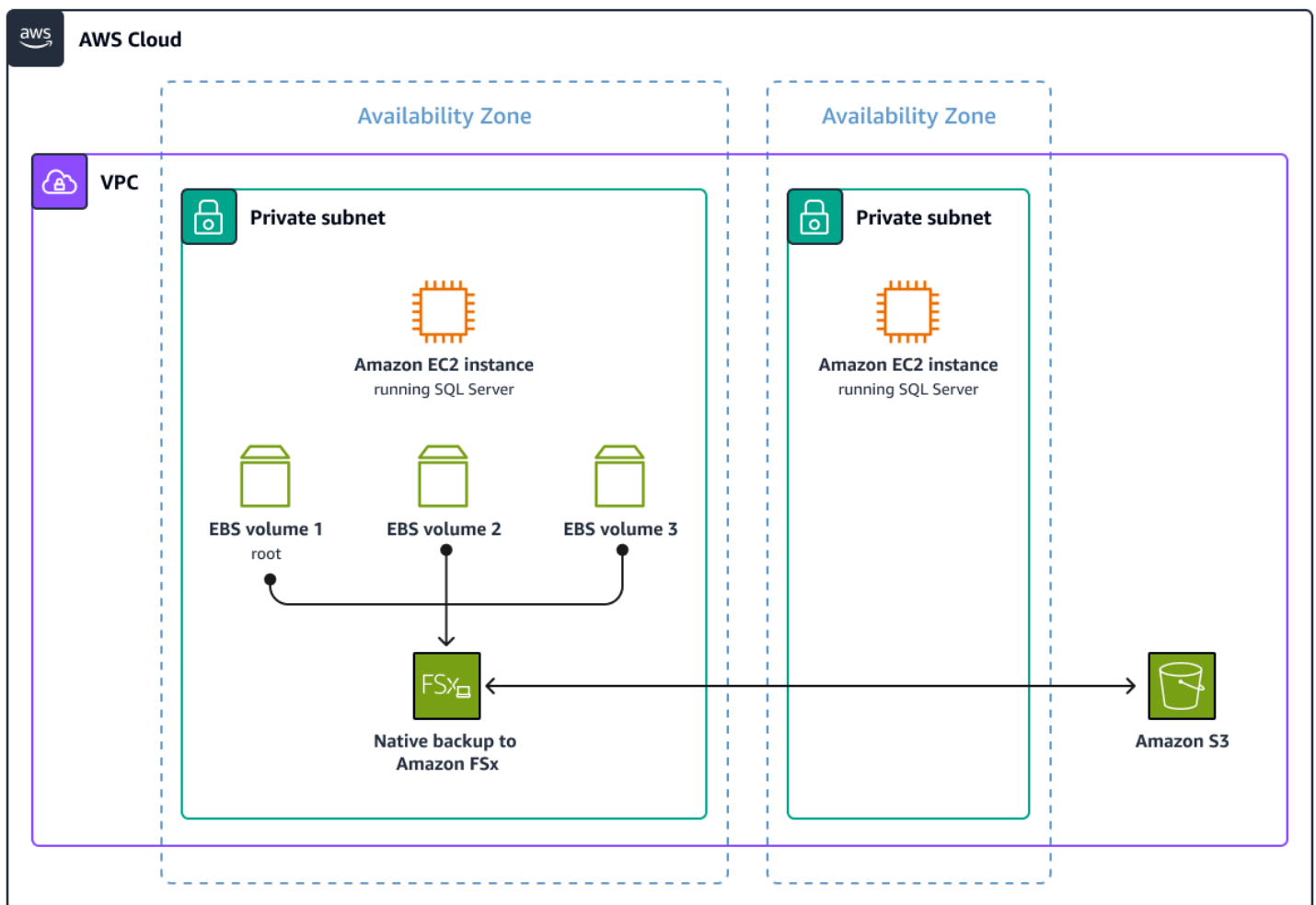
Tenez compte des limites suivantes liées à l'utilisation de la sauvegarde native sur des volumes EBS :

- Vous devez déplacer manuellement les sauvegardes vers Amazon S3 à partir du volume EBS.
- Pour les sauvegardes de grande taille, vous devez gérer l'espace disque sur Amazon EC2.
- Sur l'instance EC2, le débit d'Amazon EBS peut constituer un goulot d'étranglement.
- Un espace de stockage supplémentaire est nécessaire pour stocker les sauvegardes sur Amazon EBS.

Sauvegarde native de SQL Server sur Amazon FSx pour Windows File Server

[Amazon FSx pour Windows File Server](#) est un système de fichiers Windows natif entièrement géré qui offre jusqu'à 64 To de stockage conçu pour fournir des performances rapides, prévisibles et cohérentes. AWS a introduit le [support natif pour les déploiements de systèmes de fichiers multi-AZ](#) sur FSx Windows File Server. Le support natif facilite le déploiement du stockage de fichiers Windows AWS avec une haute disponibilité et une redondance sur plusieurs zones de disponibilité. AWS a également introduit la prise en charge des [partages de fichiers SMB disponibles en continu \(CA\)](#). Vous pouvez utiliser FSx Windows File Server comme stockage de sauvegarde pour une base de données SQL Server.

Le schéma suivant montre l'architecture d'une sauvegarde native de SQL Server FSx pour Windows File Server.



Tenez compte des avantages suivants liés à l'utilisation de la sauvegarde native de SQL Server FSx pour Windows File Server :

- Vous pouvez sauvegarder votre base de données SQL Server sur un partage de FSx fichiers Amazon.
- Vous pouvez effectuer des sauvegardes de bases de données individuelles sur une instance SQL Server et restaurer une base de données individuelle au lieu de devoir restaurer l'instance complète.
- Les sauvegardes en plusieurs parties sont prises en charge.
- Vous pouvez planifier des tâches de sauvegarde à l'aide de l'agent SQL Server et du moteur de tâches.
- Les instances disposent d'une bande passante réseau supérieure à celle d'Amazon EBS.

Tenez compte des limites suivantes liées à l'utilisation de la sauvegarde native de SQL Server FSx pour Windows File Server :

- Vous devez déplacer manuellement les sauvegardes vers Amazon S3 depuis Amazon à FSx l'aide de AWS Backup ou AWS DataSync.
- Les sauvegardes volumineuses peuvent nécessiter des frais supplémentaires pour la gestion de l'espace disque sur Amazon FSx.
- Le débit du réseau d'instances EC2 peut constituer un goulot d'étranglement.
- Un espace de stockage supplémentaire est nécessaire pour stocker les sauvegardes sur le FSx serveur de fichiers Windows.

Sauvegarde de SQL Server sur Amazon FSx pour NetApp ONTAP

Les instantanés réalisés avec FSx for ONTAP sont toujours compatibles avec les crashes, mais ils nécessitent que vous mettiez votre base de données en veille (ou suspendiez) votre base I/O de données afin de créer un instantané cohérent avec l'application. Vous pouvez utiliser NetApp SnapCenter (un outil d'orchestration avec des plug-ins pour des applications spécifiques, y compris SQL Server) avec FSx for ONTAP pour créer des instantanés cohérents avec les applications et protéger, répliquer et cloner vos bases de données sans frais supplémentaires.

NetApp SnapCenter

NetApp SnapCenter est une plate-forme unifiée pour une protection des données cohérente au niveau des applications. SnapCenter fait référence aux instantanés en tant que sauvegardes. Ce guide adopte la même convention de dénomination. SnapCenter fournit une interface unique pour gérer les sauvegardes, les restaurations et les clones cohérents avec les applications. Vous ajoutez un SnapCenter plug-in pour votre application de base de données spécifique afin de créer des sauvegardes cohérentes avec les applications. Le SnapCenter plug-in pour SQL Server fournit les fonctionnalités suivantes qui simplifient votre flux de travail de protection des données.

- Options de sauvegarde et de restauration avec granularité pour les sauvegardes complètes et les sauvegardes dans les journaux
- Restauration sur place et restauration vers un autre emplacement

Pour plus d'informations à ce sujet SnapCenter, consultez le billet « [Protégez vos charges de travail SQL Server à l' NetApp SnapCenter aide d'Amazon FSx pour NetApp ONTAP](#) » sur le blog consacré au AWS stockage.

Optimisation des coûts pour les sauvegardes

Les options suivantes peuvent vous aider à réduire le coût de stockage des sauvegardes de SQL Server sur AWS.

- Activez la [compression SQL Server](#) lors de la création du fichier de sauvegarde et envoyez le plus petit fichier possible au stockage. Par exemple, un taux de compression de 3:1 indique que vous économisez environ 66 % d'espace disque. Pour effectuer une requête sur ces colonnes, vous pouvez utiliser l'instruction Transact-SQL suivante : `SELECT backup_size/compressed_backup_size FROM msdb..backupset;`
- Pour les sauvegardes destinées à des compartiments S3, activez la classe de stockage [Amazon S3 Intelligent-Tiering](#) afin de réduire les coûts de stockage de 30 %.
- Pour les sauvegardes FSx destinées à Windows File Server ou FSx à ONTAP, utilisez une seule zone de disponibilité pour économiser 50 % (par rapport à l'utilisation de plusieurs zones de disponibilité). Pour plus d'informations sur les tarifs, consultez les tarifs des [serveurs de fichiers Amazon FSx pour Windows](#) et les tarifs [Amazon FSx pour NetApp ONTAP](#).
- L'option la plus efficace pour SQL Server 2022 est la sauvegarde directe sur Amazon S3. Vous pouvez économiser des coûts supplémentaires en évitant Storage Gateway.

Résultats des tests de référence pour les sauvegardes

Cette section compare les options suivantes du point de vue des coûts et des performances pour un exemple de base de données de 1 To, sur la base des résultats des tests de performance réalisés sur les solutions de sauvegarde décrites dans ce guide.

- Spécification de l'instance EC2 — r5d.8xlarge avec Windows Server 2019 et SQL Server 2019 Developer Edition
- Spécification de la base de données — 1 To avec TDE désactivé

Les tests ont été effectués avec une instance r5d.8xlarge et une base de données SQL Server de 1 To comme source. Le système source a été configuré conformément aux meilleures pratiques, et la base de données source contenait quatre fichiers de données (250 Go chacun) et un fichier journal (50 Go) répartis sur des volumes gp3 distincts. La BACKUP commande native de SQL Server inclut l'écriture dans 10 fichiers de sauvegarde, l'utilisation de la compression pour optimiser les performances de sauvegarde et réduire la quantité de données envoyées sur le réseau et écrites sur la cible. Dans tous les cas de test, les performances de stockage étaient le principal obstacle.

Il existe une variété presque infinie de configurations possibles pour ce type de test. Ce test s'est concentré sur l'optimisation des performances, des coûts, de l'évolutivité et des cas d'utilisation réels. Le tableau suivant présente les mesures de performance capturées pour les options de cible de sauvegarde.

Options de sauvegarde	Niveau	Durée d'exécution (Appx)	Taux de sauvegarde	Coût en dollars américains par mois*
Sauvegarde native sur disque dur EBS st1 local, 2 To	Base de données	00:30:46 min	554,7 Mbits/s	92,16\$
Sauvegarde native sur SSD EBS gp3 local, 2 To	Base de données	00:22:00 min	512 Mbits/s	193,84\$
Sauvegarde native sur disque dur FSx pour serveur de fichiers Windows, débit de 2 To @512 Mbits/s	Base de données	00:20 min 58 min	814,0 Mbit/s	1 146\$
Sauvegarde native sur SSD FSx pour serveur de fichiers Windows, débit de 2 To @512 Mbits/s	Base de données	00:20:00 minute	814,0 Mbit/s	1 326\$
Sauvegarde native vers S3	Base de données	00:23 h 20 min	731,5 Mbit/s	470,42\$

Options de sauvegarde	Niveau	Durée d'exécution (Appx)	Taux de sauvegarde	Coût en dollars américains par mois*
File Gateway m6i.4xlarge (16 vCPU, 64 Go) avec GP3 de 2 To				
Instantané EBS VSS	Volume EBS	00:00:02 sec 00:00:53 sec	Instantané N/A	<u>51\$</u>
AWS Backup (sauvegarde AMI)	AMI	00:00:04 sec 00:08:00 minute	Instantané N/A	<u>75\$</u>
Sauvegarde native de SQL Server directement sur Amazon S3 (SQL Server 2022)	Base de données	00:12 min	731,5 Mbit/s	<u>50 premiers To/ mois, 0,023\$ par Go (23,55\$ par mois)</u>
Sauvegarde native FSx pour ONTAP (en utilisant SnapCenter)	Base de données	–	–	<u>440,20\$</u>

Le tableau précédent part des hypothèses suivantes :

- Les frais de transfert de données et d'Amazon S3 ne sont pas inclus.
- Le prix du stockage est inclus dans le prix de l'instance.
- Les coûts sont basés dans la us-east-1 région.

- Le débit et les IOPS augmentent de 10 % avec des sauvegardes multiples dont le taux de variation global est de 10 % au cours du mois.

Les résultats des tests montrent que l'option la plus rapide est une sauvegarde de base de données SQL Server native vers FSx Windows File Server. La sauvegarde sur Storage Gateway et sur des volumes EBS connectés localement est l'option la plus rentable, mais ses performances sont plus lentes. Pour les sauvegardes au niveau du serveur (AMI), nous vous recommandons de les utiliser AWS Backup pour des performances, des coûts et une facilité de gestion optimaux.

Recommandations en matière d'optimisation des coûts

Comprendre les solutions possibles pour sauvegarder SQL Server sur Amazon EC2 est essentiel pour protéger vos données, garantir que vous répondez à vos besoins de sauvegarde et mettre en place un plan de restauration après un événement critique. Les différentes méthodes de sauvegarde et de restauration de vos instances et bases de données SQL Server décrites dans cette section peuvent vous aider à concevoir une stratégie de sauvegarde et de restauration qui protège vos données et répond aux exigences de votre entreprise.

Cette section couvre les options de sauvegarde suivantes :

- Compression
- Amazon S3 Intelligent Tiering
- Zone de disponibilité unique
- Backup vers une URL

Les conseils fournis pour chacune de ces options sont de haut niveau. Si vous souhaitez mettre en œuvre l'une de ces recommandations dans votre organisation, nous vous recommandons de contacter l'équipe chargée de votre compte. L'équipe peut ensuite s'adresser à un Microsoft Specialist SA pour mener la conversation. Vous pouvez également nous contacter en envoyant un e-mail à optimize-microsoft@amazon.com.

En résumé, nous recommandons ce qui suit :

- Si vous utilisez SQL Server 2022, la sauvegarde sur Amazon S3 est l'option la plus rentable.
- Si vous utilisez SQL Server 2019 et des éditions antérieures de SQL Server, considérez la sauvegarde sur Storage Gateway, soutenue par Amazon S3, comme l'option la plus rentable.

Compression

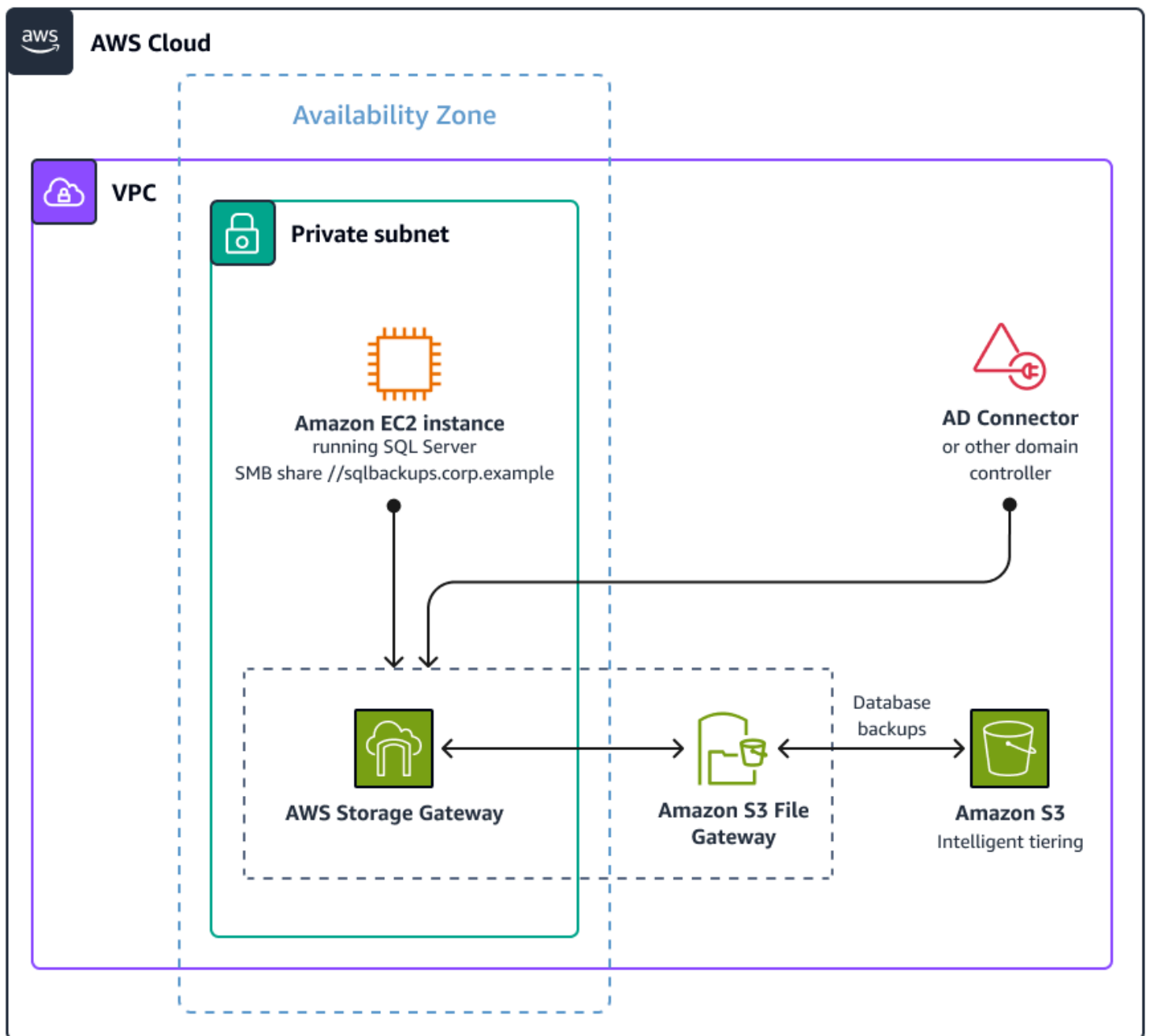
L'objectif de la compression est de réduire la consommation de stockage par chaque sauvegarde, ce qui est avantageux pour les différentes options de stockage. Vous devez activer la compression pour une sauvegarde SQL Server au niveau de l'[instance de SQL Server](#). L'exemple suivant montre comment ajouter le mot clé de compression à une base de données de sauvegarde :

```
BACKUP DATABASE <database_name> TO DISK WITH COMPRESSION (ALGORITHM = QAT_DEFLATE)
```

Amazon S3 Intelligent Tiering

[Pour les sauvegardes destinées à des compartiments Amazon S3, vous pouvez activer Amazon S3 Intelligent-Tiering comme classe de stockage Amazon S3 File Gateway.](#) Cela peut réduire les coûts de stockage jusqu'à 30 %. Vous montez ensuite S3 File Gateway sur vos serveurs SQL à l'aide d'un partage de fichiers SMB qui peut être intégré à votre [domaine Active Directory](#). Cela vous permet de contrôler l'accès à votre partage, de tirer parti des comptes de service existants et d'accéder à Amazon S3 à l'aide d'un protocole de fichiers commun axé sur Microsoft. Pour les comptes qui ne disposent pas d'une connectivité directe à un contrôleur de domaine, vous pouvez utiliser le [connecteur Active Directory](#) pour faciliter la communication avec Active Directory sur site ou dans le cloud. Pour configurer les paramètres Active Directory sur la passerelle, vous devez spécifier le connecteur Active Directory IPs pour que le contrôleur de domaine transmette les demandes par proxy à Active Directory.

Le schéma suivant montre l'architecture d'une solution basée sur S3 Intelligent-Tiering.



Par défaut, les fichiers de sauvegarde écrits dans le compartiment S3 utilisent le niveau Standard. Pour convertir les fichiers de sauvegarde du niveau Standard au niveau S3 Intelligent-Tiering, vous devez [créer une](#) règle de cycle de vie. Vous pouvez également utiliser le [AWS Management Console](#) pour activer la hiérarchisation intelligente S3. Pour plus d'informations, consultez [Getting Started Using Amazon S3 Intelligent-Tiering dans la documentation](#). AWS

Zone de disponibilité unique

Pour créer un système de fichiers à zone de disponibilité unique, choisissez l'option mono-AZ lorsque vous [créez un système de fichiers FSx pour Windows File Server](#). Amazon effectue FSx également quotidiennement des sauvegardes hautement durables (stockées dans Amazon S3) de votre système de fichiers à l'aide du service Windows Volume Shadow Copy, et vous permet d'effectuer des sauvegardes supplémentaires à tout moment. Gardez à l'esprit certains des problèmes liés à l'utilisation d'une zone de disponibilité unique. Par exemple, le partage de fichiers SMB devient inaccessible si une zone de disponibilité affectée dans laquelle le système de fichiers est approvisionné tombe en panne pendant des heures d'affilée. Si vous avez besoin d'accéder aux données, vous devez les restaurer à partir de sauvegardes dans une zone de disponibilité disponible au sein de la région source. Pour plus d'informations, consultez la section [Utiliser une seule zone de disponibilité](#) de ce guide.

Backup vers une URL

Pour SQL Server 2022, la fonctionnalité [de sauvegarde vers une URL](#) permet une sauvegarde directe sur Amazon S3. Il s'agit de l'approche de sauvegarde idéale pour l'exécution de SQL Server 2022, AWS car vous bénéficiez de l'ensemble complet des fonctionnalités d'Amazon S3 au niveau de la couche de stockage et vous supprimez le coût de l' AWS Storage Gateway appliance nécessaire dans les versions précédentes pour faciliter cette fonctionnalité. Deux coûts principaux doivent être pris en compte lors de la mise en œuvre de cette fonctionnalité : les coûts de transfert de données et la classe de stockage S3 choisie. Si vous souhaitez bénéficier des fonctionnalités natives de reprise après sinistre d'Amazon S3, vous devez tenir compte du fait que [la réplication entre régions](#) entraîne des coûts de sortie de [données](#) entre régions. Pour en savoir plus sur la configuration de cette option, consultez le billet [Backup SQL Server database to Amazon S3](#) publié sur le AWS blog Microsoft Workloads on.

Ressources supplémentaires

- [Options de sauvegarde et de restauration pour SQL Server sur Amazon EC2](#) (directives AWS prescriptives)
- [Point-in-time restauration et sauvegarde continue pour Amazon RDS avec AWS Backup](#) (AWS Storage Blog)
- [Protégez vos charges de travail SQL Server à l' NetApp SnapCenter aide d'Amazon FSx for NetApp ONTAP](#) (AWS Storage Blog)
- [Commencer à utiliser Amazon S3 Intelligent-Tiering](#) (AWS Getting Started Resource Center)

- [Stratégies de sauvegarde et de restauration pour Amazon RDS for SQL Server](#) AWS (blog de base de données)
- [Migrer une base de données Microsoft SQL Server sur site vers Amazon EC2 AWS \(directives prescriptives\)](#)
- [Bonnes pratiques pour le déploiement de Microsoft SQL Server sur Amazon EC2 \(livre blanc\)](#) AWS

Moderniser les bases de données SQL Server

Présentation de

Si vous vous lancez dans la modernisation des bases de données existantes à des fins d'évolutivité, de performance et d'optimisation des coûts, vous êtes peut-être confronté à des défis liés aux bases de données commerciales telles que SQL Server. Les bases de données commerciales sont coûteuses, bloquent les clients et proposent des conditions de licence punitives. Cette section fournit une présentation détaillée des options de migration et de modernisation de SQL Server vers des bases de données open source, ainsi que des informations sur le choix de l'option la mieux adaptée à votre charge de travail.

Vous pouvez refactoriser vos bases de données SQL Server en bases de données open source telles qu'Amazon Aurora PostgreSQL afin de réduire les coûts de licence Windows et SQL Server. Les bases de données modernes basées sur le cloud, telles qu'Aurora, allient la flexibilité et le faible coût des bases de données open source aux fonctionnalités robustes destinées aux entreprises des bases de données commerciales. Si vous avez des charges de travail variables ou des charges de travail mutualisées, vous pouvez également migrer vers [Aurora Serverless V2](#). Cela peut réduire les coûts jusqu'à 90 %, en fonction des caractéristiques de la charge de travail. AWS propose également des fonctionnalités telles que [Babelfish pour Aurora PostgreSQL](#), des outils tels que [AWS Schema Conversion Tool \(AWS SCT\)](#) et des services tels que [AWS Database Migration Service \(AWS DMS\)](#) pour simplifier la migration et la modernisation des bases de données SQL Server sur AWS.

Offres de bases de données

La migration de SQL Server sous Windows vers une base de données open source telle qu'Amazon Aurora, Amazon RDS for MySQL ou Amazon RDS for PostgreSQL peut permettre de réaliser d'importantes économies sans compromettre les performances ou les fonctionnalités. Éléments à prendre en compte :

- Le passage de l'édition SQL Server Enterprise sur Amazon EC2 à Amazon RDS for PostgreSQL ou Amazon RDS for MySQL peut permettre de réaliser des économies allant jusqu'à 80 %.
- Le passage de l'édition SQL Server Enterprise sur Amazon EC2 à l'édition compatible Amazon Aurora PostgreSQL ou à l'édition compatible Amazon Aurora MySQL peut entraîner des économies de coûts allant jusqu'à 70 %.

Pour les charges de travail de base de données traditionnelles, Amazon RDS pour PostgreSQL et Amazon RDS for MySQL répondent aux exigences et fournissent une solution rentable pour les bases de données relationnelles. Aurora ajoute de nombreuses fonctionnalités de disponibilité et de performance auparavant réservées aux fournisseurs commerciaux onéreux. Les fonctionnalités de résilience d'Aurora constituent un coût supplémentaire. Toutefois, par rapport aux fonctionnalités similaires proposées par d'autres fournisseurs commerciaux, les coûts de résilience d'Aurora restent inférieurs à ceux des logiciels commerciaux pour le même type de fonctionnalités. L'architecture Aurora est optimisée pour apporter des améliorations significatives en termes de performances par rapport aux déploiements standard de MySQL et PostgreSQL.

Aurora étant compatible avec les bases de données open source PostgreSQL et MySQL, la portabilité présente un avantage supplémentaire. Que la meilleure option soit Amazon RDS pour PostgreSQL, Amazon RDS pour MySQL ou Aurora, il faut comprendre les exigences de l'entreprise et associer les fonctionnalités nécessaires à la meilleure option.

Comparaison entre Amazon RDS et Aurora

Le tableau suivant résume les principales différences entre Amazon RDS et Amazon Aurora.

Catégorie	Amazon RDS pour PostgreSQL ou Amazon RDS pour MySQL	Aurora PostgreSQL ou Aurora MySQL
Performance	Bonne performance	Performances multipliées par 3 ou plus
Basculement	Généralement 60 à 120 secondes*	Généralement 30 secondes
Capacité de mise à l'échelle	Jusqu'à 5 répliques lues Retard en secondes	Jusqu'à 15 répliques de lecture

Catégorie	Amazon RDS pour PostgreSQL ou Amazon RDS pour MySQL	Aurora PostgreSQL ou Aurora MySQL
		Retard en millisecondes
Stockage	Jusqu'à 64 To	Jusqu'à 128 To
Stockage HA	Multi-AZ avec une ou deux unités de secours, chacune avec copie de base de données	6 copies de données réparties sur 3 zones de disponibilité par défaut
Sauvegarde	Sauvegardes quotidiennes des instantanés et des journaux	Sauvegarde continue et asynchrone vers Amazon S3
Innovations avec Aurora	NA	100 Go Clonage rapide de bases de données
	Répliques de lecture à mise à l'échelle automatique	
	Gestion de plans de requêtes	
	Aurora sans serveur	
	Répliques interrégionales avec une base de données mondiale	
	Gestion du cache du cluster**	
	Requête parallèle	
	Flux d'activité de base de données.	

*Les transactions importantes peuvent augmenter les temps de basculement

**Disponible dans Aurora PostgreSQL

Le tableau suivant indique le coût mensuel estimé des différents services de base de données couverts dans cette section.

Service de base de données	Coût en dollars américains par mois*	Calculateur de tarification AWS (nécessite Compte AWS)
Édition Amazon RDS pour SQL Server Enterprise	3 750\$	Estimate
Édition standard d'Amazon RDS pour SQL Server	2 318\$	Estimate
Édition SQL Server Enterprise sur Amazon EC2	2 835\$	Estimate
Édition standard de SQL Server sur Amazon EC2	1 345\$	Estimate
Amazon RDS pour PostgreSQL	742\$	Estimate
Amazon RDS for MySQL	712\$	Estimate
Aurora PostgreSQL	1 032\$	Estimate
Aurora MySQL	1 031\$	Estimate

* Le prix du stockage est inclus dans le prix de l'instance. Les coûts sont basés sur la us-east-1 région. Le débit et les IOPS sont des hypothèses. Les calculs concernent les instances r6i.2xlarge et r6g.2xlarge.

Recommandations en matière d'optimisation des coûts

Les migrations de bases de données hétérogènes nécessitent généralement la conversion du schéma de base de données du moteur de base de données source vers le moteur de base de

données cible et la migration des données de la base de données source vers la base de données cible. La première étape de la migration consiste à évaluer et à convertir le schéma et les objets de code SQL Server vers le moteur de base de données cible.

Vous pouvez utiliser le [AWS Schema Conversion Tool \(AWS SCT\)](#) pour évaluer la compatibilité de la base de données avec diverses options de base de données open source cibles, telles qu'Amazon RDS for MySQL ou Amazon RDS pour PostgreSQL, Aurora MySQL et PostgreSQL. Vous pouvez également utiliser l'outil Babelfish Compass pour évaluer la compatibilité avec Babelfish pour Aurora PostgreSQL. Cela fait de Compass AWS SCT et de Compass de puissants outils pour comprendre le travail initial nécessaire avant de décider d'une stratégie de migration. Si vous décidez de continuer, AWS SCT automatise les modifications nécessaires au schéma. La philosophie de base de Babelfish Compass est de permettre à la base de données SQL de migrer vers Aurora sans ou très peu de modifications. Compass évaluera la base de données SQL existante pour déterminer si cela est possible. Ainsi, le résultat est connu avant que des efforts ne soient consacrés à la migration des données de SQL Server vers Aurora.

AWS SCT automatise la conversion et la migration du schéma et du code de base de données vers le moteur de base de données cible. Vous pouvez utiliser Babelfish for Aurora PostgreSQL pour migrer votre base de données et votre application de SQL Server vers Aurora PostgreSQL sans modification de schéma ou avec un minimum de modifications. Cela peut accélérer vos migrations.

Une fois le schéma migré, vous pouvez l'utiliser AWS DMS pour migrer les données. AWS DMS peut effectuer le chargement complet des données et répliquer les modifications pour effectuer la migration avec un temps d'arrêt minimal.

Cette section explore les outils suivants de manière plus détaillée :

- AWS Schema Conversion Tool
- Babelfish for Aurora PostgreSQL
- Boussole Babelfish
- AWS Database Migration Service

AWS Schema Conversion Tool

Vous pouvez l'utiliser AWS SCT pour évaluer vos bases de données SQL Server existantes et évaluer la compatibilité avec Amazon RDS ou Aurora. Pour simplifier le processus de migration, vous pouvez également convertir le schéma d' AWS SCT un moteur de base de données à un autre

dans le cadre d'une migration de base de données hétérogène. Vous pouvez l'utiliser AWS SCT pour évaluer votre application et convertir le code d'application intégré pour les applications écrites en C#, C++, Java et dans d'autres langages. Pour plus d'informations, consultez la section [Conversion du code SQL d'une application AWS SCT à l'aide](#) de la AWS SCT documentation.

AWS SCT est un AWS outil gratuit qui prend en charge de nombreuses [sources](#) de base de données. Pour l'utiliser AWS SCT, vous le pointez vers la base de données source, puis vous exécutez une évaluation. [AWS SCT](#) Évalue ensuite le schéma et génère le rapport d'évaluation. Les rapports d'évaluation incluent un résumé, la complexité et les efforts de migration, les moteurs de base de données cibles appropriés et des recommandations de conversion. Pour le télécharger AWS SCT, consultez la section [Installation, vérification et mise à jour AWS SCT](#) dans la AWS SCT documentation.

Le tableau suivant présente un exemple de résumé analytique généré par AWS SCT pour montrer la complexité de la modification de la base de données vers différentes plateformes cibles.

Plateforme cible	Modifications automatiques ou minimales			Actions complexes			
	Objets de rangement	Objets de code	Actions de conversion	Objets de rangement		Objets de code	
Amazon RDS for MySQL	60 (98 %)	8 (35 %)	42	(12 %)	1	15 (65 %)	56
Amazon Aurora MySQL-Compatible Edition	60 (98 %)	8 (35 %)	42	(12 %)	1	15 (65 %)	56
Amazon RDS pour	60 (98 %)	12 (52 %)	54	(12 %)	1	11 (48 %)	26

PostgreSQL

L							
Amazon Aurora PostgreSQL Compatible Edition	60 (98 %)	12 (52 %)	54	(12 %)	1	11 (48 %)	26
Amazon RDS for MariaDB	60 (98 %)	7 (30 %)	42	(12 %)	1	16 (70 %)	58
Amazon Redshift	(100 %)	9 (39 %)	124	0 (0 %)	0	14 (61 %)	25
AWS Glue	0 (0 %)	17 (100 %)	0	0 (0 %)	0	0 (0 %)	0
Babelfish	59 (97 %)	10 (45 %)	20	23 (23 %)	2	12 (55 %)	30

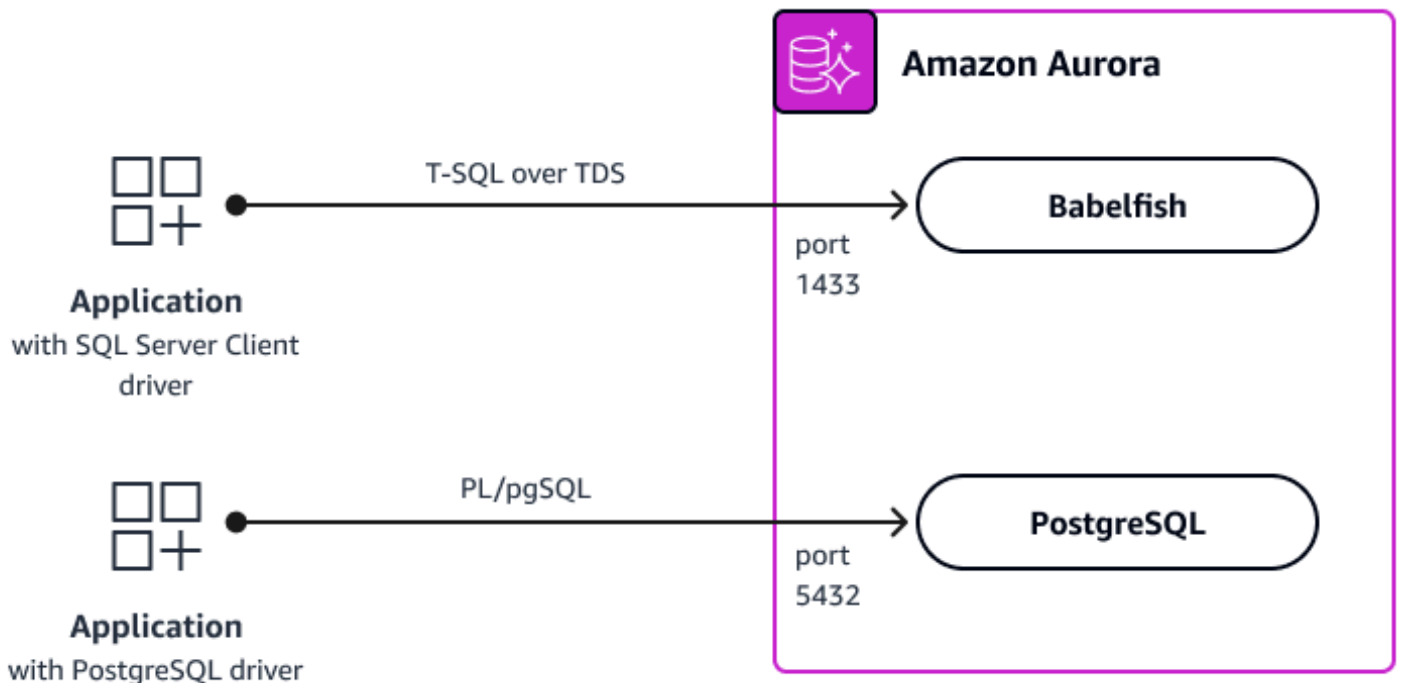
Un AWS SCT rapport fournit également des détails sur les éléments du schéma qui ne peuvent pas être convertis automatiquement. Vous pouvez combler les écarts AWS SCT de conversion et optimiser les schémas cibles en vous référant aux [playbooks de AWS migration](#). Il existe de nombreux manuels de migration de bases de données destinés à faciliter les migrations hétérogènes.

Babelfish for Aurora PostgreSQL

Babelfish for Aurora PostgreSQL étend la capacité d'Aurora PostgreSQL à accepter des connexions de base de données provenant de clients SQL Server. Babelfish permet aux applications initialement conçues pour SQL Server de fonctionner directement avec Aurora PostgreSQL, avec peu de modifications de code et sans modifier les pilotes de base de données. Babelfish rend Aurora PostgreSQL bilingue afin qu'Aurora PostgreSQL puisse fonctionner à la fois avec le T-SQL et les langages. PL/pgSQL Babelfish minimise les efforts de migration de SQL Server vers Aurora PostgreSQL. Cela accélère les migrations, minimise les risques et réduit les coûts de migration

de manière significative. Vous pouvez continuer à utiliser T-SQL après les migrations, mais il est également [possible d'utiliser les outils natifs de PostgreSQL pour le développement](#).

Le schéma suivant illustre comment une application utilisant T-SQL se connecte au port par défaut 1433 dans SQL Server et utilise le traducteur BabelFish pour communiquer avec la base de données Aurora PostgreSQL, tandis qu'une application utilisant PL/pgSQL peut se connecter directement et simultanément à la base de données Aurora PostgreSQL en utilisant le port par défaut 5432 dans Aurora PostgreSQL.



BabelFish ne supporte pas certaines fonctionnalités T-SQL de SQL Server. C'est pourquoi Amazon fournit des outils d'évaluation pour line-by-line analyser vos instructions SQL et déterminer si l'une d'entre elles n'est pas prise en charge par BabelFish.

Il existe deux options pour les évaluations de BabelFish. AWS SCT peut évaluer la compatibilité de votre base de données SQL Server avec BabelFish. Une autre option est l'outil BabelFish Compass, qui est une solution recommandée car l'outil Compass est mis à jour conformément aux nouvelles versions de BabelFish pour Aurora PostgreSQL.

Boussole BabelFish

[BabelFish Compass](#) est un outil téléchargeable gratuitement qui s'aligne sur la dernière version de BabelFish pour Aurora PostgreSQL. En revanche, les nouvelles versions de BabelFish AWS SCT seront prises en charge après un certain temps. [BabelFish Compass](#) est exécuté selon le schéma

de base de données SQL Server. Vous pouvez également extraire le schéma de base de données source SQL Server à l'aide d'outils tels que SQL Server Management Studio (SSMS). Ensuite, vous pouvez exécuter le schéma via Babelfish Compass. Cela génère le rapport détaillant la compatibilité du schéma SQL Server avec Babelfish et indiquant si des modifications sont nécessaires avant la migration. L'outil Babelfish Compass peut également automatiser bon nombre de ces changements et, à terme, accélérer vos migrations.

Une fois l'évaluation et les modifications terminées, vous pouvez migrer le schéma vers Aurora PostgreSQL à l'aide des outils natifs de SQL Server tels que SSMS ou sqlcmd. Pour obtenir des instructions, consultez l'article [Migrer de SQL Server vers Amazon Aurora à l'aide de Babelfish](#) sur le blog de AWS base de données.

AWS Database Migration Service

Une fois le schéma migré, vous pouvez utiliser AWS Database Migration Service (AWS DMS) pour effectuer la migration des données AWS avec un temps d'arrêt minimal. AWS DMS non seulement effectue un chargement complet des données, mais reproduit également les modifications de la source à la destination pendant que le système source est opérationnel. Une fois les bases de données source et cible synchronisées, l'activité de transition peut avoir lieu lorsque l'application est pointée vers la base de données cible qui termine la migration. AWS DMS effectue actuellement le chargement complet des données avec Babelfish uniquement pour une cible Aurora PostgreSQL et ne réplique pas les modifications. Pour plus d'informations, consultez la section [Utilisation de Babelfish comme cible AWS Database Migration Service dans la](#) AWS DMS documentation.

AWS DMS peut effectuer des migrations homogènes (sur le même moteur de base de données) et hétérogènes (entre différents moteurs de base de données). AWS DMS prend en charge de nombreux moteurs de base de données source et de destination. Pour plus d'informations, consultez l'article sur la [migration de votre base de données SQL Server vers Amazon RDS for SQL Server à AWS DMS l'aide du](#) blog de base de AWS données.

Ressources supplémentaires

- [Au revoir Microsoft SQL Server, bonjour Babelfish \(blog d'AWS actualités\)](#)
- [Convertissez les schémas de base de données et le code SQL de l'application à l'aide de la AWS Schema Conversion Tool CLI \(AWS Database Blog\)](#)
- [Migrez SQL Server vers Amazon Aurora PostgreSQL en utilisant les meilleures pratiques et les leçons apprises sur le terrain AWS \(blog de base de données\)](#)

- [Valider les objets de base de données après la migration de Microsoft SQL Server vers Amazon RDS for PostgreSQL et Amazon Aurora PostgreSQL](#) (blog de base de données)AWS

Optimisation du stockage pour SQL Server

Présentation de

Cette section se concentre sur l'optimisation des coûts pour le stockage SSD Amazon Elastic Block Store (Amazon EBS) pour SQL Server sur des charges de travail EC2.

Vous disposez d'un large éventail d'options de stockage pour déployer et exécuter des charges de travail SQL Server. AWS La sélection du stockage approprié doit être basée sur l'objectif, l'architecture, la durabilité, les performances, la capacité et le coût. AWS les clients exécutant des charges de travail SQL Server utilisent généralement une combinaison de solutions de stockage NVMe Amazon EBS FSx, Amazon et Amazon Simple Storage Service (Amazon S3).

Amazon EBS est un système de stockage rattaché au réseau connecté à des instances de calcul EC2 et utilisé pour stocker et traiter le système d'exploitation général, les applications, les bases de données et les fichiers de sauvegarde. Le stockage sur disque SSD Amazon EBS inclut des SSD à usage général (gp2 et gp3) et des SSD à IOPS provisionnés (io1, io2 et IO2Bx). Éléments à prendre en compte :

- Certaines instances EC2, telles que r5d, sont NVMe SSDs physiquement attachées localement à l'instance hôte. Ces volumes fournissent un stockage au niveau des blocs couramment utilisé pour SQL Server tempdb ou pour l'extension du pool de mémoire tampon.
- Amazon FSx for Windows File Server est un service de stockage de fichiers entièrement géré, tandis qu'Amazon FSx NetApp for ONTAP est un service de stockage partagé entièrement géré basé NetApp sur le célèbre système de fichiers ONTAP. Amazon FSx est fréquemment utilisé pour exécuter des charges de travail SQL Server dans une configuration FCI (SQL Server Failover Clustered Instance) à haute disponibilité. Cette solution héberge les données et les fichiers journaux SQL Server, ce qui réduit les exigences de performances EBS sur les instances EC2.
- Amazon S3 est un service de stockage d'objets offrant une évolutivité, une disponibilité des données, une sécurité et des performances de pointe. Vous pouvez stocker des fichiers de sauvegarde natifs de SQL Server AMIs, des instantanés EBS, des journaux d'applications, etc. sur Amazon S3.

Types de stockage SSD, performances et coûts pour Amazon EBS

Les coûts de stockage SSD pour Amazon EBS augmentent généralement à mesure que la durabilité et les performances augmentent. Le stockage est actuellement disponible en cinq types de volumes, chacun ayant ses [propres indicateurs de performance uniques](#). Pour un résumé des cas d'utilisation et des caractéristiques des volumes sauvegardés sur SSD, consultez le tableau de la section [Volumes SSD](#) de la documentation Amazon EBS.

Vous pouvez utiliser Amazon CloudWatch pour surveiller les performances des SSD, recueillir des données sur les tendances et définir des alarmes lorsque certains seuils sont atteints. Si vous exécutez des charges de travail SQL Server AWS, pensez à activer la [surveillance détaillée](#) et à déployer des [mesures CloudWatch personnalisées pour capturer des indicateurs](#) de performance détaillés sur les volumes, tels que la latence du disque, les IOPS, le débit, la longueur de la file d'attente du disque, la capacité utilisée par rapport à la capacité libre, etc. Vous pouvez utiliser ces indicateurs de CloudWatch performance pour identifier le stockage sous-provisionné ou surprovisionné et fournir des points de données historiques pour définir avec précision les exigences de stockage.

Les coûts de stockage SSD pour Amazon EBS varient également en fonction de la capacité allouée. Le tableau ci-dessous présente une comparaison des différents types de volumes. Tous les types de volumes ont une capacité de 1 To et des configurations de performances similaires.

Type de volume	Nombre maximal d'E/S par seconde (16 Kio d'E/S)	Débit maximal (128 KiB I/O)	Prix par 1 To	Pourcentage d'économies
gp2	3 000	250	102,40\$	
gp3	3 000	250	86,92\$	15 %
io1	16,000	500	1 168\$	
io2	16,000	500	1 168\$	
gp3	16,000	500	146,92\$	87 %
io2bx	16,000	4 000	1 168\$	

Type de volume	Nombre maximal d'E/S par seconde (16 Kio d'E/S)	Débit maximal (128 KiB I/O)	Prix par 1 To	Pourcentage d'économies
gp3	16,000	1 000	181,92\$	84 %

Note

Les indicateurs de performance et de coût présentés dans le tableau précédent sont par volume, sur la base d'une [estimation](#) du Calculateur de tarification AWS. Un Compte AWS est nécessaire pour accéder à l'estimation dans le Calculateur de tarification AWS.

Les volumes Amazon EBS SSD gp3 offrent d'excellentes performances à faible coût. Vous pouvez économiser jusqu'à 87 % si vous choisissez un volume gp3 plutôt que des volumes io1 ou io2 pour les charges de travail nécessitant moins de 16 000 IOPS et un débit de 500. MiBps

Les volumes io2 Block Express (IO2Bx) offrent des performances accrues par rapport aux volumes io2 classiques. À 16 000 IOPS, les volumes io1 ou io2 ne peuvent atteindre qu'un MiBps débit de 500, tandis que les volumes IO2Bx peuvent être configurés jusqu'à 4 000 débits. MiBps Comparés aux volumes io1 et io2, les volumes io2Bx fournissent un débit quatre fois supérieur, entre 16 000 et 64 000 IOPS, exactement au même prix. Les volumes io2 ordinaires peuvent être convertis en volumes IO2bx en les attachant à des instances EC2 prises en charge par IO2BX. Pour obtenir la liste des instances EC2 prises en charge par IO2BX, consultez la section [Volumes SSD IOPS provisionnés](#) dans la documentation Amazon EBS. Avant de déployer un nouveau système de stockage, vous pouvez utiliser le [Calculateur de tarification AWS](#) pour estimer votre coût mensuel et comprendre l'impact sur les coûts en fonction des compromis entre durabilité, performances et capacité.

Optimisation générale des coûts des SSD pour Amazon EBS

Nous vous recommandons d'évaluer ce que vous stockez et de vous assurer que vous utilisez le type et la classe de stockage appropriés. Par exemple, Amazon S3 propose un excellent rapport qualité-prix, des politiques de cycle de vie intégrées et des options de réplication idéales pour les sauvegardes SQL Server. SQL Server 2022 a la capacité de sauvegarder directement sur Amazon S3, tandis que les versions précédentes de SQL Server reposent sur des sauvegardes locales

natives. Si vous utilisez d'anciennes versions de SQL Server, pensez à effectuer une sauvegarde sur des volumes de disque dur Amazon EBS, puis à copier la sauvegarde sur Amazon S3. Cette solution permet d'économiser 53 % au lieu d'utiliser des volumes gp3 pour les sauvegardes.

Le tableau suivant indique la différence de prix pour 1 To de stockage sur Amazon EBS gp3, Amazon EBS HDD st1 et Amazon S3.

Type de stockage	Capacity	Prix par minute
EBS GP3 500 MiBps	1 To	96,92\$
EBS ST1 Burst 500 MiBps		46,08\$
S3 Standard		23,55\$
S3 Standard (accès peu fréquent)		12,80\$
S3 Glacier Deep Archive		1,03\$

Note

Les indicateurs de coûts présentés dans le tableau précédent sont basés sur une [estimation figurant](#) dans le Calculateur de tarification AWS. Un Compte AWS est nécessaire pour accéder à l'estimation dans le Calculateur de tarification AWS.

Nous vous recommandons de prendre en compte les points suivants :

- Activez une surveillance détaillée et déployez CloudWatch des métriques personnalisées pour saisir avec précision leurs exigences en matière de performances de stockage.
- Mettez à niveau le stockage Amazon EBS de gp2 à gp3 pour réduire les coûts, augmenter la flexibilité et améliorer les performances.
- Mettez à niveau le stockage Amazon EBS de io1 à io2 pour une durabilité et une flexibilité accrues en termes de performances.
- Utilisez io2Bx au lieu de io1 ou io2 lorsque cela est possible pour une durabilité et des performances accrues.

- Lorsque vous choisissez une mix-and-match solution de stockage, envisagez une approche qui contribue à réduire les besoins en capacité et le coût des volumes à hautes performances. Par exemple, vous pouvez utiliser des volumes gp3 économiques pour votre volume racine (système d'exploitation), l'installation de SQL Server, les bases de données système (à l'exception de tempdb) et les bases de données utilisateur moins performantes. Cela pourrait contribuer à réduire la capacité et le coût d'un volume io2, qui peut être dédié à des bases de données utilisateur à hautes performances.
- Si vous hébergez des bases de données SQL Server sur AWS, nous vous recommandons d'utiliser plusieurs fichiers de données SQL Server par base de données. Cela permet de répartir les read/write charges de travail sur plusieurs volumes, réduisant ainsi les exigences de performance et de capacité par volume et, par conséquent, les coûts.
- Même si les charges de travail de production nécessitent un stockage plus performant, tel que io1 ou io2/io2Bx, considérez les volumes gp3 pour les charges de travail hors production afin de réduire les coûts.
- Suivez et évaluez l'utilisation du stockage au fil du temps afin d'identifier facilement les pics d'utilisation et les coûts imprévus.
- [Optimiseur de calcul AWS](#) À utiliser pour des recommandations sur l'augmentation ou la baisse des volumes EBS en fonction de l'utilisation réelle.
- Utilisez l'élasticité de AWS pour ajuster les besoins en performances et en capacité de vos volumes SSD pour Amazon EBS. Contrairement aux environnements sur site, vous n'avez pas besoin de surprovisionner les performances et la capacité de stockage pour les charges de travail futures. Vous pouvez migrer vos charges de travail SQL Server existantes AWS et ajuster les performances ou la capacité selon vos besoins, tout en maintenant vos bases de données en ligne.

Ressources supplémentaires

- [Types de volumes Amazon EBS](#) (documentation Amazon EBS)
- [Amazon Elastic Block Store \(Amazon EBS\) \(documentation Amazon EBS\)](#)
- [Volumes SSD IOPS provisionnés \(documentation Amazon EBS\)](#)
- [Volumes de stockage d'instances SSD](#) (documentation Amazon EC2)
- [CloudWatch Métriques Amazon pour Amazon EBS](#) (documentation Amazon EBS)
- [Spécifications relatives aux instances optimisées pour le stockage Amazon EC2 \(documentation Amazon EC2\)](#)

- [Protégez vos charges de travail SQL Server à l' NetApp SnapCenter aide d'Amazon FSx for NetApp ONTAP](#) (AWS Storage Blog)
- [FAQ Amazon EC2 \(pageAWS produit\)](#)

Optimisez les licences SQL Server à l'aide de Compute Optimizer

Conseils sur la façon d'optimiser les licences pour SQL Server à l'aide de Optimiseur de calcul AWS.

Présentation de

[Optimiseur de calcul AWS](#) peut recommander des opportunités d'optimisation des licences pour les charges de travail Microsoft SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2). Compute Optimizer peut fournir des recommandations automatisées pour réduire les coûts de licence. Les recommandations de Compute Optimizer sont répertoriées à côté de chacune de vos instances EC2 avec des licences Microsoft SQL Server. Les informations fournies incluent les opportunités d'économie recommandées, les prix des instances EC2 à la demande et les prix horaires des licences BYOL (Bring Your Own License). Ces informations peuvent vous aider à décider si vous devez rétrograder votre édition de licence.

Compute Optimizer découvre automatiquement vos instances SQL Server sur Amazon EC2 par type de charge de travail déduit. Pour consulter les recommandations relatives aux licences, vous pouvez sélectionner les instances SQL Server dans Compute Optimizer, puis vous authentifier auprès d'[Amazon CloudWatch Application Insights](#) à l'aide de vos informations d'identification de base de données en lecture seule. Compute Optimizer analyse si vous utilisez des fonctionnalités de l'édition Enterprise de SQL Server. Si aucune fonctionnalité de l'édition Enterprise n'est utilisée, Compute Optimizer vous recommande de passer à l'édition Standard afin de réduire les coûts de licence.

Vous pouvez également utiliser Compute Optimizer pour faire des recommandations de dimensionnement pour vos instances Amazon EC2 qui exécutent des charges de travail SQL Server. Pour plus d'informations, consultez la section [Optimiser le dimensionnement de SQL Server à l'aide de Compute Optimizer](#) dans ce guide.

Recommandations en matière d'optimisation des coûts

Les recommandations de licence de Compute Optimizer peuvent vous aider à évaluer les fonctionnalités que vous utilisez dans Microsoft SQL Server et à choisir l'édition la plus rentable pour vos charges de travail. L'édition Enterprise de SQL Server est nettement plus chère que l'édition

Standard. Pour plus d'informations, consultez la section [Comparaison des éditions de SQL Server](#) dans ce guide et consultez la [tarification de SQL Server 2022](#) sur le site Web de Microsoft. En investissant du temps dans la configuration de Compute Optimizer afin d'évaluer votre parc SQL Server et de fournir des recommandations, vous pouvez réduire considérablement vos coûts de licence.

La page des détails de la licence fournit les informations suivantes :


- Utilisez le tableau pour comparer vos paramètres de licence actuels (tels que l'édition, le modèle et le nombre de cœurs d'instance) avec les recommandations de Compute Optimizer.
- Utilisez les graphiques d'utilisation pour examiner le nombre de fonctionnalités de l'édition Enterprise qui ont été utilisées pendant la période d'analyse.

Pour plus d'informations, consultez la section [Affichage des détails d'une recommandation de licence logicielle commerciale](#) dans la documentation de Compute Optimizer.

Configurer Compute Optimizer

Compute Optimizer analyse les licences logicielles commerciales à l'aide de cette métrique. `mssql_enterprise_features_used` Pour plus d'informations sur cette métrique, consultez la section [Mesures pour les licences logicielles commerciales](#).

1. Assurez-vous que vous disposez des autorisations appropriées pour vous inscrire à Compute Optimizer. Pour plus d'informations, consultez les ressources suivantes :
 - [Politique d'adhésion à Compute Optimizer](#)
 - [Politiques pour accorder l'accès à Compute Optimizer en mode autonome Comptes AWS](#)
 - [Politiques permettant d'accorder l'accès à Compute Optimizer pour le compte de gestion d'une organisation](#)
2. Attachez les rôles d'instance et la politique requis pour CloudWatch Application Insights. Pour obtenir des instructions, consultez la section [Politiques visant à activer les recommandations relatives aux licences logicielles commerciales](#).
3. Activez CloudWatch Application Insights en utilisant les informations d'identification de votre base de données Microsoft SQL Server. Pour obtenir des instructions, consultez la section [Configuration de l'application pour la surveillance](#) dans la CloudWatch documentation.

 Note


Pour générer des recommandations pour les licences logicielles commerciales, au moins 30 heures consécutives de données CloudWatch métriques sont nécessaires. Pour plus d'informations, consultez la section [Exigences relatives aux CloudWatch métriques](#).

4. Utilisez la requête SQL suivante pour configurer l'accès avec le moindre privilège pour CloudWatch Application Insights.

```
GRANT VIEW SERVER STATE TO [LOGIN];  
GRANT VIEW ANY DEFINITION TO [LOGIN];
```

Cela active un nouveau service, PrometheusSqlExporter SQL.

5. Depuis le compte de gestion de la cible Compte AWS ou de l'organisation, optez pour Compute Optimizer. Pour obtenir des instructions, consultez [la section Connexion à votre compte](#).

 Note

Une fois que vous vous êtes inscrit, les résultats et les recommandations d'optimisation peuvent prendre jusqu'à 24 heures pour être générés.

6. Dans la [console Compute Optimizer](#), sélectionnez Licenses dans le volet de navigation.
7. Dans la colonne Conclusions, recherchez toutes les instances pour lesquelles le résultat de mesures est insuffisant. Compute Optimizer renvoie ce résultat s'il détecte qu' CloudWatch Application Insights n'est pas activé ou ne dispose pas d'autorisations suffisantes. Pour plus d'informations, consultez la section [Trouver des raisons](#). Procédez comme suit pour résoudre ces problèmes :
 - a. Choisissez l'instance.
 - b. Ajoutez un secret.
 - c. Vérifiez que le rôle et la politique de l'instance sont attachés.
 - d. Choisissez Activer les recommandations de licence.
8. Dans la colonne Résultats, recherchez toutes les instances dont le résultat est Non optimisé. Compute Optimizer renvoie ce résultat s'il détecte que votre infrastructure Amazon EC2 n'utilise aucune des fonctionnalités de licence Microsoft SQL Server pour lesquelles vous payez. Pour plus

d'informations, consultez la section [Trouver des raisons](#). Procédez comme suit pour résoudre ces problèmes :

- a. Choisissez l'instance.
- b. Comparez l'édition de licence actuelle avec l'édition recommandée.
- c. Consultez le graphique actuel d'utilisation des licences.
- d. Si vous souhaitez rétrograder la licence, choisissez Implémenter la recommandation.
- e. Passez en revue les exigences et suivez les instructions pour rétrograder la licence. Si vous souhaitez automatiser le processus, voir [Rétrograder l'édition SQL Server Enterprise à l'aide de AWS Systems Manager Document pour réduire les coûts](#) (AWS blog).

Ressources supplémentaires

- [Réduisez les coûts de licence Microsoft SQL Server grâce à Optimiseur de calcul AWS](#) (AWS Blog)
- [Qu'est-ce que c'est Optimiseur de calcul AWS ?](#) (AWS documentation)
- [Consulter les recommandations relatives aux licences logicielles commerciales](#) (AWS documentation)
- [Rétrogradez votre édition Microsoft SQL Server](#) (AWS documentation)
- [Microsoft SQL Server activé AWS](#) (AWS)
- [Licence Microsoft sur AWS](#) (AWS)
- [Tarification de Microsoft SQL Server 2019](#) (Microsoft)
- [Tarification de Microsoft SQL Server 2022](#) (Microsoft)

Optimisez le dimensionnement de SQL Server à l'aide de Compute Optimizer

Présentation de

[Optimiseur de calcul AWS](#) aide les administrateurs de base de données (DBA) à découvrir les charges de travail Microsoft SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2) et à ajuster la taille des instances EC2 afin de réduire les coûts de licence jusqu'à 25 %. La fonctionnalité de [type de charge de travail déduit](#) de Compute Optimizer utilise l'apprentissage automatique (ML) et détecte automatiquement les applications susceptibles de s'exécuter sur AWS vos ressources. Compute Optimizer inclut la prise en charge de SQL Server en tant que type de charge de travail

inférée. En utilisant la fonctionnalité de type de charge de travail inféré, vous pouvez identifier les opportunités de réduction des coûts en fonction de la charge de travail spécifique exécutée sur vos instances Amazon EC2.

Grâce à cette fonctionnalité, vous pouvez classer les opportunités de réduction des coûts en fonction des types de charge de travail déduits pris en charge, tels que SQL Server. Compute Optimizer peut détecter automatiquement les instances SQL Server EC2 surprovisionnées. Vous pouvez passer à la console EC2 pour réduire la taille de l'instance, ce qui permet de réduire les coûts de licence et d'infrastructure.

Vous pouvez également utiliser Compute Optimizer pour recommander des licences SQL Server. Pour plus d'informations, consultez la section [Optimiser les licences SQL Server à l'aide de Compute Optimizer](#) dans ce guide.

Configurer Compute Optimizer

Pour obtenir des instructions sur l'utilisation de Compute Optimizer avec des charges de travail déduites de SQL Server, consultez [Optimisation des performances et réduction des coûts de licence : tirer parti des instances SQL Server Amazon Optimiseur de calcul AWS EC2 \(blog\)](#). AWS Vous pouvez opter pour des comptes autonomes, des comptes membres d'une organisation et des comptes de gestion d'une organisation. Pour les comptes autonomes et les comptes membres, l'inscription active Compute Optimizer uniquement pour ce compte. Pour un compte de gestion d'organisation, vous pouvez choisir d'activer Compute Optimizer uniquement pour ce compte ou pour tous les comptes membres de l'organisation.

Le processus d'opt-in de Compute Optimizer crée automatiquement un rôle lié à un service Gestion des identités et des accès AWS (IAM). Pour de plus amples informations, veuillez consulter [Utilisation des rôles liés à un service pour Optimiseur de calcul AWS](#).

Compute Optimizer analyse vos ressources en fonction des CloudWatch indicateurs Amazon, tels que le processeur, les E/S, le réseau et l'utilisation d'Amazon Elastic Block Store (Amazon EBS). Pour générer des recommandations, au moins 30 heures consécutives de données CloudWatch métriques sont nécessaires au cours des 14 derniers jours. Si vous activez la fonctionnalité de métriques d'infrastructure améliorées, elle étend les métriques d'utilisation à 93 jours. Pour plus d'informations, consultez les [exigences en matière de CloudWatch métriques](#) et les [métriques d'infrastructure améliorées](#) dans la documentation de Compute Optimizer.

Compute Optimizer propose des options et les économies associées à chaque option, en fonction du vCPU, de la mémoire, du stockage, du réseau, des risques et des efforts de migration. Vous

pouvez utiliser le tableau de bord des CloudWatch métriques pour analyser les données utilisées pour faire la recommandation. Avec ces données, vous pouvez ajuster la taille de vos instances EC2 qui exécutent des charges de travail SQL Server. Pour plus d'informations sur la modification du type d'instance, consultez [Modifier le type d'instance](#) dans la documentation Amazon EC2.

Ressources supplémentaires

- [Optimiseur de calcul AWS identifie et filtre les charges de travail Microsoft SQL Server](#) (AWS)
- [Optimisation des performances et réduction des coûts de licence : tirer parti Optimiseur de calcul AWS des instances SQL Server Amazon EC2 \(blog\)](#) AWS
- [Qu'est-ce que c'est Optimiseur de calcul AWS ?](#) (AWS documentation)
- [Afficher les recommandations relatives aux instances EC2](#) (AWS documentation)

Consultez Trusted Advisor les recommandations relatives aux charges de travail SQL Server

Présentation de

[AWS Trusted Advisor](#) fournit des recommandations qui vous aident à suivre les AWS meilleures pratiques. En analysant votre utilisation, votre configuration et vos dépenses, Trusted Advisor fournit des recommandations pratiques pour réduire vos coûts, améliorer la disponibilité et les performances du système ou contribuer à combler les failles de sécurité. Cette section se concentre sur Trusted Advisor les vérifications qui peuvent vous aider à réduire les coûts d'exploitation des charges de travail SQL Server dans le AWS Cloud.

Recommandations en matière d'optimisation des coûts

Trusted Advisor fournit des recommandations qui vous aident à optimiser vos charges de travail SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2). Les contrôles inspectent vos charges de travail SQL Server et répertorient automatiquement les instances à optimiser. La mise en œuvre Trusted Advisor des recommandations peut réduire les coûts et améliorer le niveau de sécurité de votre organisation.

Les Trusted Advisor vérifications suivantes se concentrent sur Microsoft SQL Server :

- [Instances Amazon EC2 surprovisionnées pour Microsoft SQL Server](#) : cette vérification analyse vos instances Amazon EC2 qui exécutent SQL Server et vous alerte si une instance dépasse la limite

de vCPU du logiciel SQL Server. Par exemple, une instance avec l'édition standard de SQL Server peut utiliser jusqu'à 48 CPUs v. Une instance avec SQL Server Web peut utiliser jusqu'à 32 CPUs v.

Edition	vCPU min.	Nombre maximal de vCPU
Web	4	32
Standard	4	48
Enterprise	4	Limites d'OS

- [Consolidation des instances Amazon EC2 pour Microsoft SQL Server](#) : cette vérification analyse vos instances Amazon EC2 et vous avertit si votre instance possède un nombre inférieur au nombre minimum de licences SQL Server. Vous pouvez consolider des instances SQL Server plus petites afin de réduire les coûts. Si vous disposez de nombreuses petites instances SQL Server incluses dans une licence, envisagez de les consolider. Selon le [guide des licences Microsoft SQL Server 2019](#), SQL Server nécessite un minimum de 4 licences vCPU par instance. Si vous consolidez ces bases de données, vous pouvez économiser sur les coûts de licence. Vous pouvez prendre votre décision en fonction du nombre de bases de données présentes sur l'instance, de la taille maximale de la base de données et de la taille totale des bases de données. La consolidation est prise en charge pour les éditions Web, Standard et Enterprise de SQL Server. Pour plus d'informations, consultez [Consolidation des bases de données SQL Server](#) (billet de blog Microsoft).

AWS ne recommande pas de placer de grandes bases de données de production sur un seul serveur. Cependant, vous pouvez consolider les plus petits environnements utilisés pour les environnements non liés à la production, tels que le développement, les tests et la mise en scène. Cela dépend de votre utilisation actuelle de SQL Server ; si vous avez des bases de données peu utilisées, vous pouvez les consolider sur un seul serveur.

Configuration Trusted Advisor

Procédez comme suit pour évaluer les intégrations ciblées sur SQL Server Trusted Advisor.

1. Connectez-vous au AWS Management Console.
2. Ouvrez la [AWS Trusted Advisor console](#).

3. Dans le volet de navigation, sous Recommandations, sélectionnez Optimisation des coûts.
4. Dans la liste des vérifications d'optimisation des coûts, consultez l'état de la consolidation des instances Amazon EC2 pour Microsoft SQL Server et des instances Amazon EC2 surprovisionnées pour les vérifications Microsoft SQL Server.
 - Les symboles verts indiquent que vos instances Amazon EC2 sont configurées de manière optimale.
 - Les symboles d'alerte orange indiquent qu'il existe des possibilités d'amélioration.
5. Choisissez un chèque pour voir ses détails et ses recommandations.
6. Suivez les instructions fournies par le check pour optimiser vos instances Amazon EC2 qui exécutent des charges de travail SQL Server.
7. Surveillez régulièrement vos instances et actualisez régulièrement les vérifications.

Ressources supplémentaires

- [Trusted Advisor vérifier la référence](#) (AWS documentation)
- [Microsoft SQL Server activé AWS](#) (AWS)
- [Licence Microsoft sur AWS](#) (AWS)
- [Tarification de SQL Server 2019](#) (Microsoft)
- [AWS Launch Wizard pour SQL Server](#) (AWS documentation)

Containers

La modernisation est un parcours transformationnel qui offre de nombreuses options, notamment la décomposition des monolithes en microservices, la refonte de l'architecture des applications pour qu'elles soient pilotées par les événements à l'aide de fonctions sans serveur (AWS Lambda) et la réaffectation de bases de données de SQL Server vers Amazon Aurora ou des bases de données gérées spécialement pour cela. Les processus de modernisation visant à reconfigurer les applications .NET Framework vers des conteneurs Linux et Windows nécessitent moins d'efforts que les autres options de modernisation. Les conteneurs offrent les avantages suivants :

- **Accélérez l'innovation** : le passage aux conteneurs facilite l'automatisation des étapes du cycle de développement, notamment la création, le test et le déploiement d'applications. En automatisant ces processus, les équipes de développement et d'exploitation ont plus de temps pour se concentrer sur l'innovation.
- **Réduction du coût total de possession (TCO)** : le passage aux conteneurs peut également réduire votre dépendance à l'égard des outils de gestion des licences et de protection des terminaux. Les conteneurs étant des unités de calcul éphémères, vous pouvez automatiser et simplifier les tâches de gestion telles que l'application de correctifs, le dimensionnement, la sauvegarde et la restauration. Cela réduit le coût total de possession lié à l'administration et à l'exploitation des charges de travail basées sur des conteneurs. Enfin, les conteneurs sont plus efficaces que les machines virtuelles, car vous pouvez les utiliser pour optimiser le placement de vos applications en fournissant une meilleure isolation. Cela augmente l'utilisation des ressources d'infrastructure de votre application.
- **Améliorez l'utilisation des ressources** : les conteneurs sont plus efficaces que les machines virtuelles, car vous pouvez utiliser des conteneurs pour optimiser le placement de vos applications. Cela augmente l'utilisation des ressources d'infrastructure de votre application en fournissant une meilleure isolation.
- **Comblez le déficit de compétences** : AWS propose des journées d'immersion pour améliorer les compétences de vos équipes de développement en matière de technologie et de DevOps pratiques relatives aux conteneurs.

Cette section couvre les rubriques suivantes :

- [Déplacer les applications Windows vers des conteneurs](#)
- [Optimisez les coûts AWS Fargate des tâches sur Amazon ECS](#)
- [Gagnez en visibilité sur vos coûts Amazon EKS](#)

- [Replateforme des applications Windows avec App2Container](#)

Pour plus d'informations sur les licences, consultez la section Licences d'[Amazon Web Services et Microsoft : Frequently Asked Questions](#) ou envoyez vos questions par e-mail à microsoft@amazon.com.

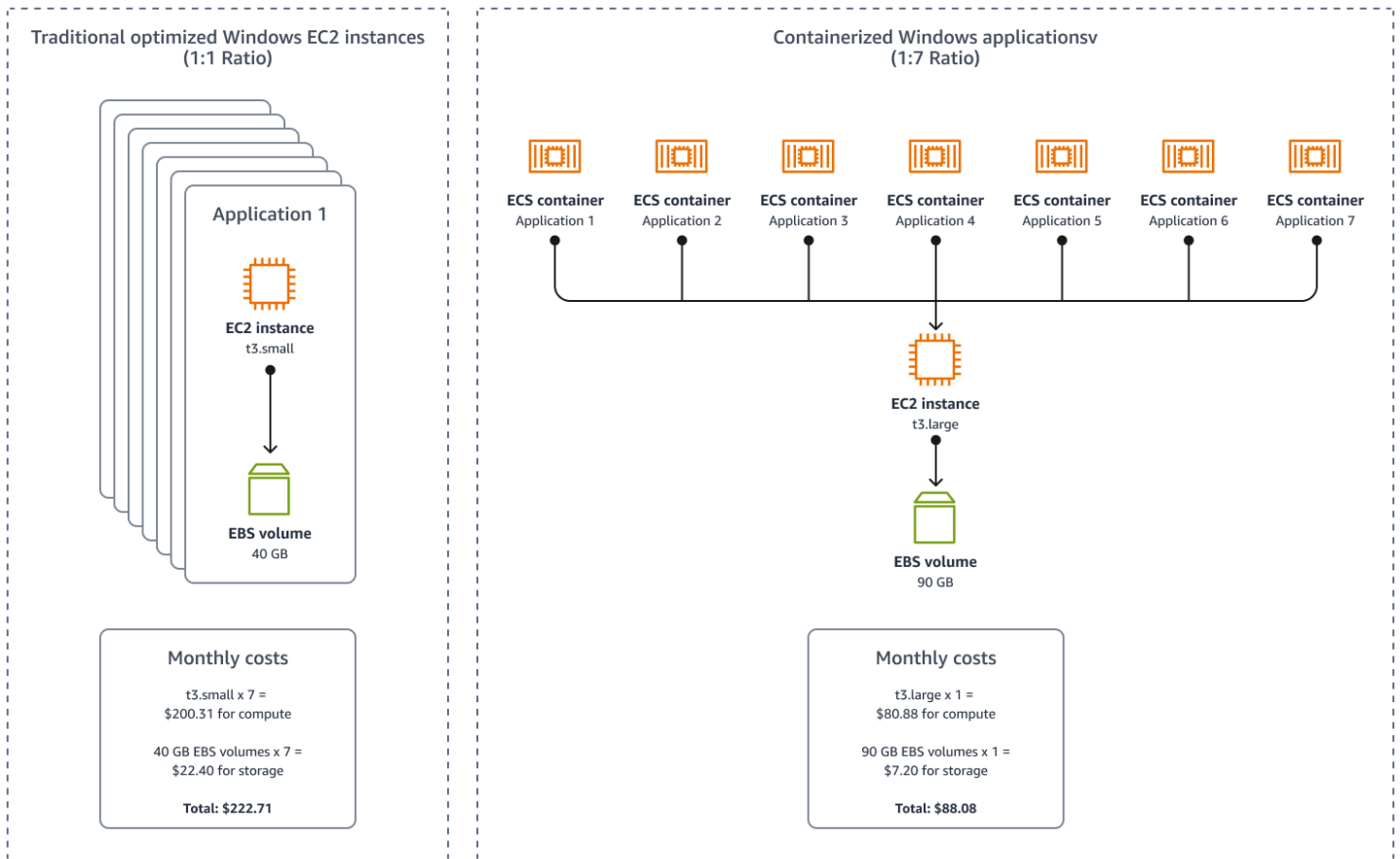
Déplacer les applications Windows vers des conteneurs

Présentation de

Selon l'[enquête annuelle 2021 de la CNCF](#), 96 % des organisations utilisent ou évaluent des conteneurs pour moderniser leur infrastructure. En effet, les conteneurs peuvent aider votre entreprise à réduire les risques, à accroître l'efficacité et la rapidité opérationnelles et à favoriser l'agilité. Vous pouvez également utiliser des conteneurs pour réduire le coût d'exécution de vos applications. Cette section propose des recommandations pour gérer des conteneurs de manière rentable sur l'ensemble des services de AWS conteneurs, notamment [Amazon Elastic Container Service \(Amazon ECS\)](#), [Amazon Elastic Kubernetes Service \(Amazon EKS\)](#) et [AWS Fargate](#)

Avantages en termes de coûts

L'infographie suivante montre les économies qu'une entreprise peut réaliser en consolidant ses applications ASP.NET Framework sur des instances Amazon Elastic Compute Cloud (Amazon EC2) sur la base d'[AWS une recommandation d'optimisation et d'évaluation des licences](#) (OLA). AWS L'infographie suivante montre les économies supplémentaires qui peuvent être réalisées en déplaçant des applications vers un conteneur Windows.



L' AWS OLA a recommandé à l'entreprise de faire un lifting et de passer aux instances individuelles de t3.small. L'entreprise pourrait réaliser ces économies en exécutant sept applications ASP.NET sur des serveurs locaux, comme le montre l'analyse d'utilisation des performances suivante.

Server name	Storage	Operating system	On-premises CPU AVG utilization	On-premises CPU peak utilization	On-premises RAM (GB)	On-premises RAM AVG utilization (GB)	On-premises RAM peak utilization (GB)	Instance size	vCPU	RAM (GB)
1 AppServer01	60	Windows Server 2012	7.00%	17.00%	8	13.50%	17.10%	t3.small	2	2
2 AppServer02	39	Windows Server 2012	20.07%	22.00%	16	7.50%	12.40%	t3.small	2	2
3 AppServer03	39	Windows Server 2012	24.00%	25.50%	16	8.80%	11.90%	t3.small	2	2
4 AppServer04	4	Windows Server 2012	21.40%	24.00%	16	7.80%	10.70%	t3.small	2	2
5 AppServer05	40	Windows Server 2012	21.30%	23.00%	16	8.20%	12.00%	t3.small	2	2
6 AppServer06	39	Windows Server 2012	21.50%	23.50%	16	7.90%	10.90%	t3.small	2	2
7 AppServer07	39	Windows Server 2012	21.60%	22.90%	16	8.40%	11.50%	t3.small	2	2

Une analyse plus approfondie a révélé que l'entreprise pourrait économiser encore davantage sur les coûts en exécutant ses charges de travail sur des conteneurs. Les conteneurs réduisent la charge du système d'exploitation sur les ressources système telles que l'utilisation du processeur, de la RAM et du disque (expliqué dans la section suivante). Dans ce scénario, l'entreprise pourrait consolider les sept applications sur une seule instance t3.large tout en disposant encore de 3 Go de RAM. La

migration vers des conteneurs peut aider l'entreprise à réaliser en moyenne 64 % d'économies sur les coûts de calcul et de stockage en utilisant des conteneurs au lieu d'Amazon EC2.

Recommandations en matière d'optimisation des coûts

La section suivante propose des recommandations pour optimiser les coûts en consolidant les applications et en utilisant des conteneurs.

Réduisez l'encombrement de Windows sur Amazon EC2

Les conteneurs Windows peuvent réduire votre empreinte Windows sur Amazon EC2 en vous permettant de consolider davantage d'applications sur un nombre réduit d'instances EC2. Supposons, par exemple, que vous disposez de 500 applications ASP.NET. Si vous exécutez un cœur par application pour Windows sur Amazon EC2, cela équivaut à 500 instances Windows (t3.small). Si vous supposez un ratio de 1:7 (qui peut augmenter de manière significative en fonction du type/de la taille de l'instance EC2) pour l'utilisation de conteneurs Windows (avec t3.large), vous n'avez besoin que d'environ 71 instances Windows. Cela représente une diminution de 85,8 % de votre empreinte Windows sur Amazon EC2.

Réduisez les coûts de licence Windows

Si vous octroyez une licence à une instance Windows, vous n'avez pas besoin de licence pour les conteneurs exécutés sur cette instance. Par conséquent, la consolidation de vos applications ASP.NET à l'aide de conteneurs Windows peut réduire considérablement vos coûts de licence Windows.

Réduisez l'encombrement de votre stockage

Chaque fois que vous lancez une nouvelle instance EC2, vous créez et payez un nouveau volume Amazon Elastic Block Store (Amazon EBS) pour héberger le système d'exploitation. Au fur et à mesure que cela évolue, le coût évolue avec lui. Si vous utilisez des conteneurs, vous pouvez réduire les coûts de stockage car tous les conteneurs partagent le même système d'exploitation de base. En outre, les conteneurs utilisent le concept de couches pour réutiliser des parties immuables d'une image de conteneur pour tous les conteneurs en cours d'exécution basés sur cette image. Dans l'exemple de scénario précédent, tous les conteneurs exécutent .NET Framework et partagent donc tous la couche intermédiaire et immuable du framework ASP.NET.

Migrer end-of-support des serveurs vers des conteneurs

Support pour Windows Server 2012 et Windows Server 2012 R2 a pris fin le 10 octobre 2023. Vous pouvez migrer vos applications exécutées sous Windows Server 2012 ou des versions antérieures en les conteneurisant pour qu'elles s'exécutent sur de nouveaux systèmes d'exploitation. Ainsi, vous évitez d'exécuter vos applications sur des systèmes d'exploitation non conformes tout en tirant parti de la rentabilité, de la réduction des risques, de l'efficacité opérationnelle, de la rapidité et de l'agilité qu'offrent les conteneurs.

Une mise en garde à prendre en compte avec cette approche est si votre application nécessite des exigences spécifiques APIs liées à la version du système d'exploitation actuellement utilisée (COM Interop, par exemple). Dans ce cas, vous devez tester le transfert de votre application vers une version plus récente de Windows. Les conteneurs Windows alignent leur image de conteneur de base (par exemple, Windows Server 2019) avec le système d'exploitation de l'hôte du conteneur (par exemple, Windows Server 2019). Les tests et le passage à des conteneurs peuvent faciliter les mises à niveau du système d'exploitation à l'avenir en modifiant l'image de base dans votre Dockerfile et en la déployant sur un nouvel ensemble d'hôtes exécutant la dernière version de Windows.

Supprimer les outils de gestion et les licences tiers

La gestion de votre parc de serveurs nécessite l'utilisation de plusieurs outils d'exploitation système tiers pour la gestion des correctifs et de la configuration. Cela peut rendre la gestion de l'infrastructure complexe et vous devez souvent faire face à des coûts de licence tiers. Si vous utilisez des conteneurs AWS, vous n'avez pas besoin de gérer quoi que ce soit du côté du système d'exploitation. Le moteur d'exécution du conteneur gère les conteneurs. Cela signifie que l'hôte sous-jacent est éphémère et peut facilement être remplacé. Vous pouvez gérer vos conteneurs sans avoir à gérer directement l'hôte du conteneur. De plus, vous pouvez utiliser des outils gratuits tels que AWS Systems Manager Session Manager l'accès facile à l'hôte et la résolution des problèmes.

Améliorez le contrôle et la portabilité

Les conteneurs vous permettent de contrôler plus précisément les ressources du serveur, telles que le processeur et la RAM, que vous n'en avez sur les instances EC2. Pour les instances EC2, vous pouvez contrôler le processeur et la RAM en sélectionnant une famille d'instances, un type d'instance et des [options de processeur](#). Cependant, avec les conteneurs, vous pouvez définir exactement la quantité de CPU ou de RAM que vous souhaitez allouer à un conteneur dans vos définitions de tâches ECS ou aux [pods dans Amazon EKS](#). En fait, nous recommandons de [spécifier le processeur et la mémoire au niveau du conteneur](#) pour les conteneurs Windows. Ce niveau de granularité présente des avantages en termes de coûts. Examinez l'exemple de code suivant :

```
json
{
  "taskDefinitionArn": "arn:aws:ecs:us-east-1:123456789012:task-definition/demo-
service:1",
  "containerDefinitions": [
    {
      "name": "demo-service",
      "image": "mcr.microsoft.com/dotnet/framework/samples:aspnetapp-
windowsservercore-ltsc2019",
      "cpu": 512,
      "memory": 512,
      "links": [],
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 0,
          "protocol": "tcp"
        }
      ],
    }
  ],
}
```

Accélérez l'innovation

Le passage aux conteneurs facilite l'automatisation des étapes du cycle de développement, notamment la création, le test et le déploiement d'applications. Si vous automatisez ces processus, vous donnez à vos équipes de développement et d'exploitation plus de temps pour se concentrer sur l'innovation.

Réduire le coût total de possession

Le passage aux conteneurs réduit souvent le recours aux outils de gestion des licences et de protection des terminaux. Les conteneurs étant des unités de calcul éphémères, vous pouvez automatiser et simplifier les tâches de gestion telles que l'application de correctifs, le dimensionnement, la sauvegarde et la restauration. Cela peut réduire le coût total de possession lié à l'administration et à l'exploitation des charges de travail basées sur des conteneurs. Les conteneurs sont plus efficaces que les machines virtuelles car ils vous permettent d'optimiser le placement de vos applications afin d'augmenter l'utilisation des ressources d'infrastructure de vos applications.

Comblent le déficit de compétences

AWS propose des programmes et des journées d'immersion pour améliorer les compétences des équipes de développement client en matière de conteneurs et de DevOps technologies. Cela inclut des services pratiques de conseil et d'habilitation.

Refactorisez vers .NET 5+ et utilisez des conteneurs Linux

Bien que vous puissiez réduire les coûts en déplaçant vos applications .NET Framework vers des conteneurs, vous pouvez réaliser des économies supplémentaires en refactorisant les anciennes applications .NET pour les remplacer par des alternatives natives au cloud sur AWS.

Supprimer les coûts de licence

La refactorisation de votre application de .NET Framework sous Windows à .NET Core sous Linux permet de réaliser des économies d'environ 45 %.

Accédez aux dernières améliorations

La refactorisation de vos applications de .NET Framework sous Windows à .NET Core sous Linux vous donne accès aux dernières améliorations telles que Graviton2. Graviton2 offre un meilleur rapport prix/performances de 40 % par rapport à des instances comparables.

Améliorez la sécurité et les performances

La refactorisation de votre application de .NET Framework sous Windows vers des conteneurs .NET Core sous Linux améliore la sécurité et les performances. Cela s'explique par le fait que vous bénéficiez des derniers correctifs de sécurité, que vous bénéficiez de l'isolation des conteneurs et que vous avez accès à de nouvelles fonctionnalités.

Utilisez des conteneurs Windows au lieu d'exécuter de nombreuses applications sur une seule instance d'IIS

Tenez compte des avantages suivants liés à l'utilisation de conteneurs Windows au lieu d'exécuter plusieurs applications sur une instance Windows EC2 avec Internet Information Services (IIS) :

- **Sécurité** — Les conteneurs fournissent un niveau de sécurité prêt à l'emploi qui n'est pas atteint par l'isolation au niveau d'IIS. Si un site Web ou une application IIS est compromis, tous les autres sites hébergés sont exposés et vulnérables. L'évasion de conteneurs est rare et constitue une

vulnérabilité plus difficile à exploiter que de prendre le contrôle d'un serveur par le biais d'une vulnérabilité Web.

- **Flexibilité** — La possibilité d'exécuter des conteneurs de manière isolée par processus et de disposer de leur propre instance permet des options de mise en réseau plus granulaires. Les conteneurs proposent également des méthodes de distribution complexes sur de nombreuses instances EC2. Vous ne bénéficiez pas de ces avantages lorsque vous consolidez des applications sur une seule instance IIS.
- **Frais de gestion** : l'indication du nom du serveur (SNI) crée une surcharge qui nécessite une gestion et une automatisation. Vous devez également vous concentrer sur les opérations classiques de gestion du système d'exploitation, telles que l'application de correctifs, le dépannage des BSOD (si le dimensionnement automatique n'est pas en place), la protection des terminaux, etc. La configuration des sites IIS conformément aux [meilleures pratiques de sécurité](#) est une activité longue et continue. Vous devrez peut-être même définir des [niveaux de confiance](#), ce qui alourdit également les frais de gestion. Les conteneurs sont conçus pour être apatrides et immuables. En fin de compte, vos déploiements sont plus rapides, plus sécurisés et reproductibles si vous utilisez plutôt des conteneurs Windows.

Étapes suivantes

Investir dans une infrastructure moderne pour exécuter vos anciennes charges de travail apporte d'immenses avantages à votre entreprise. AWS les services de conteneurs facilitent la gestion de votre infrastructure sous-jacente, que ce soit sur site ou dans le cloud, afin que vous puissiez vous concentrer sur l'innovation et les besoins de votre entreprise. Près de 80 % de tous les conteneurs du cloud fonctionnent AWS aujourd'hui. AWS fournit un ensemble complet de services de conteneurs pour à peu près tous les cas d'utilisation. Pour commencer, consultez la section [Conteneurs à l'adresse AWS](#).

Ressources supplémentaires

- [Optimisez le coût des charges de travail des conteneurs avec les fournisseurs de capacité ECS et les instances ponctuelles EC2 \(blog\)AWS](#)
- [Liste de contrôle pour l'optimisation des coûts pour Amazon ECS et AWS Fargate \(AWS blog\)](#)
- [Amazon EKS sur AWS Graviton2 est généralement disponible : considérations relatives aux applications multi-architectures \(Blog\)AWS](#)
- [Optimisation des coûts pour Kubernetes sur AWS \(Blog\)AWS](#)
- [Optimisation de vos coûts de calcul Kubernetes grâce à la consolidation Karpenter \(Blog\)AWS](#)

Optimisez les coûts AWS Fargate des tâches sur Amazon ECS

Présentation de

Le dimensionnement correct des AWS Fargate tâches est une étape importante pour l'optimisation des coûts. Trop souvent, les applications sont conçues avec un dimensionnement arbitraire pour les tâches Fargate et ne sont jamais revisitées. Cela peut entraîner un surprovisionnement des tâches Fargate et des dépenses inutiles. Cette section explique comment fournir des recommandations exploitables afin d'optimiser le processeur et la mémoire des tâches pour vos services Amazon Elastic Container Service (Amazon ECS) exécutés sur Fargate. [Optimiseur de calcul AWS](#) Compute Optimizer quantifie également l'impact financier de l'adoption de ces recommandations. Cela vous permet de hiérarchiser vos efforts d'optimisation en fonction de l'ampleur de l'opportunité d'économies. Les recommandations de Compute Optimizer fournissent des configurations de processeur et de mémoire au niveau du conteneur pour les tâches de réduction de la taille.

Avantages en termes de coûts

Le dimensionnement correct des tâches Amazon ECS sur Fargate peut réduire les coûts de 30 à 70 % pour les tâches de longue durée. Sans passer en revue les indicateurs de performance des applications pour ajuster la taille de votre tâche, vous pouvez appliquer au dimensionnement des conteneurs le même état d'esprit que celui utilisé sur les instances de calcul EC2. Cela entraîne des tâches Fargate surdimensionnées qui augmentent les coûts liés aux ressources inutilisées. Vous pouvez utiliser Compute Optimizer pour identifier les bonnes opportunités de dimensionnement de manière réactive. Idéalement, le propriétaire de l'application passe en revue les indicateurs de performance spécifiques à l'application et supprime la surcharge du système d'exploitation pour s'assurer que la taille de tâche appropriée est spécifiée. Pour plus d'informations, consultez la section [Déplacer des applications Windows vers des conteneurs](#) de ce guide.

Recommandations en matière d'optimisation des coûts

Cette section propose des recommandations pour utiliser Compute Optimizer afin de dimensionner correctement vos tâches Amazon ECS on Fargate.

Dans le cadre du processus d'optimisation des coûts, nous vous recommandons de procéder comme suit :

- Activer Compute Optimizer
- Consomme les résultats de Compute Optimizer

- Étiquetez les tâches pour qu'elles soient correctement dimensionnées
- Activez l'étiquette de répartition des coûts pour qu'elle fonctionne avec les outils AWS de facturation
- Mettre en œuvre les bonnes recommandations de dimensionnement
- Révision des coûts avant et après dans Cost Explorer

Activer Compute Optimizer

Vous pouvez l'activer [Optimiseur de calcul AWS](#) au niveau de l'organisation ou du compte unique dans AWS Organizations. La configuration à l'échelle de l'organisation fournit des rapports permanents pour les instances nouvelles et existantes sur l'ensemble de votre flotte pour tous les comptes membres. Cela permet de faire de la bonne taille une activité récurrente plutôt qu'une point-in-time activité.

Niveau de l'organisation

Pour la plupart des entreprises, le moyen le plus efficace d'utiliser Compute Optimizer est au niveau de l'organisation. Cela fournit une visibilité multicompte et multirégionale de votre organisation et centralise les données dans une seule source à des fins de révision. Pour activer cette fonctionnalité au niveau de l'organisation, procédez comme suit :

1. Connectez-vous à votre [compte AWS Organizations de gestion](#) avec un rôle disposant des [autorisations requises](#) et choisissez de vous connecter à tous les comptes de cette organisation. [Toutes les fonctions doivent être activées](#) pour votre organisation.
2. Après avoir activé le compte de gestion, vous pouvez vous connecter au compte, voir tous les autres comptes membres et consulter leurs recommandations.

Note

Il est recommandé de configurer un [compte d'administrateur délégué](#) pour Compute Optimizer. Cela vous permet d'appliquer le principe du moindre privilège, en minimisant l'accès au compte de AWS Organizations gestion tout en fournissant un accès au service à l'échelle de l'organisation.

Niveau de compte unique

Si vous ciblez un compte dont les coûts sont élevés mais que vous n'y avez pas accès AWS Organizations, vous pouvez toujours activer Compute Optimizer pour ce compte et cette région. Pour en savoir plus sur le processus d'inscription, consultez [Getting started with Optimiseur de calcul AWS](#).

Note

Les recommandations sont actualisées tous les jours et leur génération peut prendre jusqu'à 12 heures. N'oubliez pas que Compute Optimizer a besoin de 24 heures de statistiques au cours des 14 derniers jours pour générer des recommandations pour Amazon ECS sur Fargate. Pour plus d'informations, consultez la section [Exigences relatives aux services Amazon ECS sur Fargate](#) dans la documentation de Compute Optimizer.

Compute Optimizer analyse automatiquement les indicateurs d'utilisation CloudWatch Amazon et Amazon ECS suivants pour vos services Amazon ECS sur Fargate :

- CPUUtilization— Le pourcentage de capacité du processeur utilisé dans le service.
- MemoryUtilization— Le pourcentage de mémoire utilisé dans le service.


Consume les résultats de Compute Optimizer

Prenons un exemple qui met l'accent sur les modifications de taille appropriées au sein d'un seul compte et d'une seule région. Dans cet exemple, Compute Optimizer est activé au niveau de l'organisation sur tous les comptes. N'oubliez pas que le bon dimensionnement est un processus perturbateur qui, dans la plupart des cas, est effectué avec précision par les propriétaires de l'application au cours d'une période de maintenance planifiée de plusieurs semaines.

Si vous accédez à Compute Optimizer depuis le compte de gestion d'une organisation (comme indiqué dans les étapes suivantes), vous pouvez choisir le compte que vous souhaitez examiner. Dans cet exemple, une tâche est exécutée sur un seul compte surapprovisionné en us-east-1. L'accent est mis sur le redimensionnement à la taille recommandée pour le service Amazon ECS.

1. Ouvrez la console [Compute Optimizer](#).
2. Sur la page Tableau de bord, filtrez par Findings=OverProvisioned pour voir tous les services Amazon ECS sur Fargate.

3. Pour consulter les recommandations détaillées concernant les services ECS surprovisionnés sur Fargate, faites défiler la page vers le bas, puis sélectionnez Afficher les recommandations.
4. Choisissez Exporter et enregistrez le fichier pour une utilisation future.

 Note

Pour enregistrer des recommandations en vue d'une future révision, vous devez disposer d'un compartiment S3 dans lequel Compute Optimizer pourra écrire dans chaque région. Pour plus d'informations, consultez la [politique relative aux compartiments Amazon S3 Optimiseur de calcul AWS dans la](#) documentation de Compute Optimizer.

Pour consulter les recommandations de Compute Optimizer, procédez comme suit :

1. Dans la [console Compute Optimizer](#), accédez à la page des recommandations d'exportation.
2. Pour la destination du compartiment S3, choisissez votre compartiment S3.
3. Dans la section Filtres d'exportation, pour Type de ressource, choisissez les services ECS sur Fargate.
4. Sur la page Recommandations pour les services ECS sur Fargate, explorez l'un des services ECS sur Fargate et consultez les recommandations de Compute Optimizer relatives au processeur et à la mémoire. Par exemple, consultez les recommandations des sections Comparer les paramètres actuels avec la taille de tâche recommandée et Comparer les paramètres actuels avec la taille de conteneur recommandée.

Pour obtenir la liste des services ECS pour Fargate dont vous avez besoin à la bonne taille, procédez comme suit :

1. Ouvrez la [console Amazon S3](#).
2. Dans le volet de navigation, choisissez Buckets, puis choisissez le bucket dans lequel vous avez exporté vos résultats.
3. Dans l'onglet Objets, sélectionnez votre objet et choisissez Télécharger.
4. Dans les résultats que vous avez téléchargés, filtrez la colonne de recherche pour afficher uniquement les services Amazon ECS OVER_PROVISIONED sur Fargate. Cela indique les services Amazon ECS que vous prévoyez de cibler pour obtenir le bon dimensionnement.
5. Stockez les définitions de tâches dans un éditeur de texte pour une utilisation ultérieure.

Tâches de balisage à la bonne taille

Le balisage de vos charges de travail est un outil puissant pour organiser vos ressources. AWS Vous pouvez utiliser des tags pour obtenir une visibilité précise des coûts et permettre la rétrofacturation. Il existe de nombreuses méthodes et stratégies pour ajouter des balises aux AWS ressources afin de gérer la rétrofacturation et l'automatisation. Pour plus d'informations, consultez le AWS livre blanc sur les [meilleures pratiques en matière de balisage AWS des ressources](#). L'exemple suivant permet [AWS CloudShell](#) de baliser toutes les tâches faisant partie d'un service Amazon ECS au sein du compte cible et Région AWS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$( aws ecs list-clusters -query 'clusterArns' -output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$( aws ecs list-services -cluster $ClustersArn -query 'serviceArns' -
output text)
  for ServiceArn in $ServiceArns; do
    TasksArns=$( aws ecs list-tasks -cluster $ClustersArn -service-name $ServiceArn -
query 'taskArns' -output text)
    for TasksArn in $TasksArns; do
      aws ecs tag-resource -resource-arn $TasksArn -tags key=$TAG_KEY,value=$TAG_VALUE
    done
  done
done
```

L'exemple de code suivant montre comment activer la [propagation des balises](#) vers tous les services Amazon ECS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$(aws ecs list-clusters --query 'clusterArns' --output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$(aws ecs list-services --cluster $ClustersArn --query 'serviceArns' --
output text)
  for ServiceArn in $ServiceArns; do
```

```
aws ecs update-service --cluster $ClustersArn --service $ServiceArn --propagate-tags
SERVICE &>/dev/null
aws ecs tag-resource --resource-arn $ServiceArn --tags key=$TAG_KEY,value=$TAG_VALUE
done
done
```

Activez l'étiquette de répartition des coûts pour qu'elle fonctionne avec les outils AWS de facturation

Nous vous recommandons d'activer la balise de répartition des coûts définie par l'utilisateur. Cela permet à la balise Rightsizing d'être reconnue et filtrable dans les outils AWS de facturation (par exemple, AWS Cost Explorer et) AWS Cost and Usage Report. Si vous ne l'activez pas, l'option de filtrage des balises et les données ne seront pas disponibles. Pour plus d'informations sur l'utilisation des balises de répartition des coûts, consultez la section [Activation des balises de répartition des coûts définies par l'utilisateur](#) dans la AWS Billing and Cost Management documentation.

Après 24 heures d'attente, vous pouvez voir le tag dans Cost Explorer avant de mettre en œuvre les recommandations de dimensionnement appropriées dans la section suivante. Pour ce faire, recherchez la balise Rightsizing dans Cost Explorer.

Mettre en œuvre les bonnes recommandations de dimensionnement

Compute Optimizer fournira des recommandations concernant la taille des tâches ou des conteneurs. Pour mettre en œuvre les recommandations de dimensionnement appropriées, procédez comme suit.

1. Ouvrez la [console Amazon ECS](#).
2. Dans la barre de navigation, choisissez la région qui contient votre définition de tâche.
3. Dans le panneau de navigation, choisissez Task definitions (Définition des tâches).
4. Sur la page Task definitions (Définition des tâches), choisissez la tâche, puis choisissez Create new revision (Créer une nouvelle révision).
5. Sur la page Create new task definition revision (Créer une nouvelle révision de définition de tâche), effectuez les modifications souhaitées. Pour mettre à jour la recommandation de taille de conteneur, mettez-la à jour `cpu` et `memory` sous le bloc ContainerDefinitions dans [votre](#) définition de tâche ECS. Par exemple :

```
"containerDefinitions": [  
  {  
    "name": "your-container-name",
```

```
"image": "your-image",
"cpu": 1024,
"memory": 2048,
}
],
```

6. Vérifiez les informations, puis choisissez Create (Créer).

Pour mettre à jour le service Amazon ECS, procédez comme suit :

1. Ouvrez la [console Amazon ECS](#).
2. Sur la page Clusters, sélectionnez le cluster.
3. Sur la page Cluster overview (Aperçu du cluster), sélectionnez le service, puis choisissez Update (Mettre à jour).
4. Pour Task definition (Définition de tâche), choisissez la famille et la révision de définition de tâche à utiliser.

Pour les opérateurs expérimentés, vous pouvez utiliser CloudShell pour mettre à jour le service Amazon ECS. Par exemple :

```
bash
#!/bin/bash
# Set variables
ClustersName="workshop-cluster"
ServiceName="lab7-fargate-service"
TaskDefinition="lab7-fargate-demo:3"
# update the service
aws ecs update-service --cluster $ClustersName --service $ServiceName --task-definition
$TaskDefinition
```

Révision avant et après les coûts

Une fois que vous avez correctement dimensionné vos ressources, vous pouvez utiliser Cost Explorer pour afficher les coûts avant et après en utilisant la balise Rightsizing. N'oubliez pas que vous pouvez utiliser [des balises de ressources](#) pour suivre les coûts. En utilisant plusieurs couches de balises, vous pouvez obtenir une visibilité précise de vos coûts. Dans l'exemple présenté dans ce guide, la balise Rightsizing est utilisée pour appliquer une balise générique à toutes les instances ciblées. Ensuite, un tag d'équipe est utilisé pour mieux organiser les ressources. L'étape suivante

consiste à introduire des balises d'application pour mieux montrer l'impact financier de l'exploitation d'une application spécifique.

Prenons un exemple de réduction des coûts qui peut être réalisée en utilisant la balise Rightsizing pour un seul niveau de compte. Dans cet exemple, les coûts d'exploitation passent de 30,26\$ par jour à 7,56\$ par jour. En supposant 744 heures par mois, le coût annuel avant de choisir la bonne taille est de 11 044,9\$. Après le bon dimensionnement, le coût annuel tombe à 2 759,4\$. Cela se traduit par une diminution de 75 % des coûts de calcul pour ce compte. Imaginez l'impact de cette situation sur une grande organisation.

Avant de vous lancer dans l'aventure de la bonne taille, tenez compte des points suivants :

- AWS offre de nombreuses options de réduction des coûts. Cela inclut [AWS OLA](#), où AWS passe en revue vos instances sur site avant de passer à AWS. L' AWS OLA vous fournit également des recommandations sur la bonne taille et des conseils en matière de licence.
- Choisissez toutes les bonnes tailles avant de souscrire à [Savings Plans](#). Cela peut vous aider à éviter les surachats dans le cadre de votre engagement Savings Plans.

Étapes suivantes

Nous vous recommandons de suivre les étapes suivantes :

1. Passez en revue votre environnement existant et envisagez de convertir les volumes Amazon EBS gp2 en volumes gp3.
2. Passez en revue [Savings Plans](#).

Ressources supplémentaires

- [Commencer à utiliser Compute Optimizer \(documentation\)](#) AWS
- [Bonnes pratiques pour le balisage des AWS ressources](#) (livres AWS blancs)
- [Conteneurs Windows activés AWS](#) (AWS Workshop Studio)

Gagnez en visibilité sur vos coûts Amazon EKS

Présentation de

Une vision globale est nécessaire pour surveiller efficacement le coût d'un déploiement de Kubernetes. Le seul coût fixe et connu concerne le plan de contrôle Amazon Elastic Kubernetes Service (Amazon EKS). Cela inclut tous les autres composants composant le déploiement, du calcul au stockage en passant par la mise en réseau, dont le montant varie en fonction des besoins de votre application.

Vous pouvez utiliser [Kubecost](#) pour analyser le coût de votre infrastructure Kubernetes, depuis les [espaces de noms](#) et les [services jusqu'aux pods](#) individuels, puis afficher les données dans un tableau de bord. Kubecost fait apparaître les coûts liés au cluster, tels que le calcul et le stockage, et les coûts out-of-cluster tels que les [buckets Amazon Simple Storage Service \(Amazon S3\)](#) et les instances Amazon Relational Database Service ([Amazon RDS](#)). Kubecost fera des recommandations de dimensionnement sur la base de ces données et affichera les alertes critiques susceptibles d'avoir un impact sur le système. Kubecost peut [s'intégrer AWS Cost and Usage Report](#) pour montrer les économies réalisées grâce aux [Compute Savings Plans](#), [aux instances réservées](#) et à d'autres programmes de réduction.

Avantages en termes de coûts

Kubecost fournit des rapports et des tableaux de bord qui permettent de visualiser le coût de vos déploiements Amazon EKS. Il vous permet d'accéder à chacun des différents composants tels que les contrôleurs, les services, les nœuds, les pods et les volumes à partir du cluster. Cela vous donne une vue globale de vos applications exécutées dans un environnement Amazon EKS. En activant cette visibilité, vous pouvez suivre les recommandations de Kubecost ou visualiser les coûts de chaque application à un niveau granulaire. La bonne taille d'un groupe de nœuds Amazon EKS permet de réaliser les mêmes économies potentielles que les instances EC2 standard. Si vous pouvez dimensionner correctement vos conteneurs et vos nœuds, vous pouvez supprimer la charge de calcul liée à la taille de l'instance nécessaire pour exécuter le conteneur et au nombre d'instances EC2 requises dans le groupe de dimensionnement automatique.

Recommandations en matière d'optimisation des coûts

Pour profiter de Kubecost, nous vous recommandons de procéder comme suit :

1. Déployez Kubecost dans votre environnement

2. Obtenez une ventilation détaillée des coûts des applications Windows
3. Nœuds de cluster de la bonne taille
4. Demandes de conteneurs de la bonne taille
5. Gérez les nœuds sous-utilisés
6. Remédier aux charges de travail abandonnées
7. Agir sur les recommandations
8. Mettre à jour les nœuds autogérés

Déployez Kubecost dans votre environnement

L'[atelier Amazon EKS Finhack](#) vous apprend à déployer un environnement Amazon EKS configuré pour utiliser Kubecost dans un AWS compte détenu. Cela vous permet d'acquérir une expérience pratique de la technologie. Si vous souhaitez organiser cet atelier au sein de votre organisation, contactez l'équipe chargée de votre compte.

Pour déployer Kubecost sur votre cluster Amazon EKS à l'aide de [Helm](#), consultez l'article sur le blog consacré à la [collaboration entre Kubecost AWS et Kubecost pour fournir un suivi des coûts aux clients d'EKS](#). AWS Vous pouvez également consulter la [documentation officielle de Kubecost](#) pour obtenir des instructions sur l'installation et la configuration de Kubecost. Pour plus d'informations sur le support de Kubecost pour les nœuds Windows, consultez la section [Support des nœuds Windows](#) dans la documentation de Kubecost.

Obtenez une ventilation détaillée des coûts des applications Windows

Bien que vous puissiez réaliser d'importantes économies en utilisant les [instances Spot Amazon EC2](#), vous pouvez également tirer parti du fait que les charges de travail Windows ont tendance à être dynamiques. L'utilisation des instances Spot dépend de l'application, et nous vous encourageons à vérifier si elles s'appliquent à votre cas d'utilisation.

Pour obtenir une ventilation détaillée des coûts de vos applications Windows, [connectez-vous à Kubecost](#). Sur la page de navigation, sélectionnez Savings.

Nœuds de cluster de la bonne taille

Dans [Kubecost](#), choisissez Savings dans la barre de navigation, puis choisissez Rightsize your cluster node.

Prenons un exemple où Kubecost indique que le cluster est surprovisionné en termes de vCPU et de RAM. Le tableau suivant présente les détails et les recommandations de Kubecost.

	Current	Recommandation : Simple	Recommandation : Complexe
Nombre total	3462,57\$ US par mois	137,24\$ US par mois	303,68\$ US par mois
Nombre de nœuds	4	5	4
CPU	74 VCPUs	10 VCPUs	8 VCPUs
RAM	152 GO	20 Go	18 GO
Répartition des instances	2 c5.xlarge + 2 de plus	5 t3a.medium	2 c5n.large + 1 de plus

Comme décrit dans le billet de blog de Kubecost intitulé [Trouvez un ensemble optimal de nœuds pour un cluster Kubernetes](#), l'option simple utilise un seul groupe de nœuds, tandis que l'option complexe utilise une approche de groupe à nœuds multiples. Le bouton Apprenez à adopter permet de redimensionner le cluster en un clic. Cela nécessite l'installation du [Kubecost Cluster](#) Controller.

Si vous utilisez des [nœuds Windows autogérés](#) qui ne sont pas créés par [eksctl](#), consultez [Mettre à jour un groupe de nœuds autogérés existant](#). Ces instructions vous montrent comment modifier le type d'instance dans le modèle de lancement Amazon EC2 utilisé par le groupe [Auto Scaling](#).

Demandes de conteneurs de la bonne taille

Dans [Kubecost](#), choisissez Savings dans la barre de navigation, puis accédez à la page Demander des recommandations de dimensionnement correct. Cette page présente l'[efficacité](#) des capsules, les recommandations relatives à la bonne taille et les économies estimées. Vous pouvez utiliser le bouton Personnaliser pour filtrer par cluster, nœud, espace de noms \ contrôleur, etc.

Par exemple, considérez que Kubecost a calculé que certains de vos pods sont surapprovisionnés en termes de CPU et de RAM (mémoire). Kubecost vous recommande ensuite de vous adapter aux nouvelles valeurs du processeur et de la RAM pour réaliser ses économies mensuelles estimées. Pour modifier les valeurs du processeur et de la RAM, vous devez mettre à jour votre fichier [manifeste de déploiement](#).

Gérez les nœuds sous-utilisés

Dans [Kubecost](#), choisissez Savings dans la barre de navigation, puis sélectionnez Gérer les nœuds sous-utilisés.

Prenons un exemple où la page montre qu'un nœud du cluster est sous-utilisé en termes de CPU et de RAM (mémoire) et peut donc être vidé et arrêté ou redimensionné. Le choix des nœuds qui ne passent pas les vérifications des nœuds et des pods vous donnera plus d'informations sur les raisons pour lesquelles ils ne peuvent pas être vidangés.

Remédier aux charges de travail abandonnées

Dans [Kubecost](#), choisissez Savings dans la barre de navigation, puis choisissez la page Charges de travail abandonnées. Dans cet exemple, vous filtrez par espace de noms appelé fenêtrés. Cette page indique les pods qui n'ont pas atteint le seuil de trafic et qui sont considérés comme abandonnés. Les pods doivent envoyer ou recevoir une certaine quantité de trafic réseau au cours de la période définie.

Après avoir soigneusement étudié l'abandon d'un ou de plusieurs pods, vous pouvez réaliser des économies en réduisant le nombre de répliques, en supprimant le déploiement, en le redimensionnant pour consommer moins de ressources ou en informant le propriétaire de l'application que vous pensez que le déploiement est abandonné.

Agir sur les recommandations

Dans la section Dimensionnez correctement les nœuds de votre cluster, Kubecost analyse l'utilisation des nœuds de travail du cluster et formule des recommandations sur le dimensionnement correct des nœuds afin de réduire les coûts. Deux types de groupes de nœuds peuvent être utilisés avec Amazon EKS : [autogérés et gérés](#).

Mettre à jour les nœuds autogérés

Pour plus d'informations sur la mise à jour des nœuds autogérés, consultez la section [Mises à jour des nœuds autogérés](#) dans la documentation Amazon EKS. Il indique que les groupes de nœuds créés avec `ne eksctl` peuvent pas être mis à jour et doivent être migrés vers un nouveau groupe de nœuds avec la nouvelle configuration.

Par exemple, supposons que vous avez un groupe de nœuds Windows appelé `ng-windows-m5-2xlarge` (qui utilise une instance EC2 `m5.2xlarge`) et que vous souhaitez migrer les pods vers

un [nouveau groupe de nœuds](#) appelé `ng-windows-t3-large` (qui est soutenu par une instance EC2 `t3.large` pour réduire les coûts).

Pour migrer vers un nouveau groupe de nœuds lorsque vous utilisez des groupes de nœuds déployés par `kubectl`, procédez comme suit :

1. Pour trouver le nœud dans lequel se trouve actuellement le pod, exécutez la commande `kubectl describe pod <pod_name> -n <namespace>`.
2. Exécutez la commande `kubectl describe node <node_name>`. La sortie indique que le nœud s'exécute sur une instance `m5.2xlarge`. Il correspond également au nom du groupe de nœuds (`ng-windows-m5-2xlarge`).
3. Pour modifier le déploiement afin d'utiliser un groupe de nœuds `ng-windows-t3-large`, supprimez le groupe de nœuds `ng-windows-m5-2xlarge` et exécutez la commande `kubectl describe svc,deploy,pod -n windows`. Le déploiement commence immédiatement à se redéployer une fois que son groupe de nœuds a été supprimé.

Note

Le service sera interrompu lorsque vous supprimerez le groupe de nœuds.

4. Exécutez à nouveau la commande `kubectl describe svc,deploy,pod -n windows` après quelques minutes. La sortie indique que les pods sont tous à nouveau en cours d'exécution.
5. Pour montrer que les pods s'exécutent désormais sur le groupe de nœuds `ng-windows-t3-large`, réexécutez les commandes `kubectl describe node <node_name>` et `kubectl describe pod <pod_name> -n <namespace>`.

Autres méthodes de redimensionnement

Cette méthode s'applique à toute combinaison de groupes de nœuds autogérés ou gérés. Le billet de blog sur la [migration fluide des charges de travail d'un groupe de nœuds autogérés par EKS vers des groupes de nœuds gérés par EKS](#) fournit des conseils sur la façon de migrer vos charges de travail d'un groupe de nœuds avec le type d'instance surdimensionné vers le groupe de nœuds correctement dimensionné, sans interruption de service.

Étapes suivantes

Kubecost permet de visualiser facilement le coût de vos environnements Amazon EKS. L'intégration approfondie de Kubecost à Kubernetes AWS APIs peut vous aider à réaliser des économies

potentielles. Vous pouvez les consulter sous forme de recommandations dans le tableau de bord des économies de Kubecost. Kubecost peut également implémenter certaines de ces recommandations pour vous grâce à sa fonction de [contrôleur de cluster](#).

Nous vous recommandons de passer en revue le step-by-step déploiement dans le billet de blog publié sur le blog AWS Containers [AWS et Kubecost collabore pour fournir une surveillance des coûts aux clients d'EKS](#).

Ressources supplémentaires

- Atelier [Amazon EKS \(Atelier Amazon EKS\)](#)
- [AWS et Kubecost collaborent pour fournir un suivi des coûts aux clients d'EKS \(Blog\)](#) AWS
- Atelier [Amazon EKS Finhack \(AWS atelier Studio\)](#)
- [Conteneurs Windows activés AWS](#) (AWS Workshop Studio)

Replateforme des applications Windows avec App2Container

Présentation de

[AWS App2Container](#) est un outil de ligne de commande permettant de migrer et de moderniser des applications Web Java et .NET vers des conteneurs. App2Container analyse et dresse un inventaire de toutes les applications exécutées sur du matériel nu, des machines virtuelles, des instances Amazon Elastic Compute Cloud (Amazon EC2) ou d'autres fournisseurs de cloud. Vous sélectionnez l'application que vous souhaitez conteneuriser. App2Container regroupe les artefacts et les dépendances de l'application dans des images de conteneur, configure les ports réseau et génère les artefacts de déploiement Amazon Elastic Container Service (Amazon ECS) et Amazon Elastic Kubernetes Service (Amazon EKS) nécessaires au déploiement, qui sont des modèles d'infrastructure en tant que code (IaC). App2Container fournit l'infrastructure cloud et les pipelines CI \ CD nécessaires au déploiement de l'application conteneurisée dans un environnement de production. Pour plus d'informations, consultez la section [Fonctionnement d'App2Container](#) dans la documentation d'App2Container.

Avec App2Container, vous pouvez migrer AWS et moderniser vos applications sous forme de conteneurs, tout en normalisant le déploiement et les opérations de vos applications. Vous pouvez utiliser App2Container pour créer rapidement une preuve de concept (PoC) ou accélérer le déploiement de charges de travail de production dans des conteneurs.

Il y a plusieurs points à garder à l'esprit lorsque vous travaillez avec des applications Windows. App2Container prend en charge la conteneurisation des applications ASP.NET déployées sur Microsoft Internet Information Services (IIS), y compris les applications Windows Communication Foundation (WCF) hébergées par IIS qui s'exécutent sous Windows Server 2016, Windows Server 2019 ou Windows Server Core 2004. Pour plus d'informations, consultez la section [Applications prises en charge pour Windows](#) dans la documentation d'App2Container. App2Container utilise Windows Server Core comme image de base pour ses artefacts de conteneur, en faisant correspondre la version du conteneur Windows Server Core à la version du système d'exploitation (OS) du serveur sur lequel vous exécutez les commandes de conteneurisation. Cette approche dissocie l'application du système d'exploitation sous-jacent afin que vous puissiez mettre à niveau le système d'exploitation sans effectuer de migration traditionnelle.

Si vous utilisez un ordinateur subordonné pour conteneuriser votre application, l'image de base du conteneur, telle que le canal de maintenance à long terme (LTSC) de Windows Server 2019, correspond au système d'exploitation de votre ordinateur subordonné, tel que Windows Server 2019. Si vous exécutez la conteneurisation directement sur des serveurs d'applications, la version correspond au système d'exploitation de votre serveur d'applications. Si vos applications s'exécutent sous Windows Server 2008 ou 2012 R2, vous pouvez toujours utiliser App2Container en configurant un ordinateur subordonné pour les étapes de conteneurisation et de déploiement. App2Container ne prend pas en charge les applications exécutées sur les systèmes d'exploitation clients Windows, tels que Windows 7 ou Windows 10. App2Container prend en charge Tomcat, Tomee et les frameworks JBoss (en mode autonome) pour les processus Java. Pour plus d'informations, consultez la section [Compatibilité avec App2Container](#).

Avantages en termes de coûts

La conteneurisation et la consolidation de vos applications peuvent générer jusqu'à [60 % d'économies de calcul](#) par rapport à un modèle de conception de déploiement de one-application-to-one serveurs. App2Container permet d'accélérer le processus de conteneurisation des applications. Voici quelques-uns des avantages de l'utilisation d'App2Container pour vos besoins de modernisation :

- App2Container est proposé sans frais supplémentaires.
- App2Container prend en charge plusieurs applications dans une image de conteneur.
- Résolvez les problèmes des systèmes d'exploitation qui approchent de la fin du support en utilisant App2Container pour déplacer vos anciennes applications .NET vers des conteneurs. Vous pouvez

passer à un système d'exploitation plus récent, éviter de payer pour un support étendu et réduire les risques de sécurité.

- Les conteneurs constituent une méthode efficace et rentable pour emballer vos applications .NET. Passez en revue les avantages des conteneurs dans la [recommandation MACO sur le passage aux conteneurs](#).
- La consolidation et la conteneurisation des applications contribuent à réduire votre encombrement en termes de calcul, de stockage et de licences en utilisant vos ressources informatiques de manière plus efficace.
- Le passage aux conteneurs permet de réduire les frais d'exploitation et les coûts d'infrastructure et d'accroître la portabilité du développement et l'agilité du déploiement.

Recommandations en matière d'optimisation des coûts

Pour obtenir des instructions sur l'utilisation d'App2Container, consultez [Getting started with](#). AWS App2Container Pour plus d'informations sur les commandes App2Container, consultez la référence des commandes [App2Container](#).

Étapes suivantes

App2Container peut accélérer le processus de conteneurisation des applications et de déploiement sur Amazon EKS ou Amazon ECS. Le déploiement d'applications dans des conteneurs réduit les coûts de calcul, de mise en réseau et de stockage ainsi que les frais opérationnels pour les opérateurs d'applications.

Pour une expérience pratique avec App2Container, consultez l'atelier [Modernisation avec AWS App2Container](#) App2Container. Si vous souhaitez vivre une expérience d'apprentissage approfondie, demandez à l'équipe chargée de votre AWS compte d'organiser une journée d'immersion dans App2Container.

Ressources supplémentaires

- [Conteneurisation d'applications Windows multiniveaux complexes à l'aide de AWS App2Container](#)(AWS article de blog)
- [Conteneurisation d'applications ASP.NET héritées à l'aide de AWS App2Container](#)(billet de blog)AWS
- [Applications prises en charge par App2Container](#) (documentation)AWS

- [Modernisez avec AWS App2Container Workshop](#) (AWS Workshop Studio)
- [AWS App2Container FAQs](#)(AWS site Web)

Stockage

Le choix du stockage adapté à vos charges de travail Microsoft est une décision architecturale essentielle. Dans le cadre du processus décisionnel, nous vous recommandons de développer un plan de stockage et de déterminer les exigences fonctionnelles pour vos applications et services. Ce chapitre fournit une vue d'ensemble des options de stockage suivantes qui pourraient être prises en compte dans votre planification.

Rubriques :

- [Amazon EBS](#)
- [Amazon FSx](#)
- [AWS Storage Gateway](#)

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) est un service de stockage par blocs entièrement géré qui vous permet de stocker des volumes de stockage persistants au niveau des blocs que vous pouvez utiliser avec des instances Amazon Elastic Compute Cloud (Amazon EC2). Vous pouvez tirer parti de plusieurs fonctionnalités d'Amazon EBS pour gérer et optimiser efficacement vos ressources de stockage pour les charges de travail Windows dans le cloud. Par exemple, vous pouvez utiliser Amazon EBS pour fournir la quantité exacte d'IOPS et de débit dont vous avez besoin pour votre charge de travail, sélectionner parmi une gamme de types de volumes correspondant à vos exigences de charge de travail et utiliser des outils pour identifier et éliminer les ressources de stockage gaspillées. Ce contrôle granulaire des performances et de l'utilisation du stockage vous permet d'optimiser vos ressources de stockage tout en évitant des coûts inutiles.

Cette section couvre les rubriques suivantes :

- [Migrer les volumes Amazon EBS de gp2 à gp3](#)
- [Modifier les instantanés Amazon EBS](#)
- [Supprimer des volumes Amazon EBS non attachés](#)

Migrer les volumes Amazon EBS de gp2 à gp3

Présentation de

Un disque SSD est l'option de stockage standard pour la production et les charges de travail à hautes performances. Amazon EBS propose un [volume SSD à usage général](#) pour les charges de travail moyennes à hautes performances. La norme utilisée par de nombreuses Services AWS entreprises (y compris Amazon EC2) est le [gp2](#), la deuxième génération de ces volumes SSD à usage général. La troisième génération d'appareils à usage général SSDs, appelée [gp3](#), est sortie en décembre 2020.

L'offre gp3 a apporté des améliorations significatives aux aspects de personnalisation des performances par rapport à la génération précédente. Pour les volumes Amazon EBS gp2, les performances sont étroitement liées à la taille du volume. Pour chaque Go de capacité, les volumes gp2 obtiennent 3 IOPS de performance. C'est-à-dire qu'un volume gp2 de 2 000 Go est capable de 6 000 IOPS. Pour les volumes gp3, les performances peuvent être personnalisées indépendamment de la capacité de stockage. Cela permet aux volumes, même de petite capacité, d'atteindre des performances allant jusqu'à 80 000 IOPS et un débit de 2 000 Mo/s.

Un autre changement majeur apporté aux volumes gp3 concerne les performances IOPS de base. Les volumes gp3 commencent à 3 000 IOPS. À titre de comparaison, les volumes gp2 doivent atteindre 1 TiB avant d'atteindre les mêmes capacités de performance. Pour Windows Server, qui possède généralement un lecteur C : bien inférieur à 1 TiB, la mise à niveau de gp2 vers gp3 constitue une amélioration significative des performances.

Enfin, le prix des volumes gp3 est l'une des plus grandes améliorations par rapport aux volumes gp2. Les volumes gp3 offrent des capacités de performance améliorées à un coût inférieur de 20 % à celui des volumes gp2.

Impact sur les coûts

Étant donné la possibilité d'adapter les performances indépendamment de la capacité, il est important de comprendre les aspects tarifaires liés à l'ajout d'IOPS et de débit supplémentaires. Pour les volumes gp2, la tarification est basée sur une capacité provisionnée à 0,10\$ par Go par mois. Pour les volumes gp3, le prix est similaire à celui des [volumes SSD IOPS provisionnés](#) à hautes performances, qui ont un coût unique pour la capacité et un coût distinct pour les IOPS et le débit supplémentaires.

Comme indiqué dans le tableau suivant, les volumes gp3 ont un prix de capacité de 0,08 dollar par Go par mois (20 % moins cher que le volume gp2) et des coûts distincts pour les IOPS à 0,005 dollar

par mois d'IOPS provisionné sur 3 000 et à 0,04 dollar par mois provisionné sur 125 pour le débit.
MiBs MiBs

	gp3	gp2
Taille du volume	1 GiB — 64 TiB	1 Gio – 16 Tio
IOPS de référence	3 000	3 IOPS/GiB (minimum 100 IOPS) à 16 000 IOPS maximum Les volumes inférieurs à 1 TiB peuvent atteindre 3 000 IOPS
IOPS maximum/volume	80 000	16,000
Débit de référence	125 MiBs	La limite de débit est comprise entre 128 MiBs et 250 MiBs, selon la taille du volume
Débit maximal/volume	2 000 MiBs	250 MiBs
Prix	0,08 \$/GiB-mois 3 000 IOPS gratuites et 0,005 \$/IOPS provisionnées par mois pour plus de 3 000 125\$ MiBs gratuits et 0,04\$ provisionnés par mois pour plus de MiBs 125\$ MiBs	0,10 USD/Go par mois

Important

Même si les volumes gp3 ont des coûts distincts en termes de capacité et de performance, les volumes gp3 sont toujours moins chers que les volumes gp2 s'ils sont configurés aux mêmes niveaux de performance.

Les tableaux suivants présentent des exemples d'économies de coûts qui peuvent être réalisées en convertissant des volumes gp2 en volumes gp3 avec différentes configurations de capacité et de performance.

Exemple de configuration GP2

Taille du volume (Gio)	Nombre maximal d'IOPS	Débit () MiBs	Coût (USD/mois)
30	3000	128	3,00\$
100	3000	128	10,00\$
500	3000	250	50,00\$
1 000	3000	250	100,00\$
2000	6 000	250	200,00\$
6 000	16000	250	600,00\$

Exemple de configuration gp3 (ligne de base)

Nombre maximal d'IOPS	Débit () MiBs	Coût (USD/mois)	Réduction des coûts (par rapport au GP2)
3000	125	2,40\$	20 %
3000	125	8,00\$	20 %
3000	125	40,00\$	20 %
3000	125	80,00\$	20 %
3000	125	160,00\$	20 %
3000	125	480,00\$	20 %

Exemple de configuration gp3 (gp2 matching)

Nombre maximal d'IOPS	Débit () MiBs	Coût (USD/mois)	Réduction des coûts (par rapport au GP2)
3000	128	2,52\$	16 %
3000	128	8,12\$	19 %
3000	250	45,00\$	10 %
3000	250	85,00\$	15 %
6 000	250	180,00\$	10 %
16000	250	550,00\$	8 %

Pour une analyse des coûts, consultez la section relative au calculateur de réduction des coûts de migration d'EBS gp2 vers gp3 dans la ressource [Amazon EBS](#). Vous pouvez télécharger le calculateur et l'utiliser pour savoir combien vous pouvez économiser en migrant vos volumes gp2 vers gp3.

Recommandations en matière d'optimisation des coûts

Pour savoir comment effectuer le processus de migration, consultez l'article [Migrer vos volumes Amazon EBS de la version gp2 à la version gp3 et économiser jusqu'à 20 % sur les coûts sur le blog sur le AWS stockage](#).

Ressources supplémentaires

- [Migrez vos volumes Amazon EBS de gp2 à gp3 et économisez jusqu'à 20 % sur les coûts](#) (AWS Storage Blog)
- [Créez une règle AWS Config personnalisée pour optimiser les types de volumes Amazon EBS](#) (blog sur les opérations et migrations dans le AWS cloud)
- [Contrôlez vos AWS coûts en supprimant les volumes Amazon EBS inutilisés](#) (blog sur les opérations et migrations dans le AWS cloud)
- [Utilitaire de migration Amazon EBS](#) () GitHub
- [Réaliser des économies grâce aux annonces de 2020 relatives à re:Invent](#) (AWS Cloud Financial Management)

- [Atelier sur l'optimisation des coûts](#) (AWS Well-Architected Labs)
- [Calculateur de réduction des coûts de migration de GP2 vers GP3 \(téléchargement\)](#)

Modifier les instantanés Amazon EBS

Présentation de

La suppression de volumes EBS et la gestion de la conservation et de l'archivage des instantanés sont des aspects importants pour contrôler les coûts dès le départ. Vous pouvez sauvegarder les données de vos volumes EBS sur Amazon Simple Storage Service (Amazon S3) en prenant des instantanés. point-in-time Les instantanés sont des sauvegardes incrémentielles. Ils enregistrent donc uniquement les blocs sur les appareils qui ont été modifiés après votre dernier instantané. Cela réduit le temps nécessaire pour créer l'instantané, ainsi que les coûts de stockage en ne dupliquant pas les données. Chaque instantané contient toutes les informations nécessaires pour restaurer vos données (depuis sa création) sur un nouveau volume EBS.

Les frais pour les instantanés EBS sont calculés au gigaoctet par mois. Vous êtes facturé en fonction de la taille de l'instantané et de la durée pendant laquelle vous le conservez. Le prix varie en fonction du niveau de stockage. Pour le [niveau Standard](#), vous n'êtes facturé que pour les blocs modifiés qui sont stockés. Pour le niveau Archive, tous les blocs d'instantanés stockés vous sont facturés. [La récupération des instantanés depuis le niveau Archive vous est également facturée.](#) Voici des exemples de scénarios pour chaque niveau de stockage :

- Niveau standard : vous disposez d'un volume qui stocke 100 Go de données. La totalité des 100 Go de données vous est facturée pour le premier instantané (snap A). Au moment de la capture d'écran suivante (snap B), vous disposez de 105 Go de données. Vous n'êtes alors facturé que pour les 5 Go de stockage supplémentaires pour le snap B incrémentiel.
- Niveau d'archivage : vous archivez le cliché B. L'instantané est ensuite déplacé vers le niveau d'archivage, et la totalité du bloc d'instantanés de 105 Go vous est facturée.

Vous pouvez utiliser [Amazon Data Lifecycle Manager](#) pour vous aider à configurer un cycle de vie afin de conserver et de gérer vos instantanés dans les délais.

Impact sur les coûts

Les frais relatifs aux volumes EBS et aux instantanés sont gérés séparément. Les instantanés EBS sont facturés à un tarif inférieur à celui des volumes EBS actifs. Lorsqu'une instance se termine,

la valeur de l'[DeleteOnTermination attribut](#) pour chaque volume EBS attaché détermine s'il faut conserver ou supprimer le volume. Par défaut, l'`DeleteOnTermination` attribut est défini sur `True` le volume racine. Il est configuré `False` pour tous les autres types de volumes. Cela crée des situations dans lesquelles l'opérateur a l'intention de supprimer une instance EC2, mais laisse derrière lui des volumes qui ont été ajoutés à l'instance en plus du volume racine. Pour obtenir des instructions sur la vérification des volumes (et des instantanés associés) dont vous n'avez plus besoin, consultez [Afficher les informations relatives à un volume Amazon EBS](#) dans la documentation Amazon EBS.

Par défaut, lorsque vous créez un instantané, il est stocké dans le niveau Amazon EBS Snapshot Standard (niveau standard). Les instantanés stockés dans le niveau standard sont incrémentiels. Cela signifie que seuls les blocs du volume qui ont été modifiés après votre dernier instantané sont sauvegardés. L'[archive des instantanés Amazon EBS](#) est un nouveau niveau de stockage que vous pouvez utiliser pour le stockage à long terme et à faible coût de vos instantanés rarement consultés qui ne nécessitent pas de récupération fréquente ou rapide. La différence de prix entre le standard et l'archivage est significative et doit être un élément clé à prendre en compte lors de la mise en place de votre stratégie d'instantanés. Amazon EBS Snapshots Archive offre des coûts de stockage d'instantanés jusqu'à 75 % inférieurs pour les instantanés que vous prévoyez de stocker pendant 90 jours ou plus et auxquels vous n'avez que rarement besoin d'accéder.

Stockage de snapshots Amazon EBS	Cost
Standard	0,05 \$/Go par mois
Archivage	0,0125 \$/Go par mois

Dans les environnements plus restreints, les économies de coûts peuvent ne pas être significatives. Les économies sont plus importantes à grande échelle lorsque plusieurs comptes et des milliers d'instances EC2 contiennent des instantanés EBS même lorsque les volumes EBS ont été supprimés. TBs

Le tableau suivant compare les niveaux standard et d'archivage par mois avec seulement 50 To d'utilisation. Même à cette échelle inférieure, cela représente toujours des milliers de dollars d'économies par an.

Stockage de snapshots Amazon EBS	Coût par mois	Coût par an
50 To en standard	312,50\$	3 750\$
Archive 50 To	78,13\$	937,60\$
	Économies annuelles	2 812,40\$

Recommandations en matière d'optimisation des coûts

La suppression d'un instantané ne réduit pas les coûts de stockage des données de votre organisation. D'autres instantanés peuvent faire référence aux données de cet instantané, et les données référencées sont toujours conservées. Par exemple, lorsque vous prenez le premier instantané d'un volume comprenant 10 Gio de données, la taille de cet instantané est également de 10 Gio. Dans la mesure où les instantanés sont incrémentiels, le deuxième instantané du même volume contient uniquement les blocs de données modifiés dans le premier instantané. Le deuxième instantané fait également référence aux données du premier instantané. Si vous modifiez 4 GiB de données et que vous prenez un deuxième instantané, la taille du second instantané est de 4 GiB. De plus, le deuxième instantané fait référence aux 6 Gio inchangés dans le premier instantané. Pour plus d'informations, voir [Pourquoi mes coûts de stockage n'ont-ils pas diminué après avoir supprimé un instantané de mon volume EBS, puis supprimé le volume lui-même ?](#) dans le AWS Knowledge Center.

Éléments à prendre en compte :

- Les instantanés qu'un tiers Compte AWS possède et partage avec votre compte ne vous sont pas facturés. Vous êtes facturé uniquement lorsque vous copiez l'instantané partagé sur votre compte. Vous êtes également facturé pour les volumes EBS que vous créez à partir de l'instantané partagé.
- Si un instantané (instantané A) est référencé par un autre instantané (instantané B), la suppression du cliché B risque de ne pas réduire les coûts de stockage. Lorsque vous supprimez un instantané, seules les données propres à cet instantané sont supprimées. Les données référencées par d'autres instantanés sont conservées, et ces données référencées vous sont facturées. Pour supprimer un instantané incrémentiel, consultez la section Suppression [incrémentielle d'un instantané](#) dans la documentation Amazon EBS.

La propreté des snapshots est une pratique opérationnelle standard lors de l'exécution de vos charges de travail. AWS Au fil du temps, les instantanés peuvent entraîner des frais élevés pour des données dont vous n'avez pas besoin.

Ressources supplémentaires

- [Contrôlez vos AWS coûts en supprimant les volumes Amazon EBS inutilisés](#) (blog sur les opérations et migrations dans le AWS cloud)
- [Supprimer un instantané Amazon EBS](#) (documentation Amazon EBS)
- [Atelier sur l'optimisation des coûts](#) (AWS Well-Architected Labs)
- [Archivez automatiquement les instantanés Amazon EBS avec Amazon Data Lifecycle Manager](#) (AWS Storage Blog)

Supprimer des volumes Amazon EBS non attachés

Présentation de

Les volumes EBS non attachés (orphelins) peuvent entraîner des coûts de stockage inutiles dans votre environnement. AWS Il est essentiel d'intégrer la révision et la suppression régulières des volumes EBS inutilisés et inutilisés dans le cadre de l'hygiène de votre AWS environnement. Il est recommandé de mettre en place un processus permettant de contrôler en permanence l'utilisation des volumes EBS. Vous pouvez utiliser le [Optimiseur de calcul AWS](#) pour examiner les instances sous-utilisées. Cette section vous aide à identifier, gérer et supprimer les volumes EBS non connectés ou sous-utilisés.

Amazon EBS

[Amazon Elastic Block Store \(Amazon EBS\)](#) est un appareil au niveau des blocs qui propose des volumes de stockage pour les instances Amazon Elastic Compute Cloud (Amazon EC2). EBS fournit un stockage persistant, avec la flexibilité d'attacher et de détacher des instances EC2. Cela signifie que le cycle de vie des volumes EBS persiste même si une instance EC2 est résiliée. L'[DeleteOnTermination](#) attribut est une fonctionnalité qui contrôle s'il convient de conserver ou de supprimer les volumes EBS attachés lors de la fermeture de l'instance. Par défaut, l'attribut est défini sur `True` le volume racine, ce qui entraîne sa suppression. Il est configuré `False` pour les autres volumes, ce qui entraîne une préservation.

Impact sur les coûts

Les volumes EBS non attachés, également appelés volumes inutilisés ou orphelins, sont soumis aux mêmes frais que les volumes attachés en fonction de la taille et du type de stockage provisionnés. Bien que le coût moyen des frais Amazon EBS puisse sembler minime à 0,10\$ par Go par mois, il est essentiel de reconnaître que l'accumulation de volumes EBS inutilisés peut entraîner des coûts importants au fil du temps.

Par exemple, considérez les conséquences de la conservation de 50 volumes EBS inutilisés, chacun étant provisionné avec une taille de stockage de 100 Go, comme le montre le tableau suivant.

Nombre de volumes de stockage	Type de volume	Size	Coût mensuel total
50 tomes	gp2 (0,10\$ US)	100 Go	100 Go 50,00 volumes EBS par mois 0,10 USD = 500,00 USD

Le scénario du tableau précédent permet une réduction des coûts d'environ 500\$ par mois ou 6 000\$ par an. Il s'agit d'une étape efficace vers la réduction des coûts. Veillez à intégrer la suppression des volumes EBS non attachés comme pratique régulière dans le cadre de l'hygiène de votre AWS environnement.

Recommandations en matière d'optimisation des coûts

Vous pouvez l'utiliser AWS pour automatiser facilement la suppression des volumes EBS non attachés. Par exemple, vous pouvez utiliser AWS Lambda Amazon CloudWatch et pour définir des critères AWS Systems Manager de suppression de volumes non attachés en fonction de l'âge, des balises et d'autres spécifications. AWS Config Vous pouvez également les utiliser Services AWS pour automatiser le processus de nettoyage à grande échelle.

Pour éviter des conséquences imprévues, nous vous recommandons de faire preuve de diligence raisonnable avant de supprimer des volumes EBS non attachés.

Gérez les volumes EBS non attachés

Nous vous recommandons de prendre en compte les meilleures pratiques suivantes :

- Respectez les exigences de conformité : vérifiez que la suppression des volumes EBS non attachés est conforme aux exigences de gouvernance et de conformité de votre organisation.
- Définissez des politiques de sauvegarde et de conservation des données : avant de supprimer un volume EBS non attaché, sauvegardez toutes les données importantes dans un autre référentiel de stockage (par exemple, [Amazon S3](#)). Pour la conservation des données, [les instantanés Amazon EBS](#) constituent un moyen plus rentable de conserver les données que les volumes EBS, et ils peuvent restaurer le volume si nécessaire à l'avenir. Pour plus d'informations sur la gestion efficace des instantanés, consultez la section [Modifier les instantanés Amazon EBS](#) de ce guide.
- Vérifier les dépendances : vérifiez les éventuelles dépendances entre les volumes EBS non attachés et les autres AWS ressources. Vous pouvez utiliser l'API [AWS Management Console ou une API](#) pour recueillir des informations descriptives sur vos volumes EBS, telles que la taille, le statut et les ressources associées. Il s'agit d'une étape importante pour éviter de supprimer des ressources temporairement non connectées.
- Création d'une politique de rétention : établissez une période de conservation pour les volumes EBS non attachés. Cela peut vous aider à déterminer le moment approprié pour supprimer les volumes non attachés, afin de garantir que votre AWS environnement reste optimisé. Par exemple, vous pouvez créer une EventBridge règle [Amazon](#) pour lancer une fonction Lambda de manière planifiée. La fonction Lambda peut utiliser le AWS SDK pour identifier activement tout volume EBS non connecté, appliquer un mécanisme de balisage pour faciliter le suivi et envoyer des notifications lorsqu'un volume EBS non connecté atteint ou dépasse un seuil défini.
- Marquer les volumes EBS non attachés : le [balisage](#) des volumes EBS est une pratique utile qui peut aider à organiser et à identifier les volumes en fonction d'attributs tels que l'environnement, l'application ou le propriétaire. Cela peut être particulièrement utile lorsque vous décidez des volumes non attachés à supprimer, car cela vous permet d'identifier rapidement les volumes dont vous n'avez plus besoin en fonction de leurs balises.
- Garantir une suppression sécurisée : le fait de vérifier la date à laquelle un volume EBS a été connecté pour la dernière fois peut vous aider à déterminer s'il est possible de le supprimer en toute sécurité. Pour plus d'informations, consultez [Comment utiliser AWS CLI les commandes pour répertorier l'historique des pièces jointes ou des détachements d'un volume Amazon EBS spécifique ?](#) dans le AWS Knowledge Center.
- Identifier les volumes EBS sous-utilisés — L'identification et la suppression des volumes EBS sous-utilisés constituent une pratique hautement recommandée pour réduire les coûts de stockage et maintenir un environnement optimisé. AWS Trusted Advisor et [Optimiseur de calcul AWS](#) peut vous aider à identifier les volumes EBS sous-utilisés et à fournir des recommandations pour réduire les coûts et améliorer l'efficacité. Par exemple, consultez [Configuration de](#)

[l'automatisation pour optimiser les volumes EBS avec AWS Trusted Advisor \(GitHub\)](#), [Establishing a Trusted Advisor Organization \(TAO\) dashboard \(AWS Workshop Studio\)](#) et Optimisation des [coûts des volumes Amazon EBS à l'aide](#) de Optimiseur de calcul AWS(AWS Storage Blog).

Automatisez le nettoyage des volumes EBS non connectés

Nous vous recommandons d'utiliser les outils suivants pour vous aider à automatiser le nettoyage des volumes EBS non connectés :

- [AWS APIs \(DescribeVolumes\)](#) — Vous pouvez filtrer et rechercher des volumes EBS non attachés à l'aide d'AWS SDKs ou du AWS Command Line Interface (AWS CLI). Vous pouvez économiser du temps et des efforts en automatisant ce processus à l'aide d'un script ou d'une fonction [Lambda](#) qui s'exécute selon un calendrier. Un [exemple de script](#) tiré de GitHub montre comment cela fonctionne. Le script utilise Lambda pour analyser les AWS CloudTrail journaux et identifier les volumes EBS non connectés.
- [AWS Systems Manager Automatisation](#) : cela vous permet d'automatiser les tâches de maintenance et de correction de routine dans votre infrastructure. Pour commencer, [créez un runbook d'automatisation](#), qui définit une série d'étapes à exécuter dans un ordre spécifique. Par exemple, vous pouvez créer un runbook qui crée d'abord un instantané du volume EBS non connecté, puis supprime le volume lui-même. Cela peut vous aider à automatiser des tâches qui seraient autrement chronophages et sources d'erreurs si elles étaient effectuées manuellement.
- [AWS Config](#)— Cela vous permet d'évaluer, d'auditer et de suivre l'évolution de vos AWS ressources au fil du temps. En capturant les modifications de configuration, vous pouvez AWS Config évaluer la conformité, la gouvernance et l'utilisation des ressources dans votre environnement. Permet, par exemple, d' AWS Config identifier les [volumes EBS inutilisés](#). En outre, vous pouvez associer AWS Systems Manager Automation AWS Config à pour remédier automatiquement à la suppression des volumes EBS inutilisés.

Ressources supplémentaires

- [Supprimer les volumes Amazon Elastic Block Store \(Amazon EBS\) inutilisés en AWS Config utilisant AWS Systems Manager et AWS \(Prescriptive Guidance\)](#)
- [Contrôlez vos AWS coûts en supprimant les volumes Amazon EBS inutilisés](#) (blog sur les opérations et migrations dans le AWS cloud)
- [AWSConfigRemediation-DeleteUnusedEBSVolume](#)(Référence du manuel AWS Systems Manager d'automatisation)

Amazon FSx

Amazon FSx pour Windows File Server est un service de stockage de fichiers entièrement géré, optimisé pour les charges de travail Windows. Il vous fournit une solution simple et évolutive pour exécuter vos applications et charges de travail basées sur Windows, sans nécessiter de gestion complexe de l'infrastructure de stockage. Vous pouvez utiliser Windows File Server FSx pour configurer et accéder facilement au stockage de fichiers partagé qui prend en charge vos applications Windows de manière native, notamment Microsoft SQL Server SharePoint, Microsoft et les applications .NET personnalisées. En outre, FSx for Windows File Server vous aide à gérer les coûts en proposant des options de tarification flexibles, telles que les quotas de stockage, pay-as-you-go et la déduplication automatique des données afin de réduire l'encombrement du stockage et d'optimiser les performances et les coûts.

Cette section couvre les rubriques suivantes :

- [Choisissez le bon système de stockage de fichiers pour PME](#)
- [Activer la déduplication des données sur Amazon FSx](#)
- [Comprendre le partitionnement des données FSx pour le serveur de fichiers Windows](#)
- [Comprendre l'utilisation du volume du disque dur sur Amazon FSx](#)
- [Utiliser une seule zone de disponibilité](#)

Choisissez le bon système de stockage de fichiers pour PME

Présentation de

AWS propose une variété de services de stockage entièrement gérés qui vous offrent les riches fonctionnalités des services de fichiers de pointe, tout en combinant les dernières innovations en matière AWS d'infrastructure et de sécurité. Vous pouvez intégrer AWS des services dans des flux de travail d'infrastructure sous forme de code (IaC) et les intégrer à des services de AWS calcul, de surveillance et de protection des données. Pour les charges de travail Windows, vous pouvez choisir entre deux services de fichiers entièrement gérés qui peuvent être utilisés pour répondre aux besoins de vos applications : FSx pour Windows File Server et Amazon FSx pour NetApp ONTAP.

FSx pour le serveur de fichiers Windows

Amazon FSx pour Windows File Server fournit un stockage partagé entièrement géré basé sur Windows Server et propose un large éventail de fonctionnalités d'accès aux données, de gestion des données et d'administration. FSx pour Windows File Server s'intègre facilement aux environnements

Windows car il s'agit d'un service natif de Windows. Nous recommandons d'utiliser FSx pour le serveur de fichiers Windows pour les partages d'utilisateurs et de groupes, les instances de cluster Always On Failover pour SQL Server, les applications Windows et l'infrastructure de bureau virtuel (VDI). FSx pour Windows File Server s'intègre également bien à Amazon FSx File Gateway, Amazon Kendra, aux journaux d'audit pour Amazon S3 et Amazon Data Firehose.

FSx pour ONTAP

FSx for ONTAP est basé sur le système NetApp de fichiers propriétaire ONTAP. Cela demande un certain niveau de renforcement des compétences et est principalement recommandé aux utilisateurs locaux existants. NetApp Les cas d'utilisation typiques incluent les partages d'utilisateurs et de groupes, les instances de cluster Always On Failover pour SQL Server et les applications Windows. FSx for ONTAP prend en charge plusieurs protocoles, des systèmes de fichiers supérieurs à 64 To (échelle PB sans serveur d'espace de noms DFS), le clonage, la réplication, les instantanés, la compression (efficacité du stockage) et la hiérarchisation intelligente des données.

Impact sur les coûts

FSx pour le serveur de fichiers Windows

FSx pour Windows File Server a été la première solution de stockage partagé AWS permettant de déployer des instances de cluster Failover pour SQL Server. Avec FSx Windows File Server, vous pouvez lancer des instances Failover Cluster à l'aide d'une licence SQL Standard Edition. Cela vous empêche toutefois de vous fier aux groupes de disponibilité Always On, qui nécessitent des licences de l'édition Enterprise de SQL Server. En passant de l'édition SQL Server Enterprise Standard à l'édition SQL Server Standard, vous pourriez économiser 65 à 75 % sur vos licences [SQL Server](#).

Vous pouvez utiliser FSx for Windows File Server for Failover Cluster Instances pour décharger le I/O stockage du stockage EBS classique. En transférant I/O vers FSx for Windows File Server, vous pouvez réduire la taille des instances EC2, qui dépendent du débit et des IOPS élevés d'Amazon EBS, sans affecter le débit de stockage.

FSx pour ONTAP

Vous pouvez utiliser FSx for ONTAP pour exécuter votre cluster de basculement Microsoft sur le protocole iSCSI en mode bloc et bénéficier de l'initialisation instantanée des fichiers SQL Server, de l'utilisation de la réplication entre régions, de la prise en charge des antivirus et du clonage. SnapMirror Si vous créez plusieurs copies de bases de données à des fins de test, le clonage peut faire une différence significative en termes de consommation d'espace et de rapidité de création de ces copies de base de données. En outre, vous pouvez utiliser NetApp SnapCenter pour gérer

les fonctionnalités de sauvegarde, de restauration et de clonage avec vos instances EC2 pour SQL Server en utilisant FSx for ONTAP. FSx for ONTAP propose également une hiérarchisation automatique du SSD vers un stockage en pool à faible coût pour un mélange de performances et de rentabilité.

FSx car ONTAP supporte le système NetApp de fichiers (ONTAP), contrairement à FSx Windows File Server, qui prend en charge un système de fichiers NTFS natif de Windows. La taille minimale FSx pour ONTAP est de 1024 Go, tandis que FSx pour le serveur de fichiers Windows, elle peut commencer à 32 Go.

Intégration au système de fichiers distribué Microsoft

FSx pour Windows File Server et FSx ONTAP s'intègrent au [système de fichiers distribué \(DFS\)](#) de Microsoft pour une intégration fluide dans les déploiements existants. Gardez les points suivants à l'esprit lors de la planification de votre architecture :

- FSx pour Windows File Server et ONTAP supportent FSx les [espaces de noms DFS \(DFSN\)](#) sur les deux types de déploiement (zones de disponibilité multiples et zones de disponibilité uniques).
- Seul le serveur FSx de fichiers Windows prend en charge la [réplication DFS \(DFSR\)](#), et uniquement lors de l'utilisation de zones de disponibilité uniques.

Recommandations en matière d'optimisation des coûts

Les performances FSx de Windows File Server et FSx d'ONTAP dépendent en grande partie de la configuration, tout comme leur prix. FSx pour Windows File Server, la tarification dépend principalement de la capacité de stockage et du type de stockage, de la capacité de débit, de la sauvegarde et des données transférées. Avec FSx for ONTAP, vous payez pour le stockage SSD, les IOPS SSD, l'utilisation du pool de capacité, la capacité de débit et la sauvegarde.

Service de fichiers	Coût d'un stockage de 5 To	Configuration	Région
FSx pour le serveur de fichiers Windows	982,78\$	Zone de disponibilité unique SSD (15 000 IPS) 32 MBps	USA Est (Virginie du Nord)

Service de fichiers	Coût d'un stockage de 5 To	Configuration	Région
		Sauvegarde de 5 To (aucune économie de déduplication)	
FSx pour ONTAP	979,28\$	Zone de disponibilité unique 100 % SSD Niveau de capacité de lecture et d'écriture de 15 000 15 000 IOPS SSD 128 MBps Sauvegarde de 5 To (aucune économie de déduplication)	USA Est (Virginie du Nord)

Gardez à l'esprit les points suivants :

- La déduplication et la compression vous permettent de stocker davantage de données sur des appareils physiques en réduisant leur taille, mais vous payez pour le stockage sur disque SSD ou disque dur (HDD) provisionné.
- Vous pouvez utiliser ONTAP FSx pour hiérarchiser vos données. Il est extrêmement rare que 100 % de vos données soient consultées régulièrement et nécessitent un stockage SSD. Vous pouvez déplacer les données inutilisées et peu consultées vers un niveau de capacité afin de réaliser des économies.
- Les prix mentionnés ici sont calculés sur la base de données à 100 % sur le niveau SSD et de 15 000 IOPS sur le niveau SSD.

Sauvegarde

Par défaut, ONTAP et FSx Windows File Server stockent leur sauvegarde entièrement gérée sur Amazon S3. FSx Cependant, avec FSx ONTAP, il existe une option supplémentaire d'utilisation de la sauvegarde SnapVault, qui permet de configurer les sauvegardes pour qu'elles résident dans le niveau de capacité. La sauvegarde par le SnapVault biais d'un mécanisme autogéré est plus rentable que l'option de sauvegarde entièrement gérée par défaut. L'option de sauvegarde entièrement gérée est de 0,05 USD par Go par mois. La SnapVault sauvegarde activée FSx pour ONTAP (SSD 10:1 par rapport à la capacité de stockage du pool) coûte 0,03221\$ (0,9 x 0,0219+0,1 x 0,125).

Gardez à l'esprit les points suivants :

- AWS les sauvegardes gérées offrent une granularité d'une heure. [SnapVault](#) vous permet d'y aller jusqu'à cinq minutes.
- Vous pouvez utiliser NetApp les outils (tels que la CLI et l'API) pour configurer SnapVault les relations et la réplication des instantanés.
- Activez la politique all de hiérarchisation sur un SnapVault volume afin d'utiliser le niveau de capacité comme stockage pour les données de sauvegarde.
- SnapVault les destinations peuvent être identiques Région AWS, interrégionales ou locales. Il s'agit généralement d'une destination de sauvegarde du système de fichiers d'une seule zone de disponibilité ou de plusieurs zones de disponibilité. En comparaison, AWS Backup elle est soutenue par la résilience régionale d'Amazon S3.

Bonne taille

Vous pouvez également économiser sur les coûts et tirer le meilleur parti de votre système de fichiers en le dimensionnant correctement et en évitant le surprovisionnement.

Pour obtenir la bonne taille, procédez comme suit :

1. Identifiez vos besoins actuels en vous basant sur les données. Pour les charges de travail Windows classiques, vous pouvez utiliser des outils de système d'exploitation intégrés tels que [Performance Monitor](#).
2. Dans Performance Monitor, utilisez les compteurs suivants pour évaluer vos besoins actuels en termes de performances. L'intervalle de capture est défini sur une seconde, avec une taille de journal maximale de 1 000 Mo et le remplacement est activé.

```
Logman.exe create counter PerfLog-Short -o "c:\perflogs\PerfLog-Long.blg" -f bincirc  
-v mmddhhmm -max 1024 -c "\\LogicalDisk(*)\*" "\\Memory\*" "\.NET CLR Memory(*)\*"  
"\\Cache\*" "\\Network Interface(*)\*" "\\Paging File(*)\*" "\\PhysicalDisk(*)\*"  
"\\Processor(*)\*" "\\Processor Information(*)\*" "\\Process(*)\*" "\\Thread(*)\*"  
"\\Redirector\*" "\\Server\*" "\\System\*" "\\Server Work Queues(*)\*" "\\Terminal  
Services\*" -si 00:00:01
```

3. Pour démarrer la capture du journal, exécutez la `logman start PerfLog-Short` commande. Pour arrêter la capture du journal, exécutez la `logman stop PerfLog-Short` commande.

Note

Les fichiers journaux des performances se trouvent dans `c:\perflogs` sur le serveur exécutant la capture. Pour plus d'informations, consultez la section [Présentation de Windows Performance Monitor](#) dans la documentation Microsoft.

4. Après avoir identifié la configuration correcte, testez si votre estimation est correcte sur le système de FSx fichiers Amazon en utilisant des outils de gestion du stress du disque tels que Microsoft [DISKSPD](#).
5. Si vous êtes satisfait des performances, passez au partage de fichiers.

Nous recommandons d'adopter une approche prudente en matière de capacité de stockage, car elle ne peut qu'être augmentée. La capacité de débit peut être augmentée ou diminuée selon les besoins.

Ressources supplémentaires

- [Amazon FSx pour NetApp ONTAP FAQs](#) (AWS site Web)
- [Optimisation des performances du serveur de fichiers Amazon FSx pour Windows grâce à de nouvelles mesures](#) (AWS Storage Blog)

Activer la déduplication des données sur Amazon FSx

Présentation de

La déduplication des données est une fonctionnalité qui vous permet de stocker vos données de manière plus efficace tout en réduisant les besoins en capacité. Cela implique de détecter et de supprimer les doublons dans les données sans compromettre leur fidélité ou leur intégrité. La

déduplication des données utilise le découpage et la compression de sous-fichiers à taille variable, ce qui permet d'obtenir des ratios d'optimisation de 2:1 pour les serveurs de fichiers généraux et jusqu'à 20:1 pour les données de virtualisation. La déduplication des données est bien plus efficace que la compression NTFS. La résilience en cas de défaillance matérielle est inhérente à l'architecture de déduplication, avec une validation complète des données et des métadonnées, y compris la redondance des métadonnées et des segments de données les plus consultés.

FSx pour Windows File Server prend entièrement en charge la déduplication des données. Son utilisation peut permettre de réaliser des économies moyennes de 50 à 60 % sur les partages de fichiers à usage général. En ce qui concerne les actions, les économies varient de 30 à 50 % pour les documents utilisateur et jusqu'à 70 à 80 % pour les ensembles de données de développement logiciel. Il est important de comprendre que les économies de stockage que vous pouvez réaliser grâce à la déduplication des données dépendent de la nature de votre ensemble de données, notamment du niveau de duplication existant entre les fichiers. La déduplication n'est pas une bonne option si les données stockées sont de nature dynamique.

Impact sur les coûts

Pour faire face à la croissance du stockage de données dans l'entreprise, les administrateurs consolident les serveurs et fixent des objectifs clés en matière de mise à l'échelle des capacités et d'optimisation des données. Les paramètres par défaut de la déduplication des données peuvent permettre de réaliser des économies immédiates, ou les administrateurs peuvent affiner les paramètres pour obtenir des gains supplémentaires. Par exemple, vous pouvez configurer la déduplication pour qu'elle s'exécute uniquement sur certains types de fichiers, ou vous pouvez créer un calendrier de travail personnalisé.

À un niveau élevé, la déduplication comporte trois types de tâches : optimisation, collecte des déchets et nettoyage. Sachez que l'espace ne sera pas libéré tant que vous n'aurez pas exécuté une tâche de collecte des déchets après l'optimisation. Vous pouvez planifier le travail ou l'exécuter manuellement. Tous les paramètres disponibles lorsque vous planifiez une tâche de déduplication de données sont également disponibles lorsque vous démarrez une tâche manuellement (à l'exception de ceux qui sont spécifiques à la planification).

Même avec seulement 25 % d'économies effectives grâce à la déduplication, Windows File Server peut réaliser FSx des économies considérables. Ces économies prévues sont basées sur une [estimation figurant](#) dans le Calculateur de tarification AWS.

Recommandations en matière d'optimisation des coûts

La déduplication activée FSx pour les systèmes de fichiers Windows File Server n'est pas activée par défaut. Pour activer la déduplication en activant [la gestion à distance PowerShell](#), vous devez exécuter la `Enable-FSxDedup` commande, puis l'`Set-FSxDedupConfiguration` utiliser pour définir la configuration. Pour plus d'informations, consultez la section [Administration des systèmes de fichiers](#) dans la documentation relative FSx aux serveurs de fichiers Windows.

Pour activer la déduplication, exécutez la commande suivante :

```
PS C:\Users\Admin> Invoke-Command -ComputerName amznfsxzzzzzzzzz.corp.example.com -  
ConfigurationName FSxRemoteAdmin -ScriptBlock {Enable-FsxDedup }
```

Pour vérifier votre configuration de déduplication, exécutez la commande suivante :

```
Invoke-Command -ComputerName amznfsxzzzzzzzzz.corp.example.com -ConfigurationName  
FSxRemoteAdmin -ScriptBlock {  
Set-FSxDedupSchedule -Name "CustomOptimization" -Type Optimization -Days  
Mon,Tues,Wed,Sat -Start 09:00 -DurationHours 7  
}
```

En exécutant l' PowerShell `Measure-DedupFileMetadata` applet de commande, vous pouvez déterminer l'espace disque potentiel pouvant être récupéré sur un volume si vous supprimez un groupe de dossiers, un seul dossier ou un seul fichier, puis si vous exécutez une tâche de collecte des déchets. Plus précisément, la `DedupDistinctSize` valeur vous indique l'espace que vous pouvez récupérer si vous supprimez ces fichiers. Les fichiers contiennent souvent des fragments partagés entre d'autres dossiers. Le moteur de déduplication calcule donc quels fragments sont uniques et seront supprimés après le travail de collecte des déchets.

Les programmes de [déduplication des données par défaut](#) sont conçus pour fonctionner correctement pour les charges de travail recommandées et être aussi peu intrusifs que possible (à l'exception de la tâche d'optimisation prioritaire activée pour le type d'utilisation de sauvegarde). Si les charges de travail nécessitent des ressources importantes, nous vous recommandons de planifier l'exécution des tâches uniquement pendant les heures d'inactivité, ou de réduire ou d'augmenter la quantité de ressources système qu'une tâche de déduplication de données est autorisée à consommer.

Par défaut, la déduplication des données utilise 25 % de la mémoire disponible. Cependant, cela peut être augmenté en utilisant `-memory switch`. Pour les tâches d'optimisation, nous vous

recommandons de définir une plage comprise entre 15 et 50. Pour les tâches planifiées, vous pouvez utiliser une consommation de mémoire plus élevée. Par exemple, pour les tâches de collecte des déchets et de nettoyage (que vous planifiez généralement pour être exécutées en dehors des heures de bureau), vous pouvez définir une consommation de mémoire plus élevée (50 par exemple).

Pour plus d'informations sur les paramètres de déduplication des données, consultez la section [Réduction des coûts de stockage grâce à la déduplication des données](#) dans la documentation relative FSx aux serveurs de fichiers Windows.

Ressources supplémentaires

- [Comprendre la déduplication des données](#) (documentation Microsoft)
- [Réduction des coûts de stockage grâce à la déduplication des données](#) (FSx pour la documentation du serveur de fichiers Windows)

Comprendre le partitionnement des données FSx pour le serveur de fichiers Windows

Présentation de

FSx pour Windows File Server, les performances dépendent de la configuration. Il est principalement basé sur le type de stockage, la capacité de stockage et la configuration du débit. La capacité de débit que vous sélectionnez détermine les ressources de performance disponibles pour le serveur de fichiers, notamment les I/O limites du réseau, le processeur et la mémoire, ainsi que les I/O limites de disque imposées par le serveur de fichiers. La capacité de stockage et le type de stockage que vous sélectionnez déterminent les ressources de performance disponibles pour les volumes de stockage, c'est-à-dire les I/O limites de disque imposées par les disques de stockage. Outre les performances, les choix de configuration influencent également le coût. FSx pour Windows File Server, la tarification dépend principalement de la capacité de stockage et du type de stockage, de la capacité de débit, de la sauvegarde et des données transférées.

Si vous avez des exigences relativement importantes en termes de stockage de fichiers et de performances, vous pouvez tirer parti du découpage des données. Le découpage des données consiste à [diviser les données de vos fichiers](#) en ensembles de données plus petits (fragments) et à les stocker dans différents systèmes de fichiers. Les applications accédant à vos données à partir de plusieurs instances peuvent atteindre des niveaux de performance élevés en lisant et en écrivant sur ces partitions en parallèle. Dans le même temps, vous pouvez toujours présenter une vue unifiée sous un espace de noms commun à vos applications. En outre, cela peut également aider à étendre

le stockage des données de fichiers au-delà de ce que chaque système de fichiers prend en charge (64 To) pour les grands ensembles de données de fichiers, jusqu'à des centaines de pétaoctets.

Impact sur les coûts

Pour les grands ensembles de données, il est généralement plus efficace de déployer plusieurs petits systèmes de fichiers FSx pour Windows File Server, plutôt qu'un seul grand partage SSD pour atteindre le même niveau de performance. L'utilisation d'une combinaison de types de stockage HDD et SSD FSx pour Windows File Server permet de réaliser de meilleures économies et de faire correspondre la charge de travail au meilleur sous-système de disque sous-jacent. Dans les tableaux suivants, vous pouvez voir la différence entre un seul système de fichiers de 17 To et le comparer à plusieurs systèmes de fichiers plus petits qui ajoutent la même capacité.

Système de fichiers SSD de grande taille avec plusieurs charges de travail

Server name	Cost	Configuration	Région
Serveur FSx de fichiers Amazon pour Windows	5 716\$ USD	SSD DE 17 TO Déduplication de 30 % 256 Mbits/s Sauvegarde de 17 To	USA Est (Virginie du Nord)

Charge de travail partitionnée à l'aide de DFSN

Server name	Cost	Configuration	Région	Partager
Serveur FSx de fichiers Amazon pour Windows	1 024 DOLLARS AMÉRICAINS	SSD DE 2 TO 20 % de déduplication 128 Mbits/s Sauvegarde de 2 To	USA Est (Virginie du Nord)	Partagez 1

Server name	Cost	Configuration	Région	Partager
		Multi-AZ		
Serveur FSx de fichiers Amazon pour Windows	2,132\$ USD	SSD DE 5 TO 30 % de déduplication 256 Mbits/s Sauvegarde de 5 To Multi-AZ	USA Est (Virginie du Nord)	Partagez 2
Serveur FSx de fichiers Amazon pour Windows	1 036\$ USD	DISQUE DUR DE 10 TO Déduplication de 40 % 128 Mbits/s Sauvegarde de 10 To Multi-AZ	USA Est (Virginie du Nord)	Partagez 3
Instances Windows EC2 DFSN	27 DOLLARS AMÉRICAINS	t3a.medium 2 v CPUs Mémoire de 4 GiB	USA Est (Virginie du Nord)	Instances DFSN

Le coût annuel d'un système de fichiers SSD de grande taille est de 68 592\$. Le coût annuel d'une charge de travail partitionnée est de 50 640\$. Dans cet exemple, il est possible de réaliser des économies de 26 % tout en adaptant la charge de travail au stockage principal approprié. Pour plus d'informations sur l'estimation des prix, consultez l'[Calculateur de tarification AWS](#) estimation.

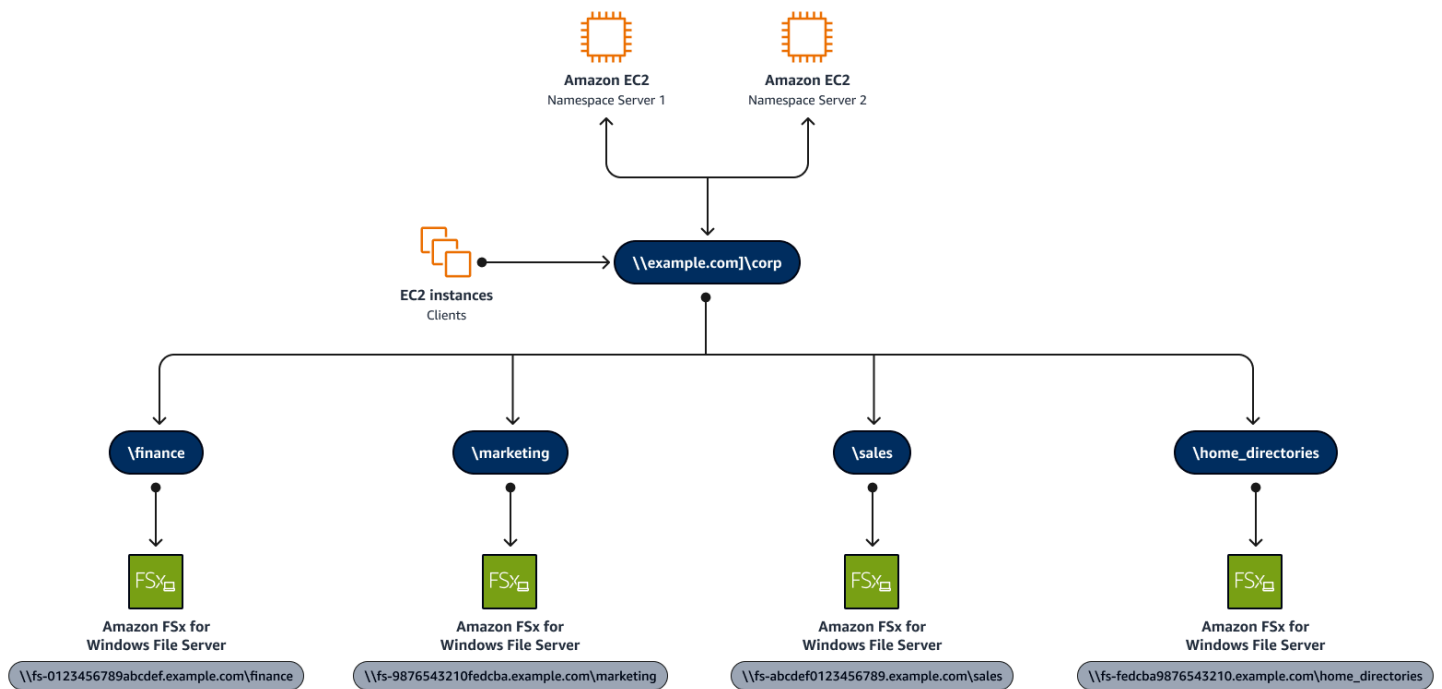
Recommandations en matière d'optimisation des coûts

Pour déployer une solution de déduplication des données, vous devez configurer un espace de [noms Microsoft DFS](#) en fonction du type de données, de I/O leur taille et I/O du modèle d'accès. Chaque espace de noms prend en charge jusqu'à 50 000 partages de fichiers et des centaines de pétaoctets de capacité de stockage au total.

Le plus efficace est de choisir une convention de partitionnement qui répartit les E/S de manière uniforme sur tous les systèmes de fichiers que vous prévoyez d'utiliser. La surveillance de votre charge de travail contribuera à une optimisation supplémentaire ou à une réduction des coûts. Si vous avez besoin d'aide pour évaluer les informations de performance du système de FSx fichiers Amazon, consultez la documentation relative aux [serveurs de fichiers Windows FSx FSx pour les performances](#) du serveur de fichiers Windows.

Après avoir choisi une stratégie de partitionnement, vous pouvez regrouper les systèmes de fichiers pour accéder facilement à vos partages à l'aide des espaces de noms DFS. Cela permet aux utilisateurs de voir un système de fichiers homogène, alors qu'en réalité ils accèdent à une variété de systèmes de fichiers différents avec des cas d'utilisation spécialement conçus. Il est important de créer les partages selon une convention de dénomination appropriée afin que vos utilisateurs finaux puissent facilement déchiffrer la charge de travail pour laquelle les partages sont conçus. Il est également important d'étiqueter les partages de production et les partages hors production, afin que les utilisateurs finaux ne placent pas les fichiers dans le mauvais système de fichiers par erreur.

Le schéma suivant montre comment un seul espace de noms DFS peut être utilisé comme point d'accès pour plusieurs systèmes de FSx fichiers Amazon.



Gardez à l'esprit les points suivants :

- Vous pouvez ajouter des partages existants FSx pour Windows File Server à une arborescence DFS.
- Amazon ne FSx peut pas être ajouté à la racine du chemin de partage DFS. Vous n'avez qu'un seul sous-dossier.
- Vous devez déployer une instance EC2 pour servir la configuration de l'espace de noms DFS.

Pour plus d'informations sur la configuration DFS-N, consultez la présentation des [espaces de noms DFS dans la documentation Microsoft](#). Pour plus d'informations sur l'utilisation des espaces de noms DFS, consultez la vidéo [Utilisation des espaces de noms DFS avec FSx Amazon pour Windows File Server](#) sur le sujet. YouTube

Ressources supplémentaires

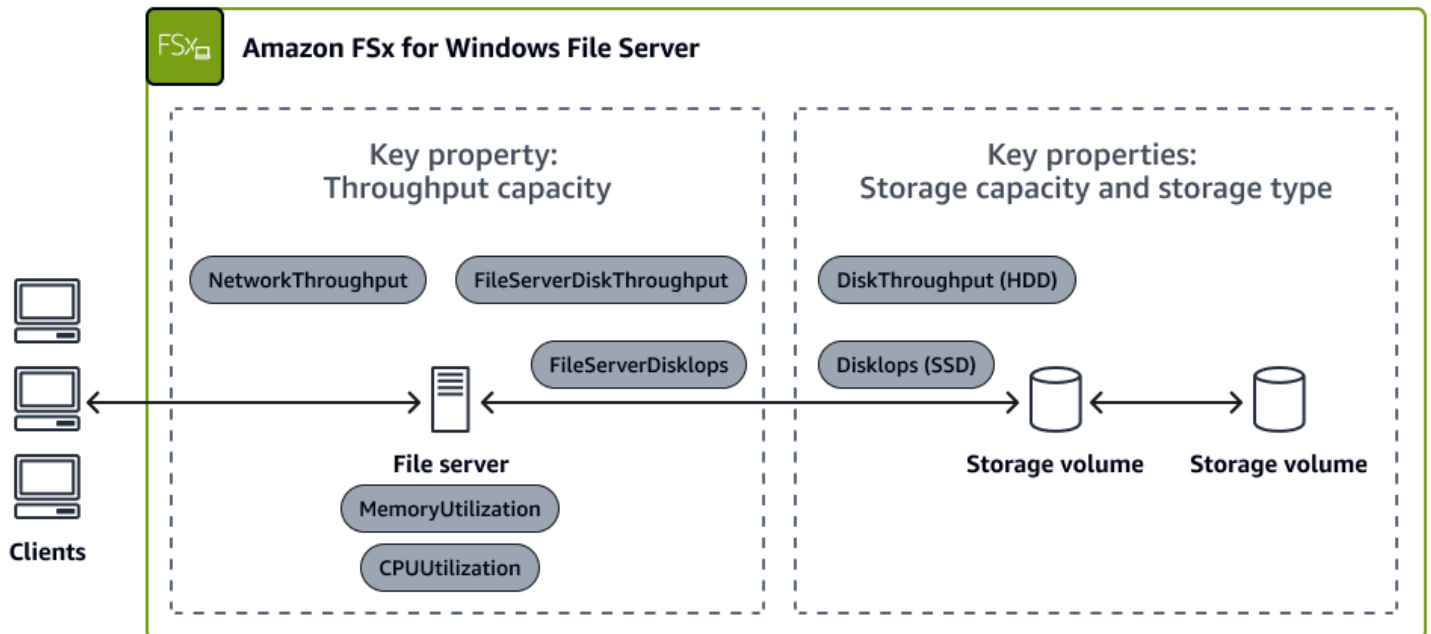
- [Regroupement de plusieurs systèmes de fichiers avec des espaces de noms DFS](#) (documentation Amazon FSx)
- [Procédure pas à pas 6 : augmenter les performances avec des partitions](#) (documentation Amazon FSx)
- [Utilisation des espaces de noms DFS avec le serveur de fichiers Amazon FSx pour Windows](#) (AWS Labs)

Comprendre l'utilisation du volume du disque dur sur Amazon FSx

Présentation de

Amazon FSx pour Windows File Server offre la flexibilité de choisir le débit indépendamment de la capacité du système de fichiers. Deux paramètres de capacité sont disponibles : disque dur (HDD) et disque SSD (Solid State Drive).

Le schéma suivant montre la relation entre le débit et les paramètres de stockage.



Avec le stockage sur disque dur, vous bénéficiez d'une ligne de base de 12 IOPS avec 80 IOPS par rafale (par IOPs TiB de stockage) et d'un débit de 12 lignes de base Megabytes/second avec 80 rafales (Megabytes/second par TiB de stockage). Par exemple, si la taille de votre partage est de 50 To, vous obtenez $50 \times 12 = 600$ comme base de référence pour le débit et les IOPS.

Le serveur de fichiers Amazon FSx pour Windows fournit 80 IOPS en rafale. Les crédits Burst sont automatiquement renouvelés lorsque votre utilisation est inférieure à votre taux de base et sont automatiquement consommés lorsque votre utilisation est supérieure à votre taux de référence. Par exemple, si votre charge de travail n'en utilise que 10 IOPS/TB pendant une heure (2 de IOPS/ TB below your baseline rate), you can then utilize 14 IOPS/TB (2 IOPS/TB plus que votre base de référence) pendant l'heure suivante avant de manquer à nouveau de crédits en rafale.

Pour les opérations de fichiers, le serveur de fichiers Amazon FSx pour Windows fournit des latences constantes inférieures à la milliseconde avec le stockage SSD et des latences à un chiffre en

millisecondes avec le stockage sur disque dur. Pour tous les systèmes de fichiers, y compris ceux dotés d'un stockage sur disque dur, le serveur de fichiers Amazon FSx pour Windows fournit un cache rapide (en mémoire) sur le serveur de fichiers, ce qui vous permet d'obtenir des performances élevées et des latences inférieures à la milliseconde pour les données auxquelles vous accédez activement, quel que soit le type de stockage.

Le cas échéant, l'utilisation du stockage sur disque dur peut contribuer à réduire le coût de votre capacité de stockage globale et à fournir une plate-forme de stockage fiable répondant à vos besoins.

Impact sur les coûts

Les performances du serveur de fichiers Amazon FSx pour Windows dépendent de trois facteurs : la capacité de stockage, le type de stockage et le débit. Les I/O performances du réseau et la taille du cache en mémoire sont uniquement déterminées par la capacité de débit, tandis que les I/O performances du disque sont déterminées par une combinaison de capacité de débit, de type de stockage et de capacité de stockage.

Bien que le SSD soit recommandé pour les charges de travail I/O intensives, il existe une variété de charges de travail dont les besoins peuvent être satisfaits grâce aux spécifications de performance des disques durs. Ce type de stockage sur HDD a été pensé pour un large éventail de charges de travail, notamment les répertoires de base, les partages entre utilisateurs et services, et les systèmes de gestion des contenus. Par exemple, si vos utilisateurs ont uniquement besoin d'un accès à faible latence aux données relatives aux projets en cours, la plupart des données que vous stockez sont rarement consultées.

Vous pouvez utiliser le [Calculateur de tarification AWS](#) pour comparer un SSD de 20 To à un système de fichiers HDD dans us-east-1. Comme le montre le tableau suivant, même en l'absence d'économies liées à la déduplication, la différence de coût est significative lorsque l'on compare les systèmes de fichiers HDD aux systèmes de fichiers SSD.

Configuration FSx du système de fichiers Amazon	Coûts mensuels
SSD Multi-AZ de 20 To () us-east-1	4 699,30\$
Disque dur multi-AZ de 20 To () us-east-1	542,88\$
Économies mensuelles estimées	4 156,42\$

Note

FSx Pour plus d'économies sur les serveurs de fichiers Windows, consultez la FSx section [Activer la déduplication des données sur Amazon](#) de ce guide.

En identifiant correctement vos besoins en termes de performances, vous pouvez sélectionner le stockage adapté à votre charge de travail et réduire vos coûts.

Recommandations en matière d'optimisation des coûts

Si vous décidez d'utiliser le stockage sur disque dur, testez votre système de fichiers pour vous assurer qu'il répond à vos exigences de performance. Le stockage sur disque dur est moins coûteux que le stockage SSD, mais avec des niveaux de débit et d'IOPS par unité de stockage inférieurs. Il peut convenir aux partages utilisateur à usage général et aux répertoires personnels peu I/O exigeants, aux systèmes de gestion de contenu volumineux dans lesquels les données sont rarement récupérées, ou aux ensembles de données contenant un petit nombre de fichiers volumineux.

Le type de stockage d'un système de fichiers existant ne peut pas être modifié. Pour convertir le type de stockage d'un système de fichiers Amazon FSx pour Windows File Server, vous devez sauvegarder votre système de fichiers existant et le restaurer sur un nouveau système de fichiers avec le type de stockage souhaité. Si vous souhaitez convertir un système de fichiers SSD existant en un système de fichiers HDD, sachez que le disque dur a une capacité minimale beaucoup plus élevée de 2 To.

Pour restaurer une sauvegarde avec un autre type de stockage, procédez comme suit :

1. [Sauvegardez votre système de fichiers existant.](#)
2. [Créez un nouveau système de FSx fichiers Amazon](#) avec le type de stockage HDD.
3. Restaurez la sauvegarde sur le nouveau système de fichiers avec le type de stockage souhaité.
4. Vérifiez que le nouveau système de fichiers dispose du type de stockage approprié et que vos données sont intactes.

Avant de transférer vos modifications vers la production, nous vous recommandons d'analyser les performances de votre système de FSx fichiers Amazon et de vérifier que les modifications sont acceptables. Pour plus d'informations, consultez l'article [Optimisation des performances des serveurs de fichiers Amazon FSx pour Windows avec de nouveaux indicateurs](#) sur le blog sur le AWS stockage.

Ressources supplémentaires

- [Optimisation des coûts avec Amazon FSx](#) (FSx documentation Amazon)

Utiliser une seule zone de disponibilité

Présentation de

Cette section explique dans quels cas il est plus avantageux d'utiliser une seule implémentation de zone de disponibilité d'[Amazon FSx pour Windows File Server](#). Il couvre les scénarios dans lesquels le passage à une zone de disponibilité unique réduit les coûts, tout en vous permettant d'utiliser Amazon FSx pour Windows File Server comme service de stockage de fichiers géré. Nous vous recommandons de mettre en œuvre une zone de disponibilité unique pour Amazon FSx pour les charges de travail de production. Cela permet de garantir la redondance de plusieurs zones de disponibilité.

Impact sur les coûts

Un système de fichiers à zone de disponibilité unique permet de réduire les coûts d'environ 40 % par rapport à une mise en œuvre de plusieurs zones de disponibilité. Avec un système de fichiers à zones de disponibilité multiples, vous payez 0,230 USD par Go par mois en SSD et 0,025 USD par Go par mois en disque dur, contre 0,130 USD par Go par mois pour le SSD et 0,013 USD par Go par mois pour le disque dur sur un système de fichiers à zone de disponibilité unique. Vous pouvez consulter une comparaison des coûts et créer vos propres estimations à l'aide du [Calculateur de tarification AWS](#).

Pour un système de fichiers de 10 To, cela peut représenter une différence d'environ 1 200\$ par mois pour plusieurs zones de disponibilité ou 680\$ par mois pour une seule zone de disponibilité. Cet [exemple](#) utilise un système de fichiers de 10 To FSx pour Windows File Server avec SSD. Les économies réalisées grâce à la déduplication sont estimées à 50 %. Dans l'ensemble, une seule zone de disponibilité présente un coût d'entrée inférieur, mais comporte quelques réserves qui sont abordées dans la section suivante.

Recommandations en matière d'optimisation des coûts

Déploiements dans une seule zone de disponibilité

Pour vous assurer qu'une seule zone de disponibilité convient, prenez en compte votre propre zone interne SLAs pour les données stockées sur FSx Windows File Server. Cela implique de

comprendre si vous devez fournir SLAs à vos clients (internes et externes) et si les trois neuf jours de disponibilité d'une zone de disponibilité FSx unique d'Amazon vous permettront toujours de les SLAs respecter. FSx pour un serveur de fichiers Windows doté d'une seule zone de disponibilité, le temps de disponibilité est toujours de 99,9 %. Le SLA d'Amazon FSx pour plusieurs zones de disponibilité est supérieur à 99,99 %. Pour les charges de travail critiques, nous vous recommandons d'utiliser plusieurs zones de disponibilité sur une seule zone de disponibilité, même moyennant des frais supplémentaires.

Les déploiements par zone de disponibilité unique sont idéaux pour les charges de travail telles que les sauvegardes de bases de données SQL Server. Ils peuvent fournir un stockage à faible coût avec un niveau de disque dur, tout en vous garantissant une disponibilité constante. Si vous avez besoin d'un niveau de disponibilité plus élevé pour une charge de travail de production, telle que des serveurs SQL hautement disponibles ou un accès aux applications de production, une zone de disponibilité unique ne convient pas à vos charges de travail. Pour les sauvegardes, les tests hors production et les environnements de développement, la mise en œuvre d'une zone de disponibilité FSx unique d'Amazon peut réduire vos coûts d'exploitation.

Un système de fichiers à zone de disponibilité FSx unique Amazon fonctionne bien dans une situation de production où plusieurs systèmes de fichiers Amazon à zone de disponibilité FSx unique sont utilisés, en tant que stockage par serveur dans un cluster SQL Server à haute disponibilité utilisant des groupes de disponibilité Always On. Pour plus d'informations, consultez l'article [Optimisation des coûts pour vos déploiements SQL Server à haute disponibilité AWS publié sur le blog sur le AWS stockage](#).

Réplication multi-régions

Une option potentielle pour réduire les coûts avec un système de fichiers de zone de disponibilité unique (un système de fichiers dans lequel un seul système de fichiers de zone de disponibilité fonctionne) consiste à tirer parti d'une réplication multirégionale avec Amazon FSx. Vous pouvez déployer des [systèmes de fichiers mono-AZ](#) compatibles avec Microsoft DFS-R natif. Le DFS-R est capable de répliquer automatiquement les données entre les régions et sur plusieurs sites. Pour plus d'informations sur la configuration du DFS-R à l'aide d'Amazon FSx, consultez la section Utilisation de la [réplication de systèmes de fichiers distribués Microsoft](#) dans la documentation Amazon FSx.

Une autre alternative pour réaliser des économies sur plusieurs régions consiste à utiliser AWS Storage Gateway. Cela vous permet d'implémenter une [passerelle de FSx fichiers Amazon](#) dans une autre région pour un accès multirégional à Amazon FSx. Pour de plus amples informations, veuillez consulter la section [AWS Storage Gateway](#) de ce guide.

Si vous travaillez dans plusieurs régions, vous devez tenir compte du coût du transfert de données pour le trafic de données entre régions. Le trafic circulant entre les régions entraîne des frais de 0,02 \$/Go. Ainsi, si vous constatez des modifications constantes de données à des volumes élevés, cela augmentera votre coût global. Par [exemple](#), 1 To de transfert de données équivaut à environ 20,48\$.

Fenêtre de maintenance

La fenêtre de maintenance est un élément clé si vous utilisez une zone de disponibilité unique avec Amazon FSx. Pendant la période de maintenance, le système de FSx fichiers Amazon est indisponible pendant environ 20 minutes, en raison de l'application de correctifs logiciels de routine pour le serveur Windows sous-jacent. Si vous utilisez le système de fichiers pour des sauvegardes pendant la nuit, ajustez la fenêtre de FSx maintenance Amazon en conséquence afin d'éviter toute interruption lors de votre sauvegarde. Vous pouvez ajuster la [fenêtre de maintenance](#) après avoir créé votre système de FSx fichiers Amazon.

Ressources supplémentaires

- [Disponibilité et durabilité : systèmes de fichiers mono-AZ et multi-AZ \(documentation Amazon FSx\)](#)
- [Tarification FSx des serveurs de fichiers Amazon pour Windows](#) (AWS site Web)

AWS Storage Gateway

AWS Storage Gateway est un service de stockage dans le cloud hybride qui connecte les environnements sur site au stockage dans AWS le cloud. Il vous permet d'intégrer facilement votre infrastructure sur site existante AWS, ce qui vous permet de stocker et de récupérer des données depuis le cloud et d'exécuter des applications dans un environnement hybride. Pour les charges de travail Windows, vous pouvez utiliser Storage Gateway pour stocker des données et y accéder à l'aide de protocoles Windows natifs tels que SMB et NFS. Vous pouvez utiliser Storage Gateway pour réduire les coûts associés à l'exécution des charges de travail Windows AWS en utilisant du matériel et des logiciels sur site comme passerelle vers le cloud. Cela vous permet de tirer parti de l'évolutivité et de la rentabilité de votre infrastructure existante AWS sans avoir à apporter de modifications importantes à votre infrastructure existante.

Sous l'égide de Storage Gateway, vous pouvez bénéficier d'Amazon S3 File Gateway, d'Amazon FSx File Gateway, de Tape Gateway et de Volume Gateway. La passerelle de fichiers S3 et la passerelle de FSx fichiers sont le plus souvent utilisées avec les charges de travail Microsoft.

Passerelle de fichiers Amazon S3

[Amazon S3 File Gateway](#) vous permet de stocker vos fichiers dans Amazon S3 tout en fournissant un accès à vos utilisateurs en utilisant des partages SMB traditionnels. Cela fournit une interface utilisateur familière et permet de réduire les coûts en stockant vos données dans Amazon S3 et en tirant parti des différents niveaux de stockage d'Amazon S3. Vous pouvez implémenter Storage Gateway avec S3 Intelligent Tiering pour vous aider à déplacer automatiquement les fichiers du cycle de vie vers les niveaux de stockage les moins coûteux afin de réduire encore vos coûts. Nous recommandons S3 File Gateway pour l'évolutivité, l'accès en lecture seule, les lectures répétées rapides (depuis le cache) et les vidages de bases de données. Il n'est généralement pas recommandé pour les écritures à hautes performances ou à haute disponibilité, l'édition de fichiers ou les partages départementaux.

Passerelle FSx de fichiers Amazon

[Amazon FSx File Gateway](#) peut également permettre de réaliser des économies lorsque vous travaillez avec des systèmes de fichiers Amazon FSx Windows. Vous pouvez configurer une passerelle de FSx fichiers pour fournir un accès localisé à un système de FSx fichiers Amazon dans une autre région afin d'éviter les coûts liés à la présence de deux systèmes de fichiers indépendants. Cela peut également être utile si vous disposez de plusieurs serveurs de fichiers sur site et que vous souhaitez les consolider afin d'éviter de payer pour plusieurs périphériques matériels.

Impact sur les coûts

Passerelle de fichiers Amazon S3

La configuration de S3 File Gateway est simple car vous pouvez utiliser l'assistant de lancement de Storage Gateway. Vous pouvez déployer la passerelle en quelques minutes en utilisant une instance EC2 dans votre AWS environnement. Une fois la passerelle configurée, vous pouvez configurer les partages Storage Gateway pour qu'ils soient accessibles via les protocoles SMB et NFS. Pour les charges de travail Windows classiques, vous pouvez également utiliser cette configuration pour tirer parti d'un environnement Active Directory et définir des autorisations sur vos partages de fichiers. Vous pouvez intégrer efficacement un Storage Gateway à votre utilisation normale, car il fonctionnera comme un partage de fichiers Windows classique. Les fichiers et les dossiers sont stockés sous forme d'objets et les listes de contrôle d'accès NTFS (ACLs) sous forme de métadonnées.

Le tableau suivant compare les coûts de 10 To de stockage avec les trois options de stockage disponibles :

- FSx pour le serveur de fichiers Windows
- Passerelle de fichiers Amazon S3
- Amazon Elastic Block Store (Amazon EBS)

Le prix du stockage de 10 To est considérablement moins élevé si vous utilisez Amazon S3, car vous pouvez partitionner vos données en différents niveaux d'utilisation. Dans l'estimation des prix, S3 Intelligent Tiering est utilisé pour sa flexibilité tarifaire. Cela inclut 80 % dans S3 Standard, 10 % dans Infrequent Access et 10 % dans Amazon Glacier. Bien que vous puissiez utiliser Amazon Glacier, il est important de définir les règles de cycle de vie appropriées pour vous assurer qu'il n'est pas nécessaire d'accéder immédiatement aux fichiers déplacés vers Amazon Glacier. Amazon Glacier est uniquement destiné à un usage archivistique, et non à un accès régulier.

Systèmes de stockage	Coût de 10 To de stockage	Région
FSx pour Windows File Server (en supposant une économie de 50 % sur la déduplication)	SSD DE 683,20\$ USD	USA Est (Virginie du Nord)
Passerelle de fichiers Amazon S3	449,51\$ USD Hiérarchisation intelligente	USA Est (Virginie du Nord)
Amazon EBS	1,335,69\$ USD GP3	USA Est (Virginie du Nord)

Éléments à prendre en compte :

- Dans Amazon Glacier, vous recevez I/O des erreurs génériques, sauf si vous utilisez l'[RestoreObject](#) API pour restaurer l'objet sur Amazon S3. Nous vous recommandons d'utiliser une notification pour cette I/O erreur via Amazon CloudWatch Events. Ainsi, votre équipe des opérations peut réagir si un utilisateur reçoit cette erreur sur un fichier auquel il pourrait avoir besoin d'accéder. Pour plus d'informations sur ces erreurs, consultez [Error : InaccessibleStorageClass](#) dans la documentation Amazon S3 File Gateway.
- Outre la limitation d'accès imposée par Amazon Glacier, [seuls 10 accès sont ACLs autorisés par objet/dossier](#) sur Storage Gateway. Avant de décider d'utiliser Storage Gateway, assurez-vous que vous n'avez pas besoin de plus de 10 entrées ACL.

Passerelle FSx de fichiers Amazon

Semblable à une passerelle de fichiers Amazon S3, une passerelle de FSx fichiers permet d'accéder à un système de fichiers qui conserve les données à long terme. Dans Amazon S3 File Gateway, les données se trouvent dans Amazon S3. Pour FSx File Gateway, vos données résident sur le serveur FSx de fichiers Windows. Bien que des options multi-AZ soient disponibles FSx pour Windows File Server, il n'existe pas d'option multi-région. Si vous avez une entreprise internationale ou un bureau distant, vous devrez peut-être fournir une plateforme de stockage partagée géographiquement plus proche de l'utilisateur final afin d'éviter les temps de latence. Si vous deviez déployer un autre système de FSx fichiers Amazon, cela augmenterait le coût d'un tout nouveau système de fichiers Amazon FSx pour Windows File Server et du stockage nécessaire. Pour éviter de créer un système de fichiers entièrement nouveau et de dupliquer les coûts, vous pouvez déployer FSx File Gateway dans la région secondaire. Cela permet aux utilisateurs d'accéder aux fichiers de manière localisée, tout en réduisant vos coûts globaux.

Systèmes de stockage	Coût de 10 To de stockage	Région
Serveur FSx de fichiers Amazon pour Windows	683,20\$ USD SSD	USA Est (Virginie du Nord)
Passerelle FSx de fichiers Amazon	503,70 \$/Passerelle unique	USA Est (Virginie du Nord)

Note

Les prix indiqués dans le tableau précédent sont basés sur les [tarifs de Storage Gateway](#).

Gardez à l'esprit les points suivants :

- FSx File Gateway peut vous aider à économiser environ 180\$ par mois (ou 2 100\$ par an) pour les charges de travail multirégionales.
- Les frais de transfert de données sont beaucoup moins élevés avec FSx File Gateway, car il suffit de mettre en cache les fichiers auxquels vous accédez régulièrement et non une copie secondaire complète.

- Bien que vous puissiez effectuer deux déploiements de FSx for Windows File Server dans différentes régions et les maintenir à jour avec AWS Backup ou AWS DataSync, aucune des deux options n'est quasiment en temps réel.

Recommandations en matière d'optimisation des coûts

Passerelle de fichiers Amazon S3

S3 File Gateway fournit une option peu coûteuse pour le stockage de fichiers, mais certains problèmes doivent être pris en compte concernant la manière dont il implémente et utilise le système de fichiers. Par exemple, S3 File Gateway nécessite l'utilisation d'une machine virtuelle pour exécuter le logiciel Storage Gateway. Dans AWS, Storage Gateway est déployé dans Amazon EC2 à l'aide d'une instance m5.xlarge, par défaut. Si vous souhaitez réduire vos coûts de stockage sur site, vous pouvez déployer Storage Gateway en tant qu'appliance virtuelle sur des plateformes de virtualisation telles que VMware Hyper-V.

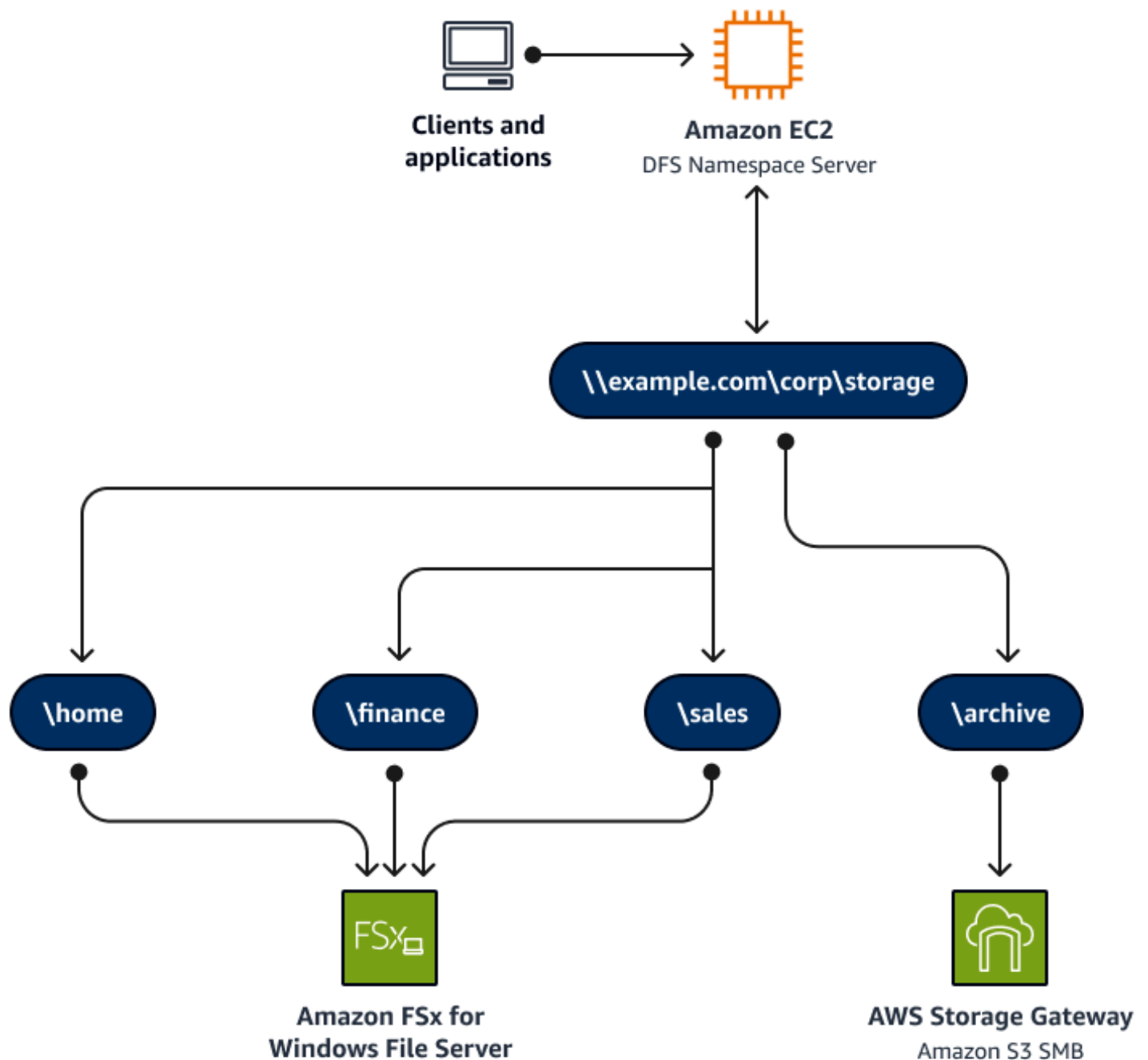
Considérations sur la haute disponibilité

L'exécution de Storage Gateway constitue un point de défaillance unique pour l'accès aux fichiers. Pour éviter les interruptions inutiles, nous vous recommandons de mettre en œuvre un contrôle d'accès strict permettant aux utilisateurs d'apporter des modifications ou d'arrêter et de démarrer l'instance Storage Gateway. En outre, pour les déploiements sur AWS, il est avantageux d'utiliser Amazon Data Lifecycle Manager pour créer des instantanés de routage afin de récupérer rapidement votre implémentation de Storage Gateway. Si vous exécutez Storage Gateway sur site à l'aide de VMware, vous pouvez le configurer pour [une haute disponibilité](#).

Exécution de plusieurs systèmes de fichiers

Séparer vos charges de travail de fichiers quotidiennes de vos charges de travail d'archivage peut vous aider à éviter des coûts de stockage inutiles. Storage Gateway peut être déployé parallèlement à un système de fichiers FSx pour Windows File Server. En utilisant les [espaces de noms DFS](#), vous pouvez présenter votre stockage principal à usage quotidien exécuté sur FSx Windows File Server et votre stockage exécuté sur Amazon S3 (accessible via Storage Gateway).

Le schéma suivant montre comment un seul espace de noms DFS peut être utilisé comme point d'accès frontal pour différentes options de stockage principal.



Les clients sont dirigés vers une structure de dossiers, telle que `\\example.com\storage`. Ce répertoire principal contient les sous-répertoires. Un système de fichiers FSx pour Windows File Server contient les partages de fichiers auxquels on accède normalement. Vous pouvez utiliser un partage de fichiers créé sur Storage Gateway pour archiver les données. Les utilisateurs peuvent archiver manuellement des éléments dans le dossier d'archive ou vous pouvez créer un processus

pour automatiser le déplacement de certains fichiers de vos partages de fichiers habituels vers le dossier d'archive.

Éléments à prendre en compte :

- Passez en revue vos besoins en stockage et fournissez un [espace de stockage adéquat pour le cache](#).
- Ajoutez votre passerelle à votre configuration Active Directory et utilisez [Windows standard ACLs pour accéder aux fichiers](#).

FSx Passerelle de fichiers

Le déploiement de FSx File Gateway est similaire au déploiement de S3 File Gateway, mais il est encore plus facile si vous utilisez l'assistant de lancement. Pour obtenir des instructions détaillées, consultez [Étape 3 : Création et activation d'Amazon FSx File Gateway](#) dans la documentation Amazon FSx File Gateway. Après avoir déployé FSx File Gateway dans votre environnement, vous pouvez l'associer à vos systèmes de FSx fichiers Amazon existants et accéder à vos fichiers.

Le stockage est la principale considération lors FSx du déploiement de File Gateway. Le stockage par défaut fournit 150 Go, ce qui est un espace décent pour la mise en cache des fichiers. La création d'alertes de surveillance en cas de manque d'espace libre peut aider à dimensionner correctement le stockage sans surallocation.

Ressources supplémentaires

- [AWS Storage Gateway ressources](#) (AWS documentation)

Active Directory

Amazon Elastic Compute Cloud (Amazon EC2) exécutant Windows Server est un environnement sécurisé, fiable et performant pour le déploiement d'applications et de charges de travail basées sur Windows. Vous pouvez approvisionner des instances rapidement et les augmenter ou les réduire selon vos besoins, tout en ne payant que pour ce que vous utilisez. Les services Active Directory sont utilisés comme source principale de gestion des identités dans les environnements Windows Server.

Cette section couvre les rubriques suivantes :

- [Active Directory autogéré sur Amazon EC2](#)
- [AWS Managed Microsoft AD](#)
- [AD Connector](#)

Active Directory autogéré sur Amazon EC2

Présentation de

Cette section fournit des recommandations pour réduire le coût d'exécution d'Active Directory sur Amazon Elastic Compute Cloud (Amazon EC2). L'objectif principal est de vous assurer que vous pouvez dimensionner les contrôleurs de domaine Active Directory de manière appropriée et utiliser la flexibilité qu'ils offrent AWS Cloud pour les ajuster en fonction des besoins de votre environnement. AWS peut vous aider à arrêter facilement une instance et à la redimensionner pour répondre à l'évolution de vos besoins, ou à réduire la taille de l'instance si vous augmentez trop rapidement. Le choix de la bonne taille et du bon type d'instance peut permettre de réaliser des économies importantes.

Impact sur les coûts

Le tableau suivant montre la différence entre le choix d'une instance de famille d'instances burstable par rapport à une instance à usage général. Ce choix peut vous faire économiser une somme d'argent considérable chaque mois. Une planification et un dimensionnement appropriés de votre instance peuvent vous aider à gérer les coûts.

Type d'instance	Nombre d'instances	vCPU	Mémoire	Cost
t3a.medium	2	2	8	81,76 \$/mois
m5a.large	2	2	8	259,88 \$/mois

Pour plus d'informations sur les coûts, consultez l' [Calculateur de tarification AWS estimation](#).

Une économie de 178,12\$ par mois représente une économie de plus de 2 000\$ par an pour vos contrôleurs de domaine. N'oubliez pas qu'il s'agit d'un faible encombrement de deux contrôleurs de domaine dans un seul compte. À grande échelle, avec plusieurs comptes et des contrôleurs de domaine supplémentaires, ces économies peuvent se traduire par une réduction significative des coûts.

Recommandations en matière d'optimisation des coûts

Microsoft fournit des [recommandations de planification des capacités](#) lorsque vous déployez votre environnement Active Directory. Nous vous recommandons de prendre en compte les principaux composants suivants lors de la planification ou de la mise à l'échelle de votre environnement Active Directory :

- Mémoire
- Réseau
- Stockage
- Processeur

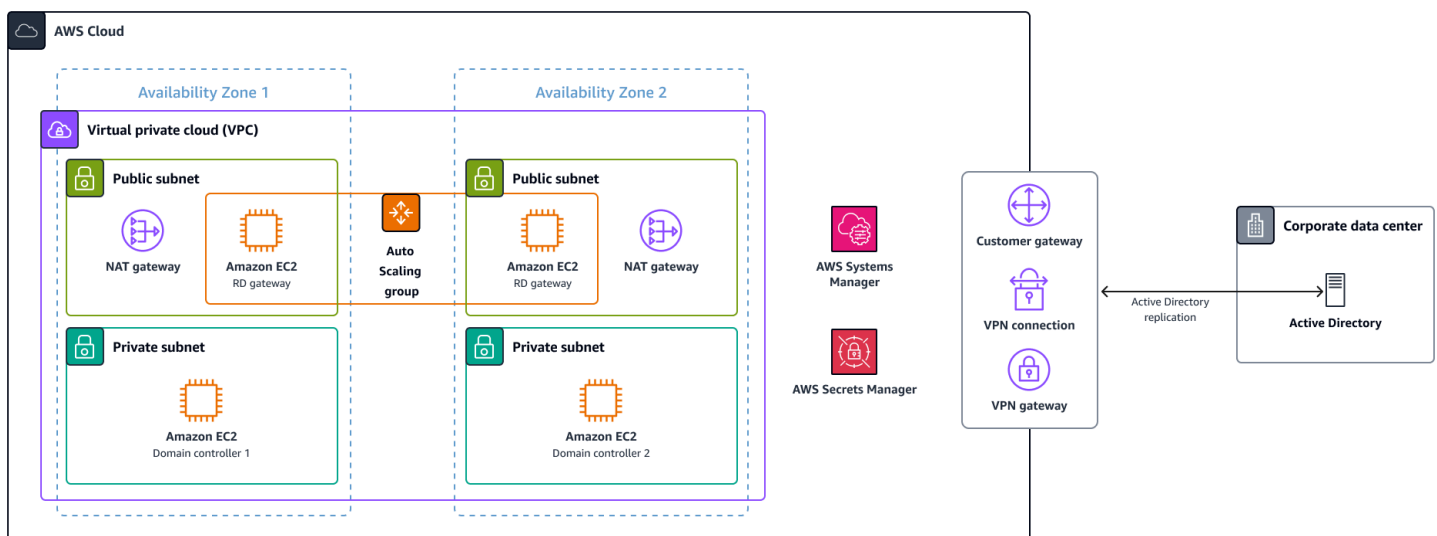
Tout en gardant ces principaux composants à l'esprit, vous pouvez sélectionner un type d'instance adapté à votre environnement Active Directory AWS. Cette section présente quelques exemples de scénarios de AWS déploiement d'Active Directory. Ces scénarios indiquent clairement qu'il n'est pas nécessaire de répliquer votre environnement sur site si vous ne prévoyez pas de gérer le même nombre d'utilisateurs et d'ordinateurs que dans votre environnement sur site. AWS

Le tableau suivant met en évidence les composants importants concernant le vCPU, la mémoire et le disque par rapport à votre AWS encombrement.

Composant	Estimations
Taille du stockage/de la base de données	40 à 60 Ko pour chaque utilisateur
RAM	Taille de base de données Recommandations relatives aux systèmes d'exploitation de base Applications tierces
Réseau	1 Go
CPU	1 000 utilisateurs simultanés pour chaque cœur

Scénario de déploiement hybride

Le schéma suivant montre un exemple d'architecture pour un déploiement hybride d'Active Directory.



Comme le montre le diagramme, vous avez généralement une empreinte sur site, puis vous l' étendez à l' AWS Cloud. Au cours des phases initiales d'une migration, tous vos utilisateurs et serveurs ne sont généralement pas déployés sur AWS. C'est pourquoi il est important de déployer dans un premier temps un encombrement réduit afin d'économiser de l'argent sur les efforts de migration.

Si vous souhaitez conserver une empreinte sur site avec des serveurs et des utilisateurs s'authentifiant sur site, vous n'aurez pas besoin de la même empreinte pour les contrôleurs de

domaine. AWS En suivant les meilleures pratiques d'Active Directory, vous pouvez implémenter des [sites et des services Active Directory](#) appropriés pour authentifier les utilisateurs et les ordinateurs selon votre empreinte sur site, tout en authentifiant votre AWS empreinte uniquement auprès des contrôleurs de domaine concernés. AWS Cela vous permet d'éviter de surdimensionner votre empreinte Active Directory AWS en limitant l'utilisation aux seules AWS ressources et non à l'ensemble de votre infrastructure sur site. Pour obtenir des conseils sur la conception d'une configuration hybride, consultez la section [Placement correct des contrôleurs de domaine et considérations relatives au site](#) dans la documentation Microsoft.

Optimisez la AWS migration grâce à la bonne taille

Si vous déployez une nouvelle instance d'Active Directory pour vos utilisateurs ou si vous envisagez de procéder à une migration complète vers AWS votre infrastructure Active Directory, nous vous recommandons de planifier le dimensionnement en fonction des recommandations de Microsoft en matière de vCPU, de mémoire et d'espace disque pour les instances choisies dans le tableau précédent.

S'il s'agit d'un nouvel encombrement, vous pouvez commencer modestement et tirer parti de la possibilité de [modifier facilement les types d'instances](#) pour redimensionner votre environnement au fur et à mesure de son évolution. AWS La section [Windows sur Amazon EC2](#) de ce guide explique comment surveiller et examiner l'utilisation de votre processeur et de votre mémoire sur. AWS Ainsi, vous savez quand augmenter la taille de votre instance EC2.

Si vous effectuez une migration complète de votre environnement Active Directory local vers AWS, vous pouvez mettre en œuvre les mêmes plans de dimensionnement pour garantir des performances optimales. Avant de dupliquer ce que vous avez sur site AWS, nous vous recommandons de procéder à un examen approfondi de votre environnement Active Directory. Cela peut vous aider à éviter le surprovisionnement. Veillez à utiliser l'Analyseur de performances pour collecter des informations sur le volume de trafic et l'utilisation de vos contrôleurs de domaine existants. Cela peut vous permettre de comprendre l'utilisation globale afin que vous puissiez choisir la bonne taille et, en fin de compte, réduire vos coûts.

Optimisez Active Directory sur AWS

Si vous exécutez Active Directory sur Active Directory AWS, il est également important de surveiller en permanence l'utilisation et de modifier la taille des instances selon les besoins afin de réduire vos dépenses. Vous pouvez l' Optimiseur de calcul AWS utiliser pour obtenir des informations sur les ressources que vous utilisez AWS. Pour plus d'informations sur l'utilisation de Compute Optimizer pour dimensionner correctement vos charges de travail Windows, consultez la section [Windows sur](#)

[Amazon EC2](#) de ce guide. Pour une analyse plus approfondie, vous pouvez utiliser l'Analyseur de performances pour surveiller l'utilisation de vos contrôleurs de domaine Active Directory, évaluer les performances, puis redimensionner en conséquence.

Vous pouvez également l'utiliser CloudWatch pour surveiller les performances des contrôleurs de domaine. Pour optimiser vos contrôleurs de domaine (à la hausse ou à la baisse), vous pouvez utiliser les métriques disponibles dans CloudWatch pour vous aider à prendre les bonnes décisions. Vous pouvez utiliser l' CloudWatchagent pour configurer les métriques personnalisées de Performance Monitor à envoyer pour la collecte de données. Pour obtenir des instructions, voir [Comment utiliser l' CloudWatch agent pour consulter les mesures de Performance Monitor sur un serveur Windows ?](#) dans le AWS Knowledge Center.

Après avoir déployé l' CloudWatch agent, vous pouvez configurer les métriques suivantes dans le fichier de configuration de l'agent situé sous `metrics_collected` :

Catégorie de la métrique	Nom de la métrique
De la base de données aux instances (NTDSA)	% de visites du cache de base de données
La base de données d'E/S lit la latence moyenne	
I/O database reads/sec	
Le journal des E/S écrit la latence moyenne	
DirectoryServices (MNT)	Heure de liaison LDAP
DRA en attente d'opérations de réplication	
DRA en attente de synchronisations de réplication	
DNS	Requêtes récursives/sec
Échec d'une requête récursive/sec	
Requête TCP reçue/sec	
Nombre total de requêtes reçues/sec	

Catégorie de la métrique	Nom de la métrique
Réponse totale envoyée/sec	
Requête UDP reçue/sec	
LogicalDisk	Longueur moyenne de la file d'attente du disque
% d'espace libre	
Mémoire	% d'octets engagés en cours d'utilisation
Durée de vie moyenne du cache de veille à long terme	
Interface réseau	Octets envoyés/s
Octets reçus/s	
Bande passante actuelle	
NTDS	Délai d'attente estimé par ATQ
Latence des requêtes ATQ	
Le répertoire DS est lu par seconde	
Recherches dans le répertoire DS/sec	
Le répertoire DS écrit par seconde	
Sessions client LDAP	
Recherches LDAP/sec	
Liaisons LDAP réussies/sec	
Processeur	% de temps du processeur
Statistiques de sécurité à l'échelle du système	Authentifications Kerberos

Catégorie de la métrique	Nom de la métrique
Authentifications NTLM	

Ressources supplémentaires

- [Services de domaine Active Directory sur AWS : Guide de déploiement de la solution destinée aux partenaires](#) (AWS documentation)
- [Planification des capacités pour les services de domaine Active Directory](#) (documentation Microsoft)
- [Considérations relatives à la conception relatives à l'exécution d'Active Directory sur des instances EC2](#) (livres AWS blancs)

AWS Managed Microsoft AD

Présentation de

AWS Directory Service for Microsoft Active Directory, également connu sous le nom de AWS Managed Microsoft AD, est alimenté par un Windows Server Active Directory et géré par AWS. Vous pouvez l'utiliser AWS Managed Microsoft AD pour migrer un large éventail d'applications compatibles avec Active Directory vers le. AWS Cloud AWS Managed Microsoft AD fonctionne avec une variété d'applications et de services Active Directory natifs. Il prend également en charge les [applications et les services AWS gérés](#). Bien qu'il n'existe pas beaucoup de leviers d'optimisation des coûts AWS Managed Microsoft AD en raison du service et de ses mécanismes de facturation, certains principes de conception peuvent vous aider à réduire les coûts au minimum.

Impact sur les coûts

Comme AWS Managed Microsoft AD il s'agit d'un service géré basé sur le présent SKUs, le dimensionnement est un processus relativement simple. Deux tailles sont actuellement SKUs disponibles : les éditions Standard et Enterprise. SKUs Les autres incluent le partage de répertoires, l'ajout de contrôleurs de domaine supplémentaires (y compris des régions supplémentaires) et le transfert de données entre régions.

Recommandations en matière d'optimisation des coûts

Il existe des différences entre l'édition AWS Managed Microsoft AD Standard et AWS Managed Microsoft AD l'édition Enterprise. L'édition Enterprise prend en charge jusqu'à 500 000 objets Active Directory, 500 partages de comptes (limite souple) et prend en charge plusieurs régions. L'édition Standard prend en charge jusqu'à 30 000 objets Active Directory, cinq partages de comptes (limite souple d'environ 25 au maximum) et ne prend pas en charge plusieurs régions.

Note

Les limites supérieures des objets Active Directory sont approximatives. Votre répertoire peut prendre en charge plus ou moins d'objets en fonction de leur taille, du comportement et des besoins en performances de vos applications.

Les questions à prendre en compte avant de sélectionner votre type de répertoire sont les suivantes :

- Une assistance multirégionale est-elle requise ?
- Le répertoire sera-t-il partagé avec plus de 25 comptes ?
- Le nombre d'objets Active Directory sera-t-il supérieur à 30 000 ?

Si la réponse est « oui » à l'une des questions ci-dessus, l'édition Enterprise est requise. Si la réponse à toutes les questions est non, nous vous recommandons de commencer par l'édition Standard.

Note

Vous pouvez mettre à niveau un répertoire de l'édition Standard vers l'édition Enterprise, mais un répertoire ne peut pas être rétrogradé. Le déploiement de l'édition standard ne se fait pas à sens unique. Si vous souhaitez mettre à jour votre annuaire vers l'édition Enterprise, contactez AWS.

Chaque partage entraîne un coût lorsque vous partagez des annuaires dans AWS Managed Microsoft AD Enterprise Edition. C'est moins que le coût du déploiement d'un annuaire dans chaque compte, mais n'oubliez pas que les coûts de partage peuvent augmenter s'ils ne sont pas contrôlés. Nous vous recommandons de ne partager des annuaires qu'avec des comptes contenant Amazon

Relational Database Service (Amazon RDS) et FSx Amazon pour Windows File Server, car seuls ces services prennent en charge cette fonctionnalité. N'oubliez pas que vous avez la possibilité d'intégrer le serveur FSx de fichiers Windows à votre Active Directory autogéré, y compris un AWS Managed Microsoft AD. Si seul Amazon FSx est requis pour un autre compte, vous pouvez effectuer un FSx déploiement Amazon autogéré par rapport au AWS Managed Microsoft AD sans avoir à partager le répertoire.

Lorsque vous décidez quand déployer des contrôleurs de domaine supplémentaires, gardez à l'esprit que seuls deux sous-réseaux sont pris AWS Managed Microsoft AD en charge dans des zones de disponibilité distinctes au sein d'un même VPC. L'ajout de contrôleurs de domaine supplémentaires ne vous permet pas d'ajouter des sous-réseaux supplémentaires. Pour déterminer si vous devez ajouter des contrôleurs de domaine supplémentaires en raison de problèmes de performances, consultez les [mesures de performance des contrôleurs de domaine dans CloudWatch](#). Cela vous indique si un ou tous les contrôleurs de domaine sont débordés. Si vous constatez qu'un seul contrôleur de domaine est surchargé, l'ajout de contrôleurs de domaine supplémentaires n'allégera pas la charge et vous devrez étudier plus en profondeur les applications et ne pas équilibrer la charge entre les contrôleurs de domaine actuellement disponibles. Si tous les contrôleurs de domaine sont fortement utilisés, l'ajout d'un contrôleur de domaine supplémentaire peut réduire la charge sur les contrôleurs de domaine existants. Pour obtenir des instructions sur la façon d'automatiser le dimensionnement, voir [Comment automatiser le AWS Managed Microsoft AD dimensionnement en fonction des indicateurs d'utilisation](#) dans le blog sur la AWS sécurité.

Si vous avez étendu votre répertoire à plusieurs régions, nous vous recommandons de ne pas utiliser les répertoires partagés NETLOGON ou SYSVOL pour le stockage de fichiers. Tous les contrôleurs de domaine répliquent le contenu de ces partages. Le fait de ne pas utiliser les partages pour le stockage de fichiers réduit les coûts de transfert de données au minimum.

Vous avez également la possibilité de souscrire un contrat d'entreprise avec AWS. Les contrats d'entreprise vous permettent de personnaliser les accords qui répondent le mieux à vos besoins. Pour plus d'informations, consultez la section [Clients d'entreprise](#).

Ressources supplémentaires

- [AWS Managed Microsoft AD quotas](#) (AWS Directory Service documentation)
- [AWS Directory Service Tarification](#) (AWS site Web)
- [Services de domaine Active Directory activés AWS](#) (livres AWS blancs)

AD Connector

Présentation de

[AD Connector](#) est un service proxy qui permet de connecter facilement votre Microsoft Active Directory sur site à des [AWS applications](#) compatibles, telles qu'Amazon, Amazon Quick WorkSpaces, et de joindre facilement des domaines pour les instances Amazon Elastic Compute Cloud (Amazon EC2), sans mettre en cache aucune information dans le cloud. Vous pouvez utiliser AD Connector pour ajouter un compte de service à votre Active Directory. AD Connector élimine le besoin de synchronisation d'annuaires ou le coût et la complexité liés à l'hébergement d'une infrastructure de fédération. Même s'il existe peu de leviers d'optimisation des coûts pour AD Connector en raison de la nature du service et de ses mécanismes de facturation, vous pouvez suivre les recommandations de conception de cette section pour réduire les coûts au minimum.

Impact sur les coûts

AD Connector est un service géré basé sur un préréglage SKUs. Cela fait du dimensionnement un processus simple. Deux tailles sont SKUs disponibles : petite et grande taille. Vous pouvez utiliser le [Calculateur de tarification AWS](#) pour les estimations de coûts impliquant AD Connector.

Recommandations en matière d'optimisation des coûts

Hormis les ressources de calcul du backend, il n'y a aucune différence entre les petites et les grandes tailles de connecteurs.

Les questions à prendre en compte avant de sélectionner votre type de répertoire sont les suivantes :

- Y a-t-il un grand nombre (plus de 10 000) d'utilisateurs actifs utilisant des AWS applications intégrées à l'AD Connector ?
- L'utilisateur est-il membre de nombreux groupes imbriqués, profonds ou circulaires ?

Si la réponse à ces deux questions est négative, nous vous recommandons de commencer par le petit format. Si vous répondez oui à l'une des questions ci-dessus, une taille plus importante pourrait être envisagée. Vous pouvez commencer avec un AD Connector de petite taille et, si le répertoire est altéré en raison des performances, vous pouvez demander qu'il soit mis à niveau vers le répertoire de grande taille.

Note

Vous pouvez mettre à niveau un AD Connector de petite taille vers un AD Connector de grande taille, mais un AD Connector ne peut pas être rétrogradé.

La plupart des problèmes de performances ne sont pas liés à l'AD Connector, mais les contrôleurs de domaine Active Directory locaux sont débordés par le fait que de nombreux utilisateurs sont membres de nombreux groupes imbriqués, profonds ou circulaires.

Vous avez également la possibilité de souscrire un contrat d'entreprise avec AWS. Les contrats d'entreprise vous permettent de personnaliser les accords qui répondent le mieux à vos besoins. Pour plus d'informations, consultez la section [Clients d'entreprise](#).

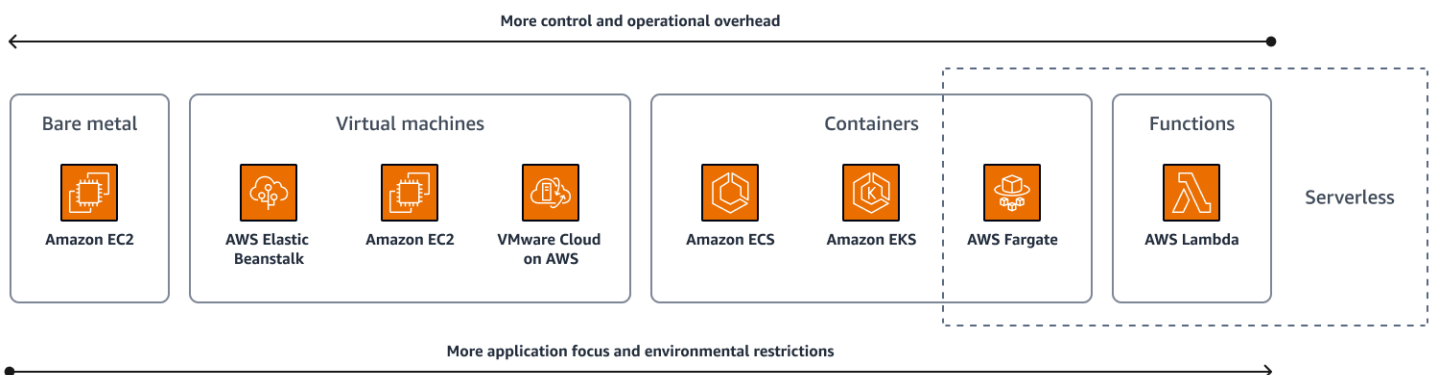
Ressources supplémentaires

- [Quotas AD Connector](#) (AWS Directory Service documentation)
- [Tarification des autres types d'annuaires](#) (AWS site Web)
- [Services de domaine Active Directory activés AWS](#) (livres AWS blancs)

.NET

Le développement et le déploiement d'applications .NET sont essentiels pour vous aider à atteindre l'évolutivité et l'agilité offertes par le cloud computing. Pour de nombreuses applications .NET existantes, le choix de calcul le plus approprié pour exécuter des applications AWS consiste à utiliser des machines virtuelles, soit via Amazon Elastic Compute Cloud (Amazon EC2), AWS Elastic Beanstalk soit via Amazon Elastic Compute Cloud (Amazon EC2). Il est également possible d'exécuter des applications .NET dans des conteneurs Windows et Linux.

L'introduction de .NET Core vous permet de concevoir des applications .NET modernes qui tirent parti de tous les avantages du cloud. Les applications modernes peuvent utiliser l'ensemble traditionnel de choix de calcul et également cibler différents types d'environnements sans serveur, notamment AWS Fargate ou AWS Lambda. .NET 6+ offre désormais un hébergement performant des charges de travail sur des instances ARM64 EC2 telles que les familles Graviton2 EC2. Cela permet d'accéder à la dernière génération de processeurs disponibles sur Amazon EC2. Cela signifie que vos applications peuvent être hébergées sur des ordinateurs spécialisés adaptés à votre type de charge de travail, tels que l'encodage vidéo, les serveurs Web et le calcul haute performance (HPC).



Cette section fournit des recommandations pour vous aider à adapter vos applications .NET afin de tirer parti des avantages du cloud en mettant l'accent sur la rentabilité.

Cette section couvre les rubriques suivantes :

- [Passez au .NET moderne et passez à Linux](#)
- [Conteneuriser les applications .NET](#)
- [Utiliser des instances et des conteneurs Graviton](#)
- [Support du dimensionnement dynamique pour les applications .NET Framework statiques](#)

- [Utiliser la mise en cache pour réduire la demande de base de données](#)
- [Envisagez le .NET sans serveur](#)
- [Envisagez des bases de données spécialement conçues](#)

Passez au .NET moderne et passez à Linux

Présentation de

La modernisation des anciennes applications .NET Framework peut vous aider à améliorer la sécurité, les performances et l'évolutivité. Un moyen efficace de moderniser les applications .NET Framework consiste à les migrer vers une version .NET moderne (6+). Voici quelques-uns des principaux avantages de la migration de ces applications vers le .NET open source :

- Pour réduire les coûts de licence Windows en les exécutant sur un système d'exploitation Linux
- Profitez de la disponibilité des langues modernes
- Bénéficiez de performances optimisées pour fonctionner sous Linux

De nombreuses entreprises utilisent toujours d'anciennes versions du .NET Framework. Cela peut présenter des risques de sécurité, car les vulnérabilités des anciennes versions ne sont plus corrigées par Microsoft. Microsoft a mis fin au support des versions récentes de .NET Framework 4.5.2, 4.6 et 4.6.1. Il est très important d'évaluer les risques et les avantages liés à la poursuite de l'exécution des anciennes versions du framework. Pour réduire les risques et les coûts, il peut être intéressant d'investir du temps et des efforts dans le refactoring vers une version moderne de .NET.

Impact sur les coûts

Envisagez un type d'instance EC2 à usage général (m5), qui offre un équilibre entre les ressources de calcul, de mémoire et de réseau. Ces instances conviennent à diverses applications telles que les serveurs Web, les bases de données de taille moyenne et les référentiels de code source.

Par exemple, une instance m5.xlarge à la demande avec 4 V CPUs et 16 Go de mémoire sur Windows Server (licence incluse) dans l'est des États-Unis (Virginie du Nord) coûte 274,48\$ par mois. Les mêmes ressources sur un serveur Linux coûtent 140,16\$ par mois. Dans cet exemple, les coûts sont réduits de 49 % lorsque vous migrez votre application de .NET Framework vers une version moderne de .NET et que vous exécutez votre application sur un serveur Linux. Votre coût peut varier en fonction des options (par exemple, type d'instance, système d'exploitation, stockage)

que vous choisissiez lors de la sélection d'une [instance EC2](#). Vous pouvez optimiser davantage les coûts en utilisant [Savings Plans](#) ou [Reserved Instances](#). Pour plus de détails, utilisez le [Calculateur de tarification AWS](#) pour établir des estimations de coûts. Pour les instances incluses dans Windows, le coût de licence est de [0,046 USD par vCPU par](#) heure, quel que soit le modèle de tarification.

Le portage de ces applications .NET Framework vers le .NET moderne nécessite des efforts de développement. Vous devez évaluer vos applications et leurs dépendances pour voir si elles sont compatibles avec la version de la plateforme cible. [AWS L'assistant de portage pour .NET](#) est un outil d'assistance qui analyse les applications .NET Framework et génère une évaluation de compatibilité .NET, vous aidant ainsi à porter vos applications pour qu'elles soient compatibles avec Linux plus rapidement. L'assistant de portage pour .NET identifie les incompatibilités avec .NET, trouve les remplacements connus et génère une évaluation détaillée de la compatibilité. Après avoir porté votre solution, vous devez modifier le code manuellement pour que votre projet soit correctement compilé avec les dépendances. Cela réduit les efforts manuels nécessaires à la modernisation de vos applications vers Linux. Si votre application prend en charge les processeurs ARM, le passage à Linux permet d'utiliser des instances Graviton. Cela peut vous aider à réaliser des réductions de coûts supplémentaires de 20 %. Pour plus d'informations, consultez la section [Powering .NET 5 with AWS Graviton2 : Benchmarks](#) dans le blog AWS Compute.

Il existe d'autres outils, tels que [AWS Toolkit for .NET Refactoring](#) et [.NET Upgrade Assistant](#), qui peuvent vous aider à porter les anciennes applications .NET Framework vers le .NET moderne.

Recommandations en matière d'optimisation des coûts

Pour migrer des applications .NET Framework, procédez comme suit :

1. Conditions préalables — Pour utiliser l'assistant de portage pour .NET, vous devez installer .NET 5+ sur la machine sur laquelle vous prévoyez d'analyser le code source de l'application. Les ressources de la machine doivent avoir une vitesse de GHz traitement minimale de 1,8, 4 Go de mémoire et 5 Go d'espace de stockage. Pour plus d'informations, consultez la section [Conditions requises](#) dans la documentation de l'assistant de portage pour .NET.
2. Évaluation — Téléchargez l'assistant de portage pour .NET sous forme de fichier [exécutable](#) (téléchargement). Vous pouvez télécharger et installer l'outil sur votre ordinateur pour démarrer l'évaluation de vos applications. La page d'évaluation contient des projets portés, des packages APIs incompatibles avec le .NET moderne. Pour cette raison, des erreurs de construction apparaissent dans la solution après l'évaluation. Vous pouvez consulter ou télécharger les résultats de l'évaluation dans un fichier CSV. Pour plus d'informations, consultez [Porter une solution](#) dans la documentation de l'assistant de portage pour .NET.

3. Refactorisation — Après avoir évalué l'application, vous pouvez porter vos projets vers la version cible du framework. Lorsque vous portez une solution, vos fichiers de projet et une partie du code sont modifiés par l'assistant de portage. Vous pouvez consulter les journaux pour vérifier les modifications apportées à votre code source. Dans la plupart des cas, le code nécessitera des efforts supplémentaires pour terminer la migration et les tests afin de le préparer à la production. Selon l'application, certaines des modifications peuvent inclure le cadre d'entité, l'identité et l'authentification. Pour plus d'informations, consultez [Porter une solution](#) dans la documentation de l'assistant de portage pour .NET.

Il s'agit d'une première étape pour moderniser vos applications en conteneurs. La modernisation de vos applications .NET Framework vers des conteneurs Linux peut nécessiter un certain nombre de facteurs commerciaux et techniques. L'un des principaux moteurs est la réduction du coût total de possession en abandonnant le système d'exploitation Windows au profit de Linux. Cela réduit les coûts de licence lors de la migration de votre application vers une version multiplateforme de .NET et vers des conteneurs afin d'optimiser l'utilisation des ressources.

Une fois votre application portée sur Linux, vous pouvez l'utiliser [AWS App2Container](#) pour conteneuriser votre application. App2Container utilise Amazon ECS ou Amazon EKS comme services de point de terminaison sur lesquels vous pouvez les déployer directement. App2Container fournit tous les artefacts de déploiement de l'infrastructure sous forme de code (IaC) nécessaires pour conteneuriser vos applications à plusieurs reprises.

Considérations et ressources supplémentaires

- Si vous possédez des applications basées sur VB.NET (un framework existant datant de 2002) et que vous souhaitez les porter vers .NET 6, consultez le billet [Port des anciennes applications VB.NET vers .NET 6.0 avec l'assistant de portage pour .NET](#) sur le blog Microsoft Workloads on. AWS
- Si vous possédez des applications existantes sur Windows Communication Foundation (WCF) et que vous souhaitez les exécuter sur un .NET moderne, vous pouvez adopter CoreWCF. Pour plus d'informations, consultez le billet sur la [modernisation des anciennes applications WCF vers CoreWCF à l'aide de l'assistant de portage pour .NET](#) sur le blog Microsoft Workloads on. AWS
- Vous pouvez ajouter un assistant de portage en tant qu'extension à votre IDE Visual Studio. Cela vous permet d'effectuer toutes les tâches nécessaires à la conversion de votre code sans avoir à basculer entre votre IDE et l'outil Porting Assistant for .NET. Pour plus d'informations, consultez le billet [Accélérer la modernisation des applications .NET avec l'assistant de portage pour l'extension IDE .NET Visual Studio](#) sur le AWS blog Microsoft Workloads on.

- [AWS L'assistant de portage pour .NET est désormais un outil open source](#) avec le code source et les composants d'analyse de compatibilité de l'évaluation. Cela peut encourager vos développeurs à utiliser et à partager les connaissances et les meilleures pratiques en matière de portage .NET.
- Vous pouvez porter des applications .NET Framework vers le .NET moderne sous Linux à l'aide du AWS Toolkit for .NET Refactoring. Pour plus d'informations, consultez le billet [Accélérer la modernisation de .NET avec AWS Toolkit for .NET Refactoring](#) sur le blog Microsoft Workloads on AWS
- Vous pouvez [accélérer la conteneurisation et la migration des applications ASP.NET Core](#) vers l'utilisation. AWS App2Container

Conteneuriser les applications .NET

Présentation de

Les conteneurs constituent un moyen léger et efficace d'empaqueter et de déployer des applications de manière cohérente et reproductible. Cette section explique comment utiliser AWS Fargate un service de conteneur sans serveur pour réduire les coûts de vos applications .NET tout en fournissant une infrastructure évolutive et fiable.

Impact sur les coûts

Parmi les facteurs qui influencent l'efficacité de l'utilisation des conteneurs pour réduire les coûts, citons la taille et la complexité de l'application, le nombre d'applications à déployer, ainsi que le niveau de trafic et de demande sur les applications. Pour les applications petites ou simples, les conteneurs peuvent ne pas permettre de réaliser des économies significatives par rapport aux approches d'infrastructure traditionnelles, car les frais généraux liés à la gestion des conteneurs et des services associés peuvent en fait augmenter les coûts. Toutefois, pour les applications plus importantes ou plus complexes, l'utilisation de conteneurs peut permettre de réaliser des économies en améliorant l'utilisation des ressources et en réduisant le nombre d'instances requises.

Nous vous recommandons de garder à l'esprit les points suivants lorsque vous utilisez des conteneurs afin de réaliser des économies :

- Taille et complexité des applications — Les applications plus volumineuses et plus complexes sont mieux adaptées à la conteneurisation, car elles ont tendance à nécessiter davantage de ressources et peuvent bénéficier davantage d'une meilleure utilisation des ressources.

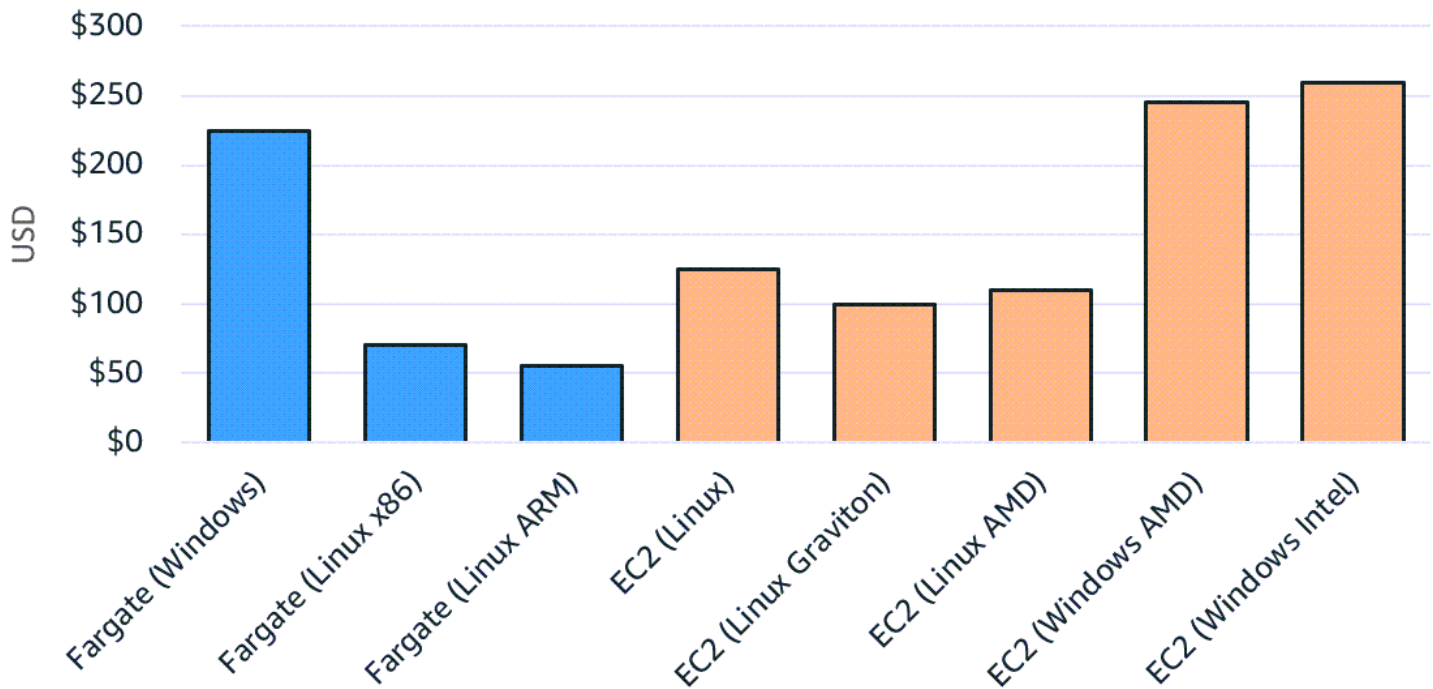
- Nombre d'applications : plus votre entreprise doit déployer d'applications, plus la conteneurisation permet de réaliser des économies.
- Trafic et demande : les applications soumises à un trafic et à une demande élevés peuvent bénéficier de l'évolutivité et de l'élasticité fournies par les conteneurs. Cela peut permettre de réaliser des économies.

Les différentes architectures et systèmes d'exploitation influent sur les coûts des conteneurs. Si vous utilisez des conteneurs Windows, il est possible que les coûts ne diminuent pas pour des raisons de licence. Les coûts de licence sont inférieurs ou absents avec les conteneurs Linux. Le tableau suivant utilise une configuration de base AWS Fargate dans la région USA Est (Ohio) avec les paramètres suivants : 30 tâches par mois exécutées chacune pendant 12 heures avec 4 V CPUs et 8 Go de mémoire alloués.

Vous pouvez choisir entre deux plateformes de calcul principales sur lesquelles exécuter vos conteneurs AWS : les [hôtes de conteneurs basés sur EC2 et les plateformes sans serveur ou. AWS Fargate](#) Si vous utilisez Amazon Elastic Container Service (Amazon ECS) au lieu de Fargate, vous devez continuer à exécuter le calcul (instances) pour permettre au moteur de placement d'instancier les conteneurs en cas de besoin. Si vous utilisez Fargate à la place, seule la capacité de calcul nécessaire est provisionnée.

Le graphique suivant montre la différence entre les conteneurs équivalents utilisant Fargate et Amazon EC2. Grâce à la flexibilité de Fargate, les tâches d'une application peuvent être exécutées 12 heures par jour, sans aucune utilisation en dehors des heures de bureau. Toutefois, pour Amazon ECS, vous devez contrôler la capacité de calcul à l'aide d'un [groupe Auto Scaling](#) d'instances EC2. Cela peut conduire à une capacité opérationnelle 24 heures sur 24, ce qui peut en fin de compte augmenter les coûts.

Monthly costs of Fargate and Amazon EC2



Recommandations en matière d'optimisation des coûts

Utiliser des conteneurs Linux plutôt que Windows

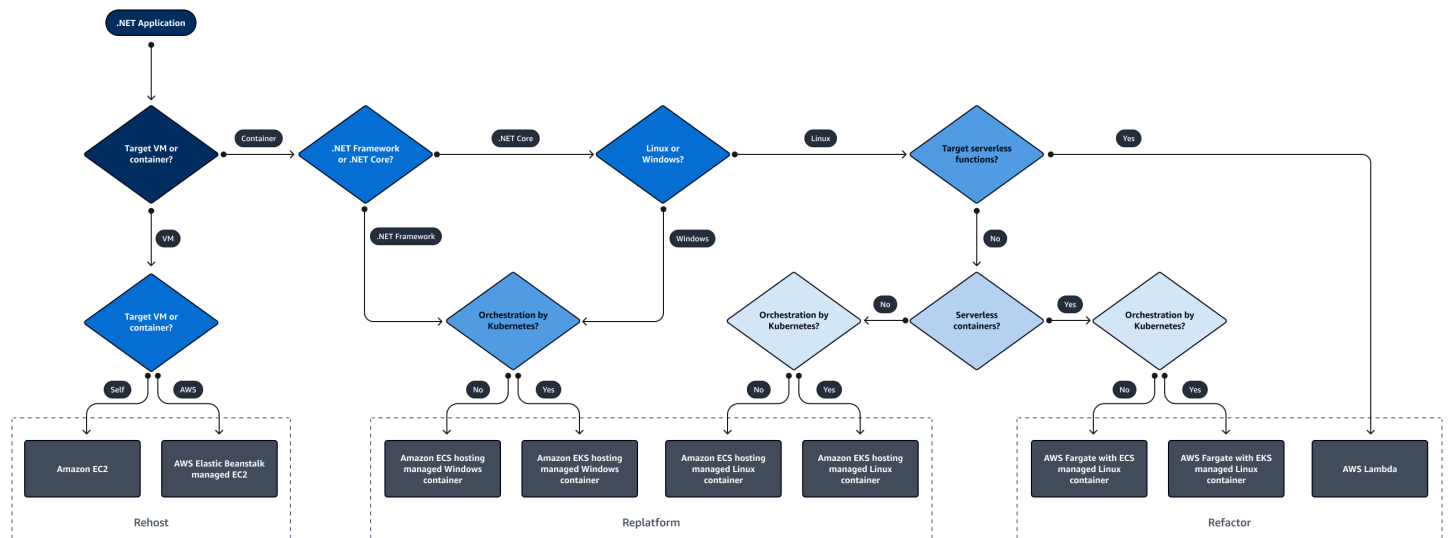
Vous pouvez réaliser des économies importantes si vous utilisez des conteneurs Linux au lieu de conteneurs Windows. Par exemple, vous pouvez réaliser des économies d'environ 45 % sur les coûts de calcul si vous exécutez le .NET Core sous EC2 Linux au lieu d'exécuter le .NET Framework sous Windows EC2. Vous pouvez réaliser des économies supplémentaires de 40 % si vous utilisez l'architecture ARM (AWS Graviton) au lieu de x86.

Si vous envisagez d'exécuter des conteneurs basés sur Linux pour des applications .NET Framework existantes, vous devez porter ces applications vers des versions multiplateformes modernes de .NET ([telles que .NET 6.0](#)) afin d'utiliser des conteneurs Linux. L'un des principaux facteurs à prendre en compte est de comparer le coût de la refactorisation aux économies réalisées grâce à la réduction du coût des conteneurs Linux. Pour plus d'informations sur le portage de vos applications vers le .NET moderne, consultez [l'assistant de portage pour .NET](#) dans la AWS documentation.

Un autre avantage du passage au .NET moderne (c'est-à-dire en s'éloignant du .NET Framework) est que des opportunités de modernisation supplémentaires deviennent disponibles. Par exemple, vous

pouvez envisager de réorganiser l'architecture de votre application vers une architecture basée sur les microservices, plus évolutive, plus agile et plus rentable.

Le schéma suivant illustre le processus de prise de décision pour explorer les opportunités de modernisation.



Profitez des Savings Plans

Les conteneurs peuvent vous aider à tirer parti des [Compute Savings Plans](#) pour réduire vos coûts liés à Fargate. Le modèle de réduction flexible offre les mêmes remises que les instances réservées convertibles. La tarification de Fargate est basée sur le vCPU et les ressources de mémoire utilisées entre le moment où vous commencez à télécharger votre image de conteneur et la fin de la tâche Amazon ECS (arrondi à la seconde près). [Les Savings Plans for Fargate](#) offrent des économies allant jusqu'à 50 % sur l'utilisation de Fargate en échange d'un engagement à utiliser une quantité spécifique d'utilisation informatique (mesurée en dollars par heure) pour une durée d'un an ou trois ans. Vous pouvez l'utiliser [AWS Cost Explorer](#) pour vous aider à choisir un Savings Plans.

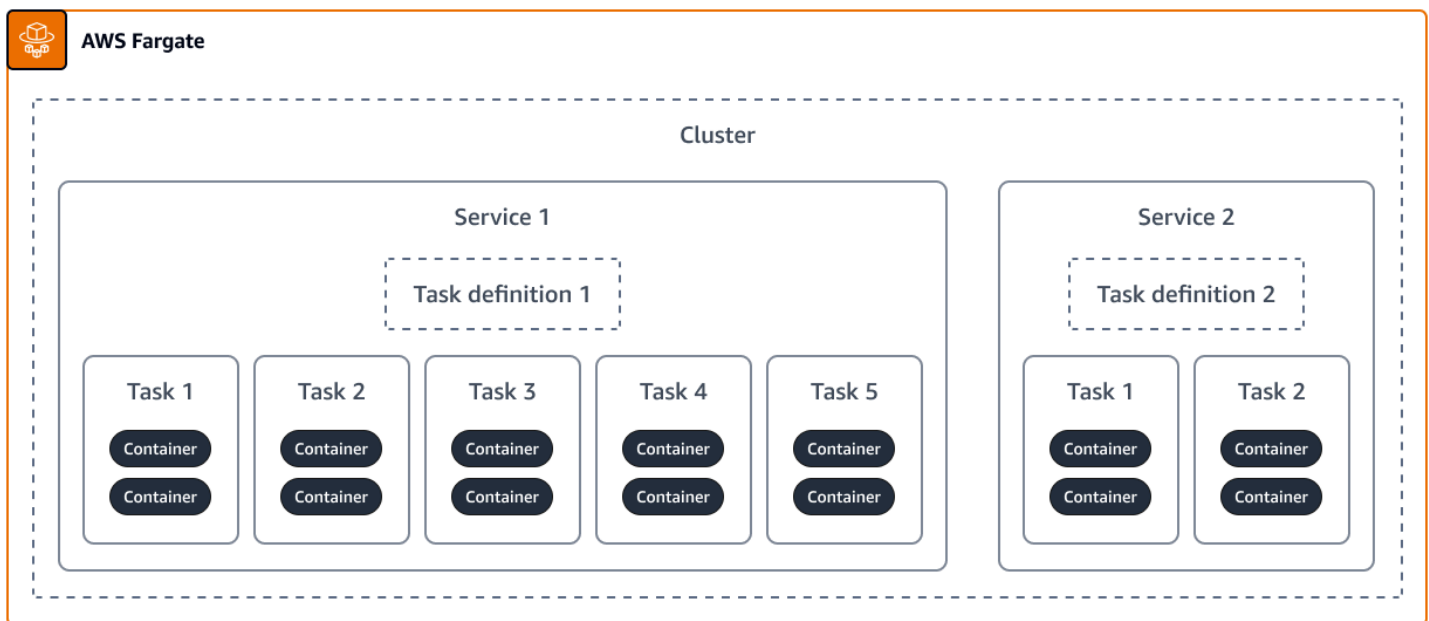
Il est important de comprendre que les Compute Savings Plans s'appliquent d'abord à l'utilisation qui vous permet de réaliser les économies les plus importantes. Par exemple, si vous exécutez une instance Linux t3.medium dans us-east-2 une instance Windows t3.medium identique, l'instance Linux bénéficie en premier de l'avantage Savings Plans. Cela est dû au fait que l'instance Linux a un potentiel d'économie de 50 % alors que la même instance Windows a un potentiel d'économie de 35 %. Si vous disposez d'autres ressources éligibles aux Savings Plans Compte AWS, telles qu'Amazon EC2 ou Lambda, il n'est pas nécessaire que vos Savings Plans soient d'abord appliqués à Fargate. Pour plus d'informations, consultez la section [Comprendre comment les plans d'épargne](#)

[s'appliquent à votre AWS utilisation](#) dans la documentation relative aux plans d'épargne et dans la section [Optimize spending for Windows on Amazon EC2](#) de ce guide.

Tâches Fargate de la bonne taille

Il est important de s'assurer que les tâches Fargate sont correctement dimensionnées afin d'optimiser au maximum les coûts. Souvent, les développeurs ne disposent pas de toutes les informations d'utilisation nécessaires lorsqu'ils déterminent initialement les configurations des tâches Fargate utilisées dans leurs applications. Cela peut entraîner un surprovisionnement des tâches et entraîner des dépenses inutiles. Pour éviter cela, nous vous recommandons de charger des applications de test exécutées sur Fargate afin de comprendre les performances d'une configuration de tâche spécifique dans différents scénarios d'utilisation. Vous pouvez utiliser les résultats des tests de charge, les vCPU, l'allocation de mémoire des tâches et les politiques d'autoscaling pour trouver le juste équilibre entre performances et coûts.

Le schéma suivant montre comment Compute Optimizer génère des recommandations pour la taille optimale des tâches et des conteneurs.



L'une des approches consiste à utiliser un outil de test de charge, tel que celui décrit dans [Distributed Load Testing on AWS](#), pour établir une base de référence pour l'utilisation des vCPU et de la mémoire. Après avoir effectué le test de charge pour simuler une charge d'application typique, vous pouvez affiner la configuration du vCPU et de la mémoire pour la tâche jusqu'à ce que l'utilisation de base soit atteinte.

Ressources supplémentaires

- [Liste de contrôle pour l'optimisation des coûts pour Amazon ECS et AWS Fargate](#) (article de blog sur AWS les conteneurs)
- [Optimisation théorique des coûts par type de lancement d'Amazon ECS : Fargate vs EC2 AWS](#) (article de blog sur les conteneurs)
- [Assistant de portage pour .NET](#) (AWS documentation)
- [Test de charge distribué sur AWS](#) (bibliothèque de AWS solutions)
- [Optimiseur de calcul AWS lance le support pour les services Amazon ECS sur AWS Fargate](#) (article de blog sur la gestion financière AWS dans le cloud)

Utiliser des instances et des conteneurs Graviton

Présentation de

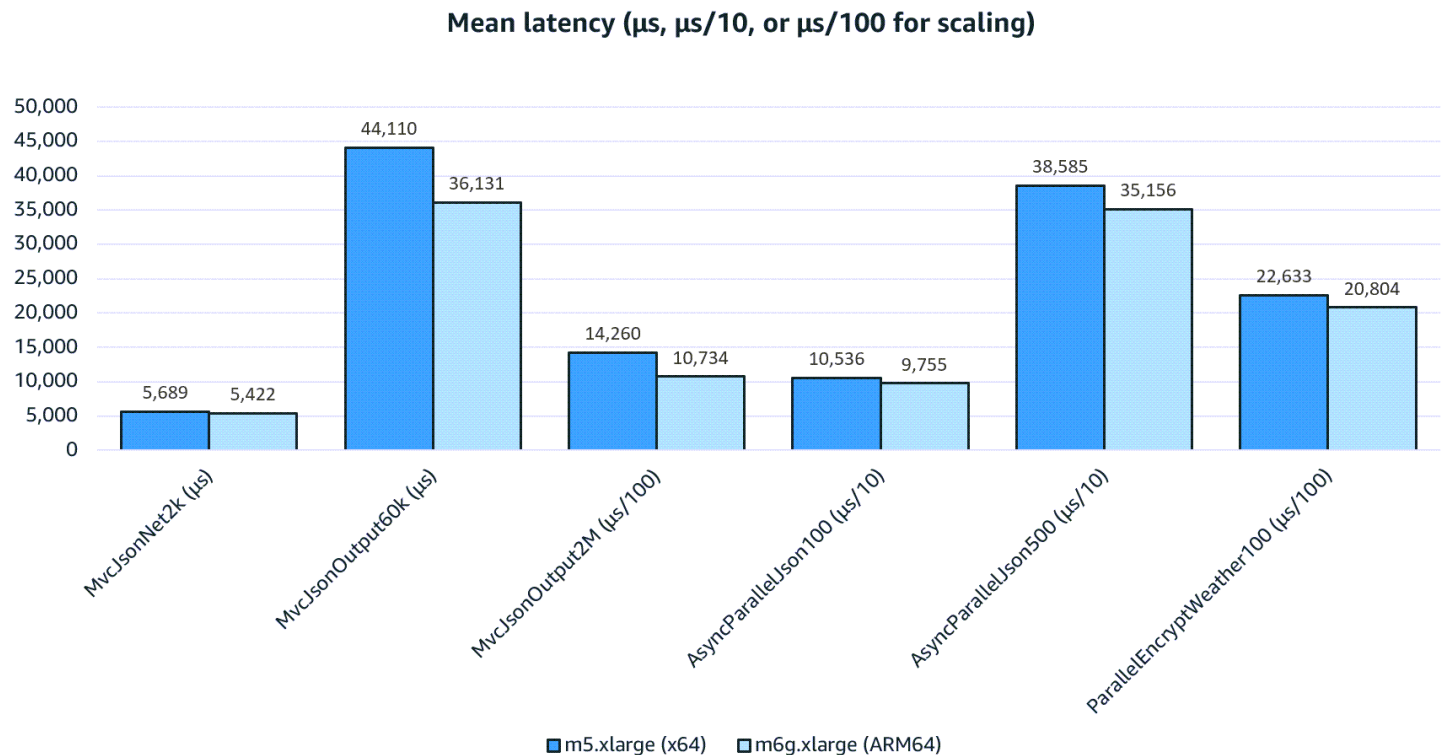
AWS Les instances Graviton sont alimentées par des processeurs ARM conçus AWS pour offrir le meilleur rapport prix/performances pour vos charges de travail cloud exécutées sur Amazon Elastic Compute Cloud (Amazon EC2), y compris les conteneurs qui s'exécutent. AWS Trois générations de Graviton sont actuellement disponibles pour une utilisation sur Amazon EC2. Ce guide se concentre sur l'utilisation de Graviton 2 et 3 avec les applications .NET, car l'utilisation des dernières versions de Graviton permet de réaliser d'importantes économies. N'oubliez pas que les instances Graviton n'exécutent que le système d'exploitation Linux. Par conséquent, les instances Graviton constituent une offre intéressante pour .NET s'exécutant sous Linux, mais ne constituent pas une option pour le système d'exploitation Windows ou les applications .NET Framework existantes.

Graviton 3 est 60 % plus efficace que les instances EC2 comparables, avec des performances jusqu'à 40 % supérieures. Ce guide met l'accent sur les avantages financiers liés à l'utilisation de Graviton, mais il est important de noter que le graviton offre des avantages supplémentaires en termes d'amélioration des performances et de durabilité environnementale.

Impact sur les coûts

Vous pouvez réaliser jusqu'à 45 % d'économies en optant pour Graviton. Après avoir refactorisé toutes les applications .NET Framework existantes vers une version .NET moderne, vous débloquez la possibilité d'utiliser des instances Graviton. Le passage à Graviton est une technique d'optimisation des coûts efficace pour les développeurs .NET.

L'exemple du tableau suivant montre le potentiel d'amélioration des performances que vous pouvez obtenir en migrant vers des instances Graviton.



Pour une analyse complète et une explication de l'approche d'analyse comparative utilisée pour créer les résultats du schéma précédent, voir [Powing .NET 5 with AWS Graviton2 : Benchmarks](#) in the AWS Compute Blog.

L'une des raisons de l'amélioration de l'efficacité est la différence de signification du vCPU entre x86 et Graviton. Dans l'architecture x86, un vCPU est un cœur logique réalisé par hyperthreading. Dans Graviton, le vCPU correspond à un cœur physique qui permet au vCPU d'être entièrement affecté à la charge de travail.

Le résultat avec Graviton2 est un rapport prix/performances 40 % supérieur à celui des instances x86/x64 comparables. Graviton3 offre les avantages suivants par rapport à Graviton2 :

- Un profil de performance amélioré avec des performances jusqu'à 25 % supérieures
- Performances en virgule flottante jusqu'à deux fois supérieures
- Performances de charge de travail cryptographique jusqu'à deux fois plus rapides
- Performances d'apprentissage automatique jusqu'à trois fois supérieures

De plus, Graviton3 est la première instance du cloud à intégrer de la mémoire. DDR5

Les tableaux suivants montrent la différence en termes d'économies de coûts entre les instances basées sur Graviton et les instances x86 équivalentes.

Ce tableau montre les économies de 19,20 % réalisées par Graviton.

Type d'instance	Architecture	vCPU	Mémoire (Go)	Coût horaire (sur demande)
t4g.xlarge	ARM	4	16	0,1344\$
t3.xlarge	x86	4	16	0,1664\$

Ce tableau montre les économies de 14,99 % réalisées par Graviton.

Type d'instance	Architecture	vCPU	Mémoire (Go)	Coût horaire (sur demande)
c7g.4xlarge	ARM	16	32	0,5781\$
c6i.4xlarge	x86	16	32	0,6800\$

Il est important de tester le profil de performance de votre application lorsque vous envisagez Graviton. Le graviton ne remplace pas de solides pratiques de développement logiciel. Vous pouvez utiliser les tests pour vérifier si vous tirez le meilleur parti de vos ressources informatiques sous-jacentes.

Recommandations en matière d'optimisation des coûts

Il existe plusieurs manières de tirer parti des processeurs/instances Graviton. Cette section décrit les modifications nécessaires pour passer d'une machine à architecture x86 à des instances Graviton (ARM).

Modifier le paramètre d'exécution dans Lambda

Nous vous recommandons d'activer les paramètres d'exécution AWS Lambda. Pour plus d'informations, consultez la section [Modification de l'environnement d'exécution](#) dans la

documentation Lambda. .NET étant un langage compilé, vous devez suivre un processus de compilation pour que cela fonctionne. Pour un exemple de la procédure à suivre, voir. [NET sur Graviton](#) in. GitHub

Containers

Pour une charge de travail conteneurisée, créez une image de conteneur multi-architecture. Vous pouvez le faire en spécifiant plusieurs architectures dans la commande Docker build. Par exemple :

```
docker buildx build -t "myImageName:latest" --platform linux/amd64,linux/arm64 --push .
```

Vous pouvez également utiliser un outil tel que AWS Cloud Development Kit (AWS CDK) pour aider à [orchestrer le build](#). Pour des exemples tirés de Docker, consultez la section [Création d'images multi-arches pour Arm et x86 avec Docker Desktops](#) dans la documentation Docker.

Amazon EC2

Pour migrer vers ARM depuis x86/x64, ciblez l'architecture ARM lors de l'étape de compilation. Dans Visual Studio, vous pouvez créer un ARM64 processeur. Pour obtenir des instructions, consultez [la section Pour configurer un projet pour cibler Arm64 et d'autres plateformes](#) dans la documentation Microsoft.

Si vous utilisez la CLI .NET, l'exécution de la compilation sur une machine ARM produit une version compatible avec Graviton. Pour voir une démonstration, regardez [Accelerate .NET 6 performance with Arm64 on AWS Graviton2 on](#). YouTube Les problèmes de dépendance entraîneront des erreurs de compilation qui pourront ensuite être traitées individuellement. Tant qu'il existe des bibliothèques ARM pour chaque dépendance, la transition devrait être relativement simple.

Ressources supplémentaires

- [Comment créer vos conteneurs pour ARM et économiser avec les instances Graviton et Spot sur Amazon ECS](#) (AWS blog)
- [AWS Lambda Fonctions optimisées par le processeur AWS Graviton2 : exécutez vos fonctions comme bon vous semble et bénéficiez d'un rapport qualité-prix jusqu'à 34 % supérieur](#) (blog)AWS
- [Migration des AWS Lambda fonctions vers les processeurs AWS Graviton2 basés sur ARM](#) (blog)AWS
- [Créez et déployez des applications Web .NET sur des clusters Amazon ECS AWS Graviton 2 alimentés par ARM à l'aide AWS CDK](#) de (blog)AWS

- [Graviton Fast Start — Un nouveau programme pour vous aider à transférer vos charges de travail vers AWS Graviton](#) (blog)AWS
- [Alimenter .NET 5 avec AWS Graviton2 : Benchmarks](#) (blog)AWS

Support du dimensionnement dynamique pour les applications .NET Framework statiques

Présentation de

L'un des principaux avantages de l'utilisation du cloud pour les applications est l'élasticité, c'est-à-dire la capacité de faire évoluer le calcul en fonction de la demande. Cela vous permet de payer uniquement pour la capacité de calcul dont vous avez besoin, plutôt que de la provisionner en cas de pic d'utilisation. Le Cyber Monday, au cours duquel les détaillants en ligne peuvent rapidement obtenir beaucoup plus de trafic que d'habitude (par exemple, [des milliers de pour cent en quelques minutes](#)), est un bon exemple d'élasticité.

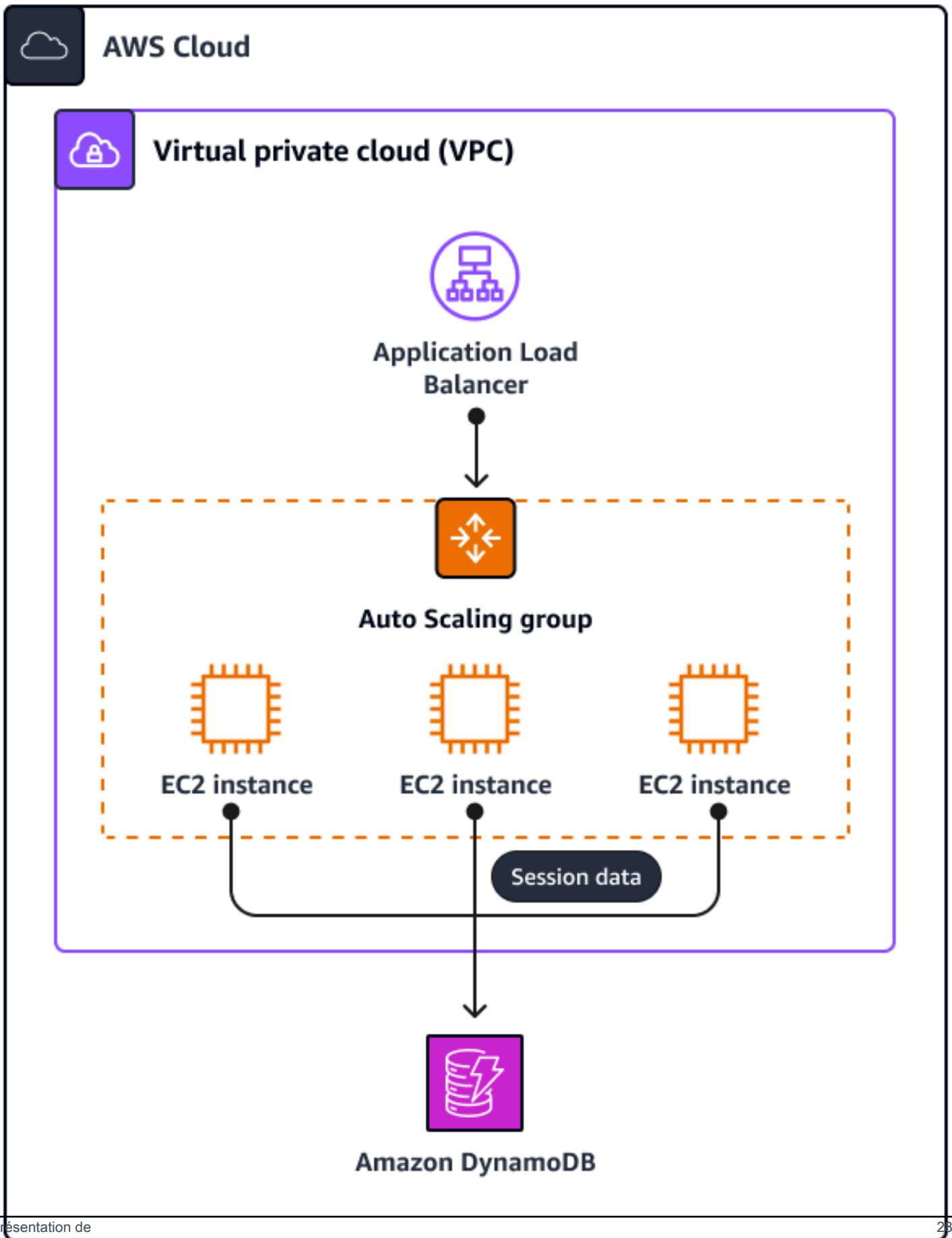
Si vous transférez des applications Web .NET héritées vers le cloud (par exemple, des applications ASP.NET Framework exécutées sur IIS), il peut être difficile, voire impossible, de dimensionner rapidement des batteries de serveurs à charge équilibrée en raison de la nature dynamique de l'application. Les données de session utilisateur sont stockées dans la mémoire de l'application, généralement avec l'[état de session ASP.NET](#) ou des variables statiques contenant des données de requêtes croisées qui doivent être conservées. L'affinité de session utilisateur est généralement maintenue par le biais de sessions permanentes de l'équilibreur de charge.

Cela s'avère difficile sur le plan opérationnel. Lorsqu'une capacité accrue est requise, vous devez intentionnellement provisionner et ajouter des serveurs. Ce processus peut être lent. La mise hors service des nœuds en cas d'application de correctifs ou de défaillances inattendues peut nuire à l'expérience de l'utilisateur final et entraîner une perte d'état pour tous les utilisateurs associés aux nœuds concernés. Dans le meilleur des cas, cela obligerait les utilisateurs à se reconnecter.

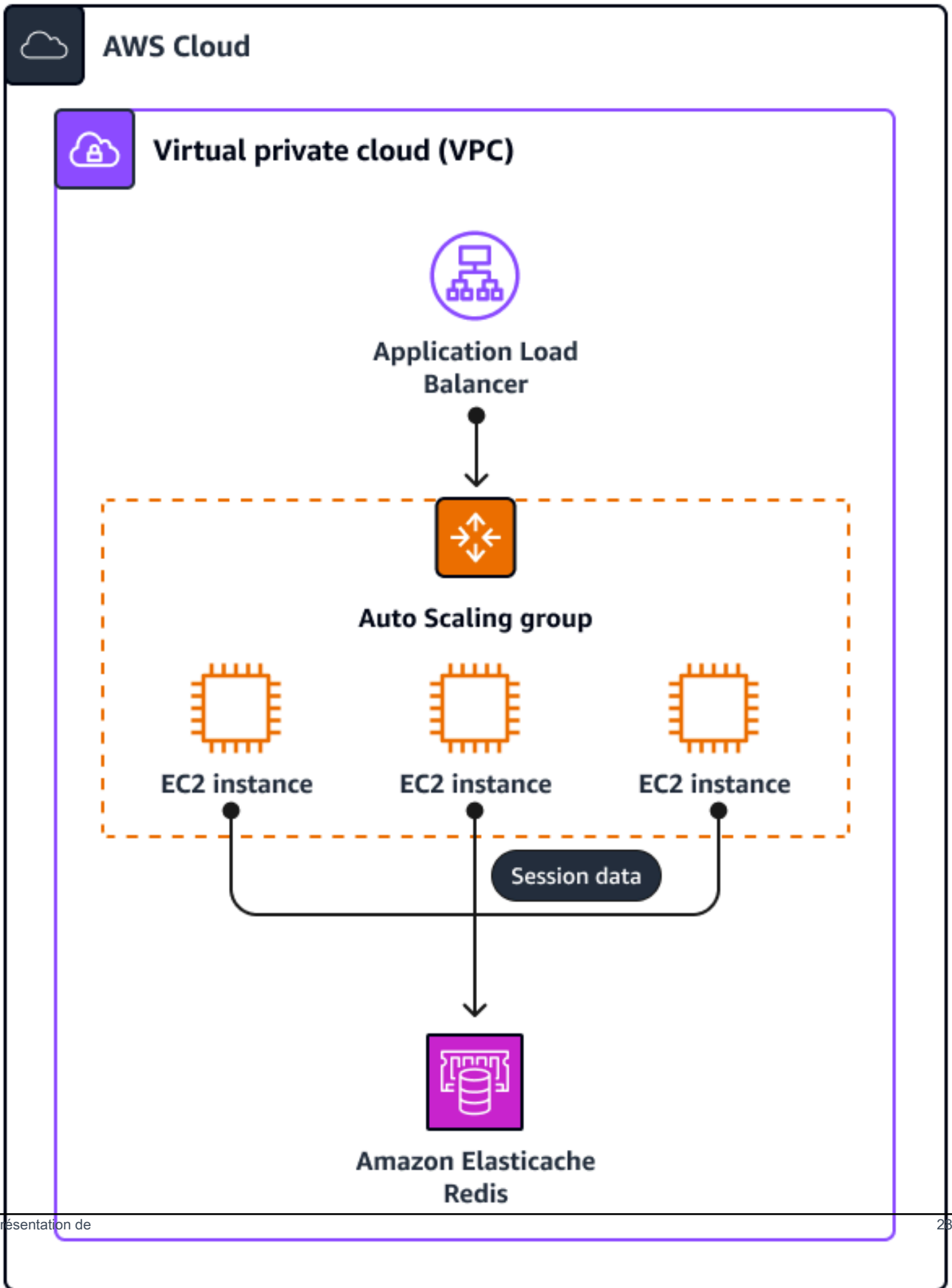
En centralisant l'état de session pour les applications ASP.NET et en appliquant des règles de mise à l'échelle automatique aux applications ASP.NET existantes, vous pouvez tirer parti de l'élasticité du cloud et éventuellement réaliser des économies lors de l'exécution d'applications. Par exemple, vous pouvez obtenir des réductions de coûts grâce à l'évolutivité du calcul, mais vous pouvez également choisir parmi les différents modèles de tarification disponibles, tels que la réduction de l'utilisation des [instances réservées et l'utilisation](#) de la [tarification des instances Amazon Spot](#).

Deux techniques courantes incluent l'utilisation d'[Amazon DynamoDB en tant que fournisseur d'état de session et l'utilisation d' ElastiCache Amazon \(Redis OSS\) en tant que](#) magasin de sessions ASP.NET.

Le schéma suivant montre une architecture qui utilise DynamoDB comme fournisseur d'état de session.



Le schéma suivant montre une architecture qui utilise ElastiCache (Redis OSS) comme fournisseur d'état de session.



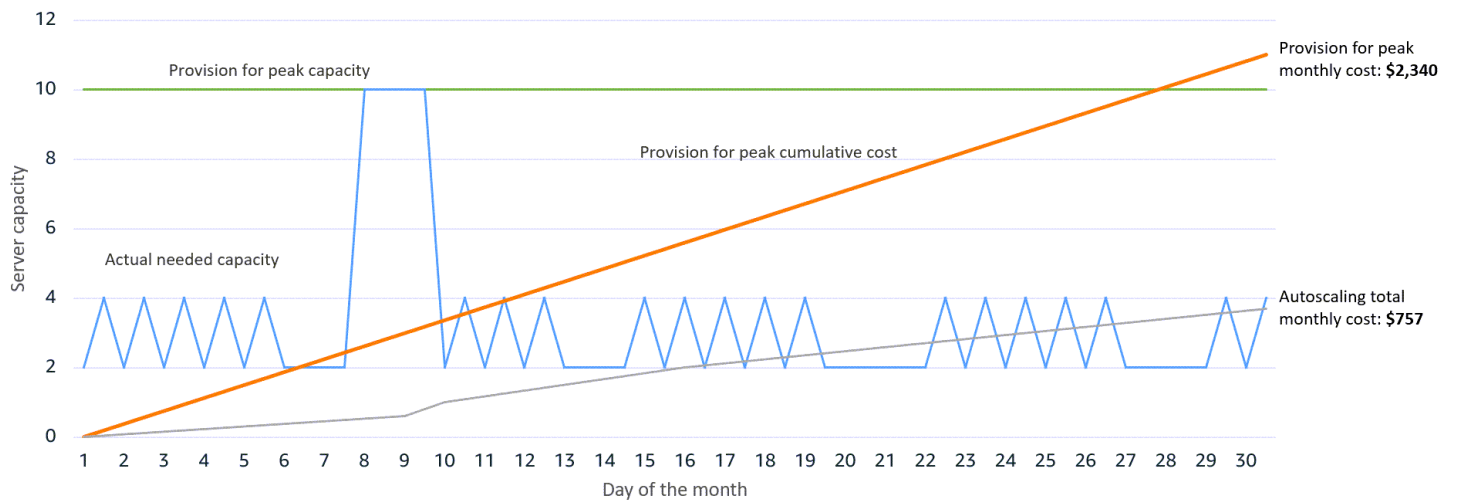
Impact sur les coûts

Pour déterminer les avantages de la mise à l'échelle pour une application de production, nous vous recommandons de modéliser votre demande réelle. Cette section repose sur les hypothèses suivantes pour modéliser un exemple d'application :

- Les instances ajoutées et supprimées de la rotation sont identiques et aucune variation de taille d'instance n'est introduite.
- Le taux d'utilisation des serveurs ne descend jamais en dessous de deux serveurs actifs afin de maintenir la haute disponibilité de l'application.
- Le nombre de serveurs évolue de façon linéaire en fonction du trafic (c'est-à-dire que deux fois plus de trafic nécessitera deux fois plus de calcul).
- Le trafic est modélisé au cours d'un mois par tranches de six heures, avec des variations au cours d'une journée et un pic de trafic anormal (par exemple, une vente promotionnelle) pour une journée où le trafic est multiplié par 10. Le trafic du week-end est modélisé en fonction de l'utilisation de base.
- Le trafic nocturne est modélisé en fonction de l'utilisation de base, tandis que le trafic en semaine est modélisé selon un taux d'utilisation multiplié par 4.
- La tarification des instances réservées utilise une tarification d'un an, sans frais initiaux. La tarification diurne normale utilise la tarification à la demande, tandis que la demande en rafale utilise la tarification par instance Spot.

Le schéma suivant montre comment ce modèle tire parti de l'élasticité d'une application .NET plutôt que de le provisionner en cas de pic d'utilisation. Cela se traduit par des économies d'environ 68 %.

Comparison of cumulative costs for peak provisioning and autoscaling



Si vous utilisez DynamoDB comme mécanisme de stockage de l'état de session, utilisez les paramètres suivants :

Storage: 20GB
 Session Reads: 40 million
 Session Writes: 20 million
 Pricing Model: On demand

Le coût mensuel estimé de ce service est d'environ 35\$ par mois.

Si vous utilisez ElastiCache (Redis OSS) comme mécanisme de stockage de l'état de session, utilisez les paramètres suivants :

Number of Nodes: 3
 Node size: cache.t4g.medium
 Pricing Model: 1y reserved

Le coût mensuel estimé de ce service est d'environ 91\$ par mois.

Recommandations en matière d'optimisation des coûts

La première étape consiste à implémenter l'état de session dans une ancienne application .NET. Si vous l'utilisez ElastiCache comme mécanisme de stockage d'état, suivez les instructions fournies dans le blog AWS Developer Tools en [ElastiCache tant que magasin de sessions ASP.NET](#). Si

vous utilisez DynamoDB, suivez les instructions de la section [What is AWS SDK pour .NET the documentation](#). SDK pour .NET

Si l'application utilise la InProcSession pour commencer, assurez-vous que tous les objets que vous prévoyez de stocker dans la session peuvent être sérialisés. Pour ce faire, utilisez l'SerializableAttribute attribut pour décorer les classes dont les instances seront stockées dans la session. Par exemple :

```
[Serializable()]
public class TestSimpleObject {
    public string SessionProperty {get;set;}
}
```

En outre, le .NET MachineKey doit être le même sur tous les serveurs utilisés. C'est généralement le cas lorsque les instances sont créées à partir d'une Amazon Machine Image (AMI) commune. Par exemple :

```
<machineKey
    validationKey="some long hashed value"
    decryptionKey="another long hashed value"
    validation="SHA1"/>
```

Cependant, il est important de s'assurer que si une image de base est modifiée, elle est configurée avec la même image de machine .NET (configurable au niveau IIS ou au niveau du serveur). Pour plus d'informations, consultez [SystemWebSectionGroup. MachineKey](#) Propriété figurant dans la documentation Microsoft.

Enfin, vous devez déterminer le mécanisme d'ajout de serveurs à un groupe Auto Scaling en réponse à un événement de dimensionnement. Il existe plusieurs moyens d'y parvenir. Nous recommandons les méthodes suivantes pour déployer de manière fluide des applications .NET Framework sur une instance EC2 d'un groupe Auto Scaling :

- Utilisez [EC2 Image Builder](#) pour configurer une AMI contenant le serveur et l'application entièrement configurés. Vous pouvez ensuite utiliser cette AMI pour configurer le [modèle de lancement de votre groupe Auto Scaling](#).
- [AWS CodeDeploy](#) À utiliser pour déployer votre application. CodeDeploy permet une intégration directe avec [Amazon EC2 Auto Scaling](#). Cela constitue une alternative à la création d'une nouvelle AMI pour chaque version de l'application.

Ressources supplémentaires

- [Création d'images avec EC2 Image Builder](#) (documentation EC2 Image Builder)
- [Déploiement d'applications Web .NET à l' AWS CodeDeploy aide de Visual Studio Team Services](#) (blog AWS consacré aux outils de développement)

Utiliser la mise en cache pour réduire la demande de base de données

Présentation de

Vous pouvez utiliser la mise en cache comme stratégie efficace pour réduire les coûts de vos applications .NET. De nombreuses applications utilisent des bases de données principales, telles que SQL Server, lorsqu'elles nécessitent un accès fréquent aux données. Le coût de maintenance de ces services principaux pour répondre à la demande peut être élevé, mais vous pouvez utiliser une stratégie de mise en cache efficace pour réduire la charge sur les bases de données principales en réduisant les exigences de dimensionnement et de dimensionnement. Cela peut vous aider à réduire les coûts et à améliorer les performances de vos applications.

La mise en cache est une technique utile pour réduire les coûts liés aux charges de travail lourdes en lecture qui utilisent des ressources plus coûteuses telles que SQL Server. Il est important d'utiliser la bonne technique adaptée à votre charge de travail. Par exemple, la mise en cache locale n'est pas évolutive et vous oblige à maintenir un cache local pour chaque instance d'une application. Vous devez évaluer l'impact sur les performances par rapport aux coûts potentiels, afin que le moindre coût de la source de données sous-jacente compense les coûts supplémentaires liés au mécanisme de mise en cache.

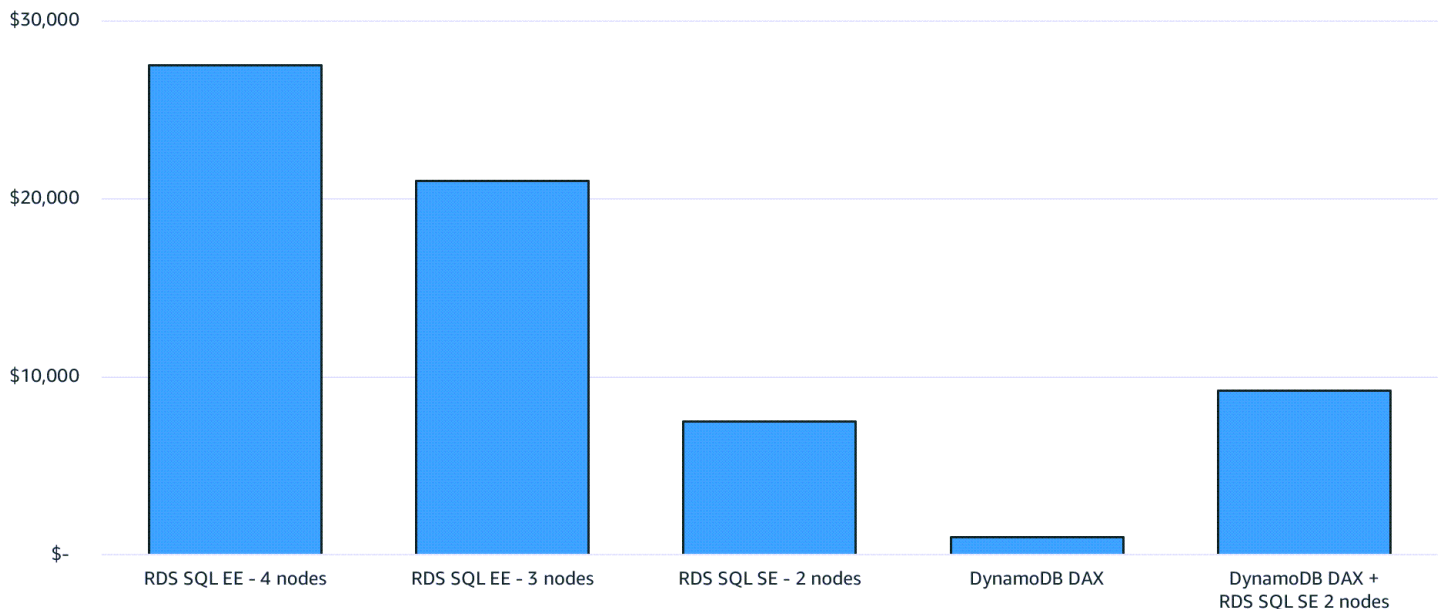
Impact sur les coûts

SQL Server vous oblige à prendre en compte les demandes de lecture lors du dimensionnement de votre base de données. Cela peut avoir une incidence sur les coûts, car vous devrez peut-être introduire des répliques de lecture pour faire face à la charge. Si vous utilisez des répliques de lecture, il est important de comprendre qu'elles ne sont disponibles que sur l'édition Enterprise de SQL Server. Cette édition nécessite une licence plus onéreuse que l'édition standard de SQL Server.

Le schéma suivant est conçu pour vous aider à comprendre l'efficacité de la mise en cache. Il montre Amazon RDS for SQL Server avec quatre nœuds db.m4.2xlarge exécutant SQL Server Enterprise

Edition. Il est déployé dans une configuration multi-AZ avec une seule réplique en lecture. Le trafic de lecture exclusif (par exemple, les requêtes SELECT) est dirigé vers les répliques de lecture. En comparaison, Amazon DynamoDB utilise un cluster DynamoDB Accelerator (DAX) r4.2xlarge à deux nœuds.

Le graphique suivant montre les résultats de la suppression du besoin de répliques de lecture dédiées qui gèrent un trafic de lecture élevé.



Vous pouvez réaliser d'importantes économies en utilisant la mise en cache locale sans répliques en lecture ou en introduisant DAX côte à côte avec SQL Server sur Amazon RDS en tant que couche de mise en cache. Cette couche se décharge de SQL Server et réduit la taille du serveur SQL requis pour exécuter la base de données.

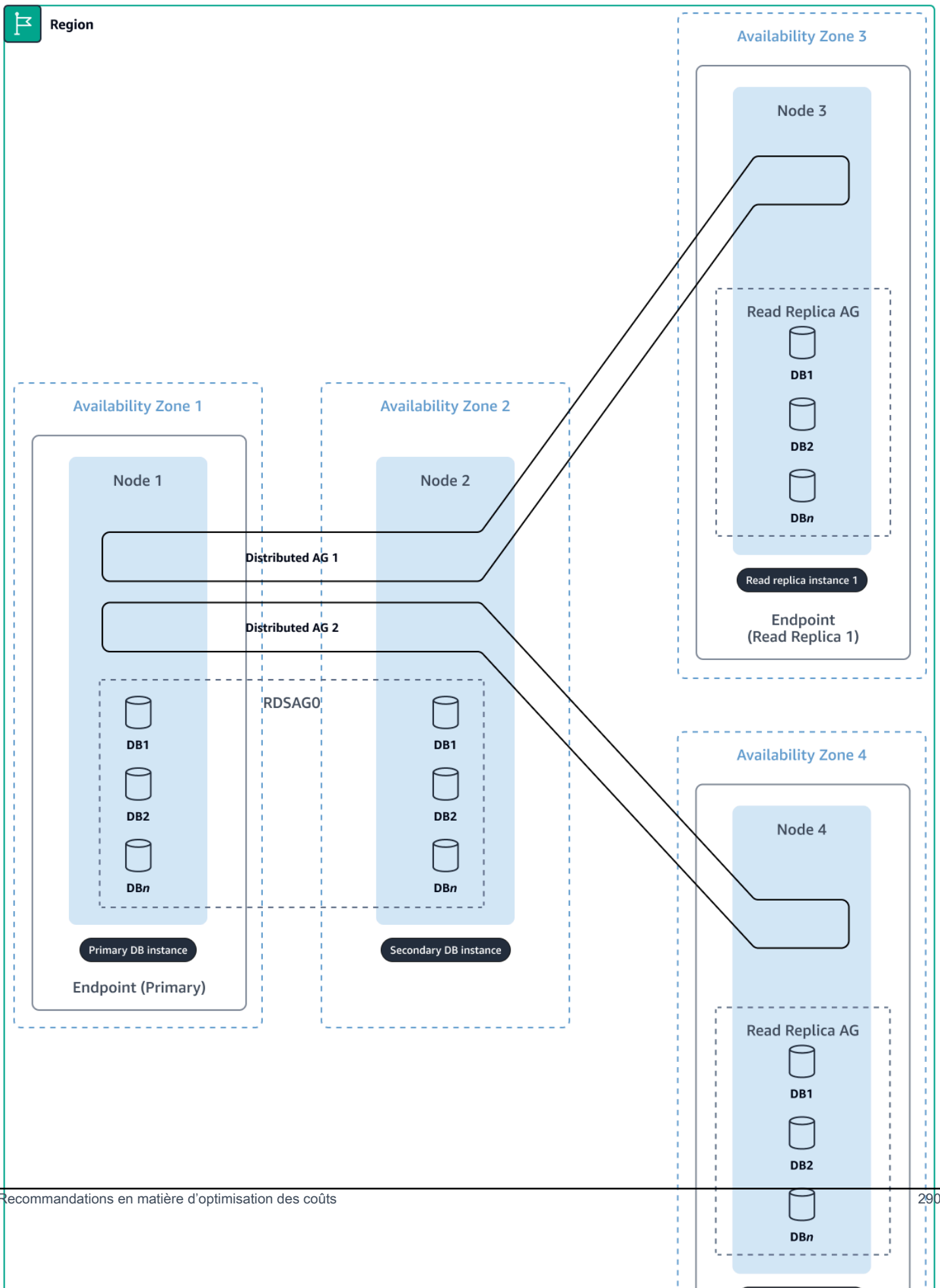
Recommandations en matière d'optimisation des coûts

Mise en cache locale

La mise en cache locale est l'une des méthodes les plus couramment utilisées pour mettre en cache le contenu des applications hébergées à la fois dans des environnements sur site ou dans le cloud. Cela s'explique par le fait qu'il est relativement facile et intuitif à mettre en œuvre. La mise en cache locale consiste à récupérer le contenu d'une base de données ou d'une autre source et à le mettre en cache localement en mémoire ou sur disque pour un accès plus rapide. Cette approche, bien que facile à mettre en œuvre, n'est pas idéale pour certains cas d'utilisation. Cela inclut, par exemple, les cas d'utilisation où le contenu de la mise en cache doit persister dans le temps, par exemple

pour préserver l'état de l'application ou l'état de l'utilisateur. Un autre cas d'utilisation est celui où le contenu mis en cache doit être accessible à partir d'autres instances d'application.

Le schéma ci-dessous illustre un cluster SQL Server à haute disponibilité avec quatre nœuds et deux répliques en lecture.



Avec la mise en cache locale, vous devrez peut-être équilibrer la charge du trafic entre plusieurs instances EC2. Chaque instance doit gérer son propre cache local. Si le cache stocke des informations dynamiques, des validations régulières doivent être effectuées dans la base de données, et les utilisateurs devront peut-être être redirigés vers la même instance pour chaque demande suivante (session persistante). Cela représente un défi lorsque vous essayez de faire évoluer les applications, car certaines instances peuvent être surutilisées, tandis que d'autres sont sous-utilisées en raison de la répartition inégale du trafic.

Vous pouvez utiliser la mise en cache locale, en mémoire ou via le stockage local, pour les applications .NET. Pour ce faire, vous pouvez ajouter des fonctionnalités permettant soit de stocker des objets sur disque et de les récupérer si nécessaire, soit d'interroger des données dans la base de données et de les conserver en mémoire. Pour effectuer une mise en cache locale en mémoire et sur le stockage local de données à partir d'un serveur SQL en C#, par exemple, vous pouvez utiliser une combinaison de `MemoryCache` bibliothèques et `LiteDB`. `MemoryCache` fournit une mise en cache en mémoire, tandis qu'`LiteDB` s'agit d'une base de données NoSQL intégrée sur disque rapide et légère.

Pour effectuer une mise en cache en mémoire, utilisez la bibliothèque .NET.

`System.Runtime.MemoryCache` L'exemple de code suivant montre comment utiliser la `System.Runtime.Caching.MemoryCache` classe pour mettre en cache des données en mémoire. Cette classe permet de stocker temporairement des données dans la mémoire de l'application. Cela peut contribuer à améliorer les performances d'une application en réduisant le besoin de récupérer des données à partir d'une ressource plus coûteuse, telle qu'une base de données ou une API.

Voici comment fonctionne le code :

1. Une instance statique privée de `MemoryCache` named `_memoryCache` est créée. Le cache reçoit un nom (`dataCache`) pour l'identifier. Ensuite, le cache stocke et récupère les données.
2. La `GetData` méthode est une méthode générique qui prend deux arguments : une `string` clé et un `Func<T>` délégué appelé `getData`. La clé est utilisée pour identifier les données mises en cache, tandis que le `getData` délégué représente la logique de récupération des données qui est exécutée lorsque les données ne sont pas présentes dans le cache.
3. La méthode vérifie d'abord si les données sont présentes dans le cache à l'aide de la `_memoryCache.Contains(key)` méthode. Si les données se trouvent dans le cache, la méthode les récupère en utilisant `_memoryCache.Get(key)` et les convertit dans le type attendu `T`.

4. Si les données ne sont pas dans le cache, la méthode appelle le `getData` délégué pour récupérer les données. Ensuite, il ajoute les données au cache en utilisant `_memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10))`. Cet appel indique que l'entrée du cache doit expirer au bout de 10 minutes, date à laquelle les données sont automatiquement supprimées du cache.
5. La `ClearCache` méthode prend une `string` clé comme argument et supprime les données associées à cette clé du cache en utilisant `_memoryCache.Remove(key)`.

```
using System;
using System.Runtime.Caching;

public class InMemoryCache
{
    private static MemoryCache _memoryCache = new MemoryCache("dataCache");

    public static T GetData<T>(string key, Func<T> getData)
    {
        if (_memoryCache.Contains(key))
        {
            return (T)_memoryCache.Get(key);
        }

        T data = getData();
        _memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10));

        return data;
    }

    public static void ClearCache(string key)
    {
        _memoryCache.Remove(key);
    }
}
```

Vous pouvez utiliser le code suivant :

```
public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";
    }
}
```

```
Func<string> getSampleData = () =>
{
    // Replace this with your data retrieval logic
    return "Sample data";
};

string data = InMemoryCache.GetData(cacheKey, getSampleData);
Console.WriteLine("Data: " + data);
}
}
```

L'exemple suivant montre comment utiliser [LiteDB](#) pour mettre en cache des données dans le stockage local. Vous pouvez utiliser LiteDB comme alternative ou complément à la mise en cache en mémoire. Le code suivant montre comment utiliser la bibliothèque LiteDB pour mettre en cache des données dans le stockage local. La `LocalStorageCache` classe contient les principales fonctions de gestion du cache.

```
using System;
using LiteDB;

public class LocalStorageCache
{
    private static string _liteDbPath = @"Filename=LocalCache.db";

    public static T GetData<T>(string key, Func<T> getData)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
            var item = collection.FindOne(Query.EQ("_id", key));

            if (item != null)
            {
                return item;
            }
        }

        T data = getData();

        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
```

```
        collection.Upsert(new BsonValue(key), data);
    }

    return data;
}

public static void ClearCache(string key)
{
    using (var db = new LiteDatabase(_liteDbPath))
    {
        var collection = db.GetCollection("cache");
        collection.Delete(key);
    }
}

public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";

        Func<string> getSampleData = () =>
        {
            // Replace this with your data retrieval logic
            return "Sample data";
        };

        string data = LocalStorageCache.GetData(cacheKey, getSampleData);
        Console.WriteLine("Data: " + data);
    }
}
```

Si vous disposez d'un cache statique ou de fichiers statiques qui ne changent pas fréquemment, vous pouvez également stocker ces fichiers dans le stockage d'objets Amazon Simple Storage Service (Amazon S3). L'application peut récupérer le fichier de cache statique au démarrage pour l'utiliser localement. Pour plus de détails sur la façon de récupérer des fichiers depuis Amazon S3 à l'aide de .NET, consultez la section [Téléchargement d'objets](#) dans la documentation Amazon S3.

Mise en cache avec DAX

Vous pouvez utiliser une couche de mise en cache qui peut être partagée entre toutes les instances d'application. L'[accélérateur DynamoDB \(DAX\)](#) est un cache en mémoire entièrement géré et

hautement disponible pour DynamoDB qui peut décupler les performances. Vous pouvez utiliser le DAX pour réduire les coûts en réduisant le besoin de surprovisionner les unités de capacité de lecture dans les tables DynamoDB. Cela est particulièrement utile pour les charges de travail qui sont lourdes en lecture et nécessitent des lectures répétées pour des clés individuelles.

DynamoDB est facturé à la demande ou avec une capacité allouée, de sorte que le nombre de lectures et d'écritures par mois contribue au coût. Si vous avez de lourdes charges de travail en lecture, les clusters DAX peuvent contribuer à réduire les coûts en réduisant le nombre de lectures sur vos tables DynamoDB. Pour obtenir des instructions sur la configuration du DAX, consultez la section [Accélération en mémoire avec DynamoDB Accelerator \(DAX\) dans la documentation DynamoDB](#). Pour plus d'informations sur l'intégration des applications .NET, consultez [Intégrer Amazon DynamoDB DAX dans](#) votre application ASP.NET sur YouTube

Ressources supplémentaires

- [Accélération en mémoire avec DynamoDB Accelerator \(DAX\) - Amazon DynamoDB \(documentation DynamoDB\)](#)
- [Intégration d'Amazon DynamoDB DAX dans](#) votre application ASP.NET () YouTube
- [Téléchargement d'objets](#) (documentation Amazon S3)

Envisagez le .NET sans serveur

Présentation de

L'informatique sans serveur est devenue une approche populaire pour créer et déployer des applications. Cela est principalement dû à l'évolutivité et à l'agilité qu'offre l'approche sans serveur lors de la création d'une architecture moderne. Cependant, il est important de prendre en compte l'impact financier de l'informatique sans serveur dans certains scénarios.

Lambda est une plate-forme informatique sans serveur qui permet aux développeurs d'exécuter du code sans avoir besoin de serveurs dédiés. Lambda est une option particulièrement intéressante pour les développeurs .NET qui cherchent à réduire les coûts d'infrastructure. Avec Lambda, les développeurs .NET peuvent développer et déployer des applications hautement évolutives et potentiellement rentables. En utilisant une approche sans serveur, les développeurs ne fournissent plus de serveurs pour traiter les demandes d'applications. Les développeurs peuvent plutôt créer des fonctions exécutées à la demande. Cela rend une approche sans serveur plus évolutive, gérable et potentiellement plus rentable que l'exécution, la gestion et le dimensionnement de machines

virtuelles. Par conséquent, vous ne payez que pour les ressources utilisées par l'application, sans avoir à vous soucier de la sous-utilisation des ressources ou des coûts de maintenance du serveur.

Les développeurs peuvent utiliser des versions .NET modernes et multiplateformes pour créer des applications sans serveur rapides, efficaces et économiques. Le .NET Core et les versions plus récentes constituent un framework gratuit et open source mieux adapté à l'exécution sur des plateformes sans serveur que les versions précédentes de .NET Framework. Cela permet aux développeurs de réduire le temps de développement et d'améliorer les performances des applications. Le .NET moderne prend également en charge une gamme de langages de programmation, notamment le C# et le F#. C'est pourquoi il s'agit d'une option intéressante pour les développeurs qui cherchent à créer des architectures modernes dans le cloud.

Cette section explique comment vous pouvez réaliser des économies en utilisant Lambda comme option sans serveur. Vous pouvez optimiser davantage les coûts en affinant les profils d'exécution de vos fonctions Lambda, en dimensionnant correctement l'allocation de mémoire de vos fonctions Lambda, en [utilisant l'AOT natif](#) et en passant aux fonctions basées sur Graviton.

Impact sur les coûts

La mesure dans laquelle vous pouvez réduire les coûts dépend de plusieurs facteurs, notamment le nombre d'exécutions que vos fonctions sans serveur exécuteront, en plus de la quantité de mémoire allouée et de la durée de chaque fonction. AWS Lambda propose un niveau gratuit, qui inclut un million de requêtes gratuites par mois et 400 000 Go de secondes de temps de calcul par mois. Vous pouvez réduire de manière significative vos coûts mensuels pour les charges de travail correspondant ou non à ces limites du niveau gratuit.

L'utilisation d'un équilibreur de charge avec les fonctions Lambda comme cible peut également entraîner des coûts supplémentaires. Ceci est calculé comme la quantité de données traitées par l'équilibreur de charge pour les cibles [Lambda](#).

Recommandations en matière d'optimisation des coûts

Dimensionnez correctement vos fonctions Lambda

Le dimensionnement correct est une pratique essentielle pour optimiser les coûts dans les fonctions Lambda basées sur .NET. Ce processus consiste à identifier la configuration de mémoire optimale qui équilibre les performances avec la rentabilité, sans qu'il soit nécessaire de modifier le code.

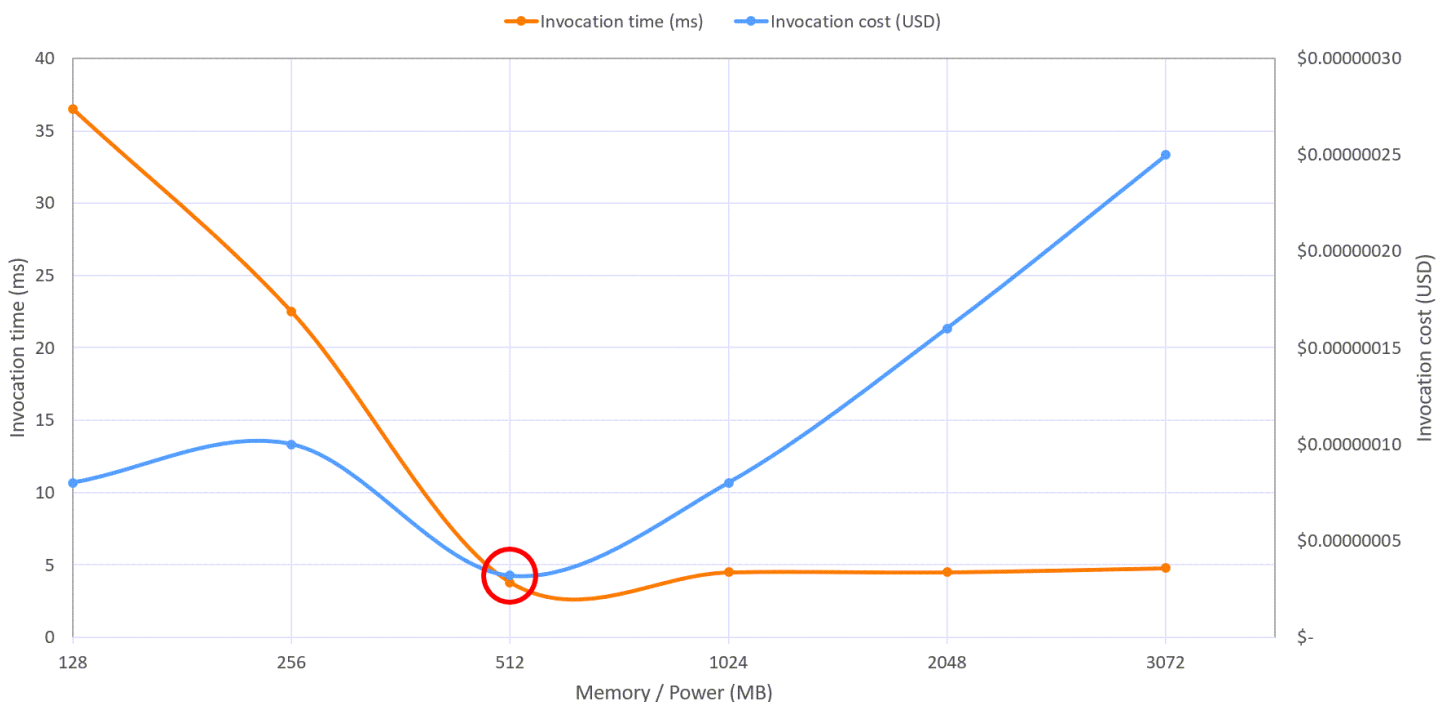
En configurant la mémoire pour une fonction Lambda, comprise entre 128 Mo et 10 240 Mo, vous ajustez également la quantité de vCPU disponible lors de l'invocation. Cela permet aux

applications liées à la mémoire ou au processeur d'accéder à des ressources supplémentaires pendant l'exécution, ce qui entraîne une réduction potentielle de la durée d'appel et du coût global.

Cependant, l'identification de la configuration optimale pour vos fonctions Lambda basées sur .NET peut être un processus manuel et chronophage, en particulier si les modifications sont fréquentes. L'[outil AWS Lambda Power Tuning](#) peut vous aider à identifier la configuration appropriée en analysant un ensemble de configurations de mémoire par rapport à un exemple de charge utile.

Par exemple, l'augmentation de la mémoire pour une fonction Lambda basée sur .NET peut améliorer le temps total d'invocation et réduire les coûts sans affecter les performances. La configuration de mémoire optimale pour une fonction peut varier. L'outil AWS Lambda Power Tuning peut aider à identifier la configuration la plus rentable pour chaque fonction.

Dans l'exemple de graphique suivant, le temps d'appel total augmente à mesure que la mémoire augmente pour cette fonction Lambda. Cela permet de réduire le coût total d'exécution sans affecter les performances initiales de la fonction. Pour cette fonction, la configuration de mémoire optimale est de 512 Mo, car c'est là que l'utilisation des ressources est la plus efficace pour le coût total de chaque appel. Cela varie selon les fonctions, et l'utilisation de l'outil sur vos fonctions Lambda permet de déterminer si elles bénéficient d'un dimensionnement correct.



Nous vous recommandons de réaliser cet exercice régulièrement, dans le cadre de tout test d'intégration lors de la publication de nouvelles mises à jour. En cas de mise à jour peu fréquente,

effectuez cet exercice régulièrement pour vous assurer que les fonctions sont réglées et correctement dimensionnées. Après avoir identifié le paramètre de mémoire approprié pour vos fonctions Lambda, vous pouvez ajouter le bon dimensionnement à vos processus. L'outil AWS Lambda Power Tuning génère une sortie programmatique qui peut être utilisée par vos flux de travail CI/CD lors de la publication du nouveau code. Cela vous permet d'automatiser la configuration de la mémoire.

Vous pouvez télécharger gratuitement l'[outil AWS Lambda Power Tuning](#). Pour obtenir des instructions sur l'utilisation de l'outil, reportez-vous [à la section Comment exécuter la machine à états](#) dans GitHub.

Lambda prend également en charge l'AOT natif, qui permet aux applications .NET d'être précompilées. Cela peut contribuer à réduire les coûts en réduisant les temps d'exécution des fonctions .NET. Pour plus d'informations sur la création de fonctions AOT natives, consultez la section [Fonctions .NET avec compilation AOT native](#) dans la documentation Lambda.

Évitez les temps d'attente inactifs

La durée de la fonction Lambda est une dimension utilisée pour calculer la facturation. Lorsque le code de fonction effectue un appel bloquant, le temps d'attente avant de recevoir une réponse vous est facturé. Ce temps d'attente peut augmenter lorsque les fonctions Lambda sont enchaînées ou lorsqu'une fonction agit en tant qu'orchestrateur pour d'autres fonctions. Si vous avez des flux de travail tels que des opérations par lots ou des systèmes de livraison de commandes, cela augmente les frais de gestion. En outre, il se peut qu'il ne soit pas possible de terminer l'ensemble de la logique du flux de travail et de gérer les erreurs dans le délai Lambda maximal de 15 minutes.

Au lieu de gérer cette logique dans le code de fonction, nous vous recommandons de redéfinir l'architecture de votre solution pour l'utiliser [AWS Step Functions](#) en tant qu'orchestrateur du flux de travail. Lorsque vous utilisez un flux de travail standard, vous êtes facturé pour chaque transition d'[état](#) au sein du flux de travail plutôt que pour la durée totale du flux de travail. En outre, vous pouvez placer la prise en charge des nouvelles tentatives, des conditions d'attente, des flux de travail liés aux erreurs et des [rappels](#) dans la condition d'état afin de permettre à vos fonctions Lambda de se concentrer sur la logique métier. Pour plus d'informations, consultez la section [Optimisation de vos AWS Lambda coûts — partie 2](#) du blog AWS Compute.

Passez aux fonctions basées sur la gravitation

Les fonctions Lambda alimentées par les processeurs Graviton2 de nouvelle génération sont désormais généralement disponibles. Les fonctions Graviton2, utilisant une architecture de

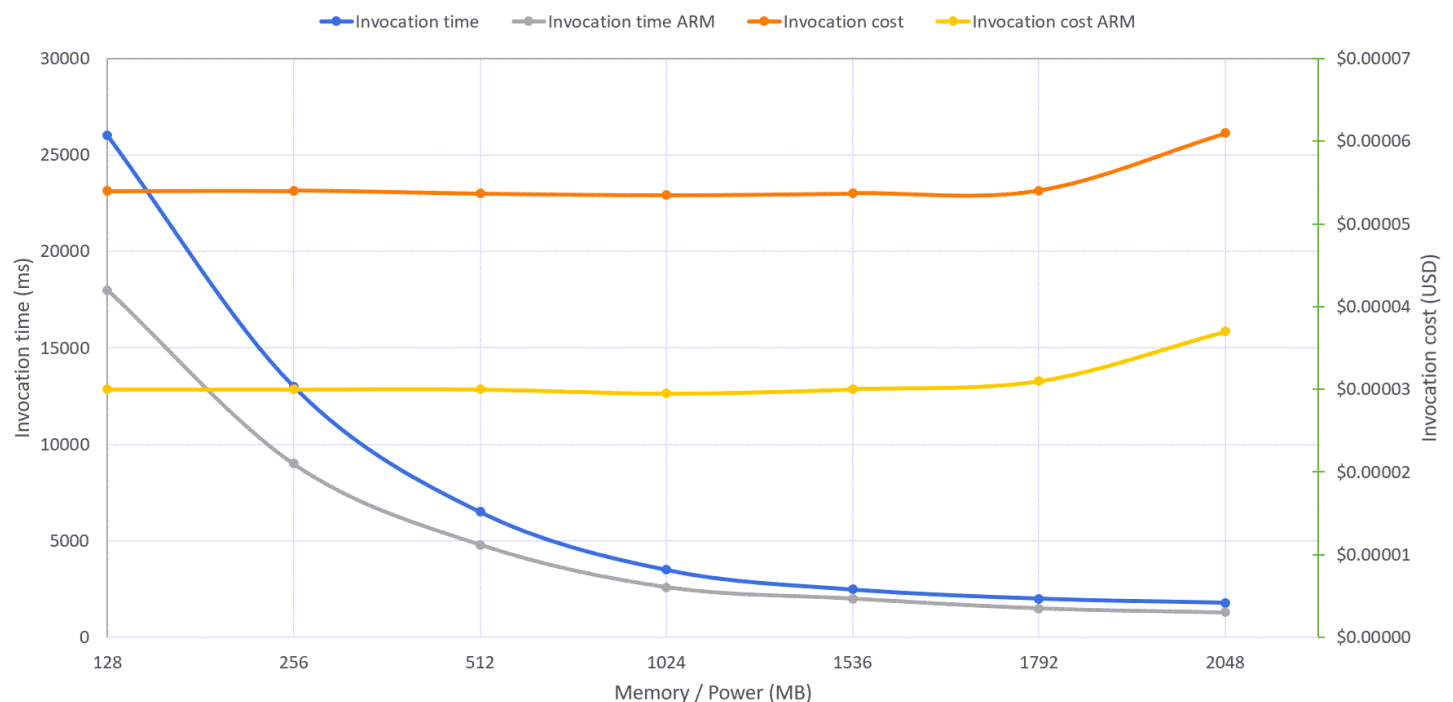
processeur basée sur ARM, sont conçues pour offrir des performances jusqu'à 19 % supérieures à un coût inférieur de 20 % pour une variété de charges de travail sans serveur. Avec une latence plus faible et de meilleures performances, les fonctions alimentées par les processeurs Graviton2 sont idéales pour alimenter les applications sans serveur critiques.

La migration vers les fonctions Lambda basées sur Graviton peut être une option rentable pour les développeurs .NET qui cherchent à optimiser leurs coûts Lambda. Les fonctions basées sur le graviton utilisent des processeurs ARM au lieu des processeurs x86 traditionnels. Cela peut permettre de réaliser d'importantes économies sans pour autant sacrifier les performances.

Bien que le passage aux fonctions basées sur Graviton présente plusieurs avantages, nous vous recommandons de prendre en compte plusieurs défis et considérations. Par exemple, les fonctions basées sur Graviton nécessitent l'utilisation d'Amazon Linux 2, qui peut ne pas être compatible avec toutes les applications .NET. En outre, il peut y avoir des problèmes de compatibilité avec des bibliothèques tierces ou des dépendances qui ne sont pas compatibles avec les processeurs ARM.

Si vous exécutez des applications .NET Framework et que vous souhaitez tirer parti du mode sans serveur avec Lambda, vous pouvez envisager de porter les applications vers le .NET moderne à l'aide de [l'assistant de portage](#) pour .NET. Cela peut vous aider à accélérer le portage des anciennes applications .NET vers le .NET moderne, permettant ainsi à l'application de s'exécuter sous Linux.

Le graphique suivant compare les résultats de l'architecture x86 et ARM/Graviton2 pour une fonction qui calcule des nombres premiers.



La fonction utilise un seul thread. La durée la plus faible pour les deux architectures est indiquée lorsque la mémoire est configurée avec 1,8 Go. Au-delà, les fonctions Lambda ont accès à plus d'un vCPU, mais dans ce cas, la fonction ne peut pas utiliser la puissance supplémentaire. Pour la même raison, les coûts sont stables avec une mémoire allant jusqu'à 1,8 Go. Avec plus de mémoire, les coûts augmentent car cette charge de travail ne présente aucun avantage supplémentaire en termes de performances. Le processeur Graviton2 fournit clairement de meilleures performances et des coûts réduits pour cette fonction gourmande en ressources informatiques.

Pour configurer votre fonction afin d'utiliser un processeur ARM avec Graviton, procédez comme suit :

1. Connectez-vous à la console [Lambda](#), AWS Management Console puis ouvrez-la.
2. Choisissez Créer une fonction.
3. Pour Nom de la fonction, entrez un nom.
4. Pour Runtime, choisissez .NET 6 (C#/ PowerShell).
5. Pour Architecture, sélectionnez arm64.
6. Effectuez les configurations supplémentaires dont vous avez besoin, puis choisissez Create function.

Ressources supplémentaires

- Les [fonctions Lambda en tant que cibles \(documentation\)](#) AWS
- [Optimisation AWS Lambda des coûts et des performances à l'aide](#) de Optimiseur de calcul AWS (AWS Compute Blog)
- [Optimisation de vos AWS Lambda coûts — Partie 1](#) (AWS Compute Blog)
- [Optimisation de vos AWS Lambda coûts — Partie 2](#) (AWS Compute Blog)
- [Création d'applications .NET sans serveur à l' AWS Lambda aide de .NET 7](#) (AWS Compute Blog)

Envisagez des bases de données spécialement conçues

Présentation de

L'un des aspects les plus coûteux de l'exécution de charges de travail basées sur Microsoft provient de l'octroi de licences pour des bases de données commerciales, telles que SQL Server. Les entreprises optent souvent pour SQL Server comme plate-forme de base de données de choix et

cette solution est ancrée dans la culture de développement de l'organisation. Les développeurs choisissent généralement un modèle basé sur SQL Server relationnel, quel que soit le cas d'utilisation. Les raisons sont les suivantes :

- L'entreprise dispose déjà de and/or licences d'instances SQL Server disponibles.
- Les équipes se sont habituées au modèle de programmation SQL grâce à l'utilisation de bibliothèques partagées et à une logique métier. ORMs
- La direction n'est pas au courant des alternatives.
- Les développeurs ne sont pas au courant des alternatives.

Les bases de données spécialement conçues peuvent s'adapter aux modèles d'accès aux données de votre cas d'utilisation. Ces bases de données sont de plus en plus adoptées par les entreprises à mesure qu'elles adoptent des architectures plus modernes (telles que les microservices) et que le champ d'application des applications individuelles se rétrécit.

Une base de données créée spécialement n'exclut pas un modèle relationnel, ni ne nécessite un modèle NoSQL (non relationnel). En fait, une base de données relationnelle est considérée comme spécialement conçue lorsqu'elle est sélectionnée en réponse aux besoins spécifiques d'une charge de travail. L'utilisation de bases de données spécialement conçues peut aider les équipes à réduire les coûts de base de données associés à leurs applications .NET, tout en bénéficiant des avantages standard du cloud, tels que l'évolutivité, la résilience et la réduction des charges lourdes indifférenciées.

Le tableau suivant présente les bases de données spécialement conçues proposées par AWS

Base de données	Type	Caractéristiques
Amazon Aurora PostgreSQL ou Amazon Aurora MySQL	Relationnel	Cas d'utilisation où les données ont une structure fixe Les bases de données relationnelles maintiennent naturellement la cohérence des données grâce aux transactions ACID
Amazon DynamoDB	Paire clé-valeur	Base de données NoSQL qui stocke les données à l'aide

Base de données	Type	Caractéristiques
		<p>d'une structure de données de table de hachage</p> <p>Stockage et extraction de données non structurées à hautes performances</p> <p>Les cas d'utilisation incluent les profils utilisateur, l'état de la session et les données du panier</p>
Amazon ElastiCache	En mémoire	<p>Base de données NoSQL hautes performances qui stocke des données non structurées en mémoire avec un temps d'accès inférieur à la milliseconde</p> <p>Utilisé pour les données éphémères fréquemment consultées, telles que les sessions utilisateur, et comme couche de mise en cache devant d'autres magasins de données plus lents</p> <p>Inclut le support pour ElastiCache (Redis OSS) et ElastiCache (Memcached)</p>
Amazon MemoryDB	Mémoire intégrée durable	Base de données spécialisée compatible Redis avec stockage durable

Base de données	Type	Caractéristiques
Amazon Timestream	Séries chronologiques	<p>Base de données conçue pour l'ingestion de données à haut débit dans l'ordre temporel</p> <p>Les cas d'utilisation incluent les applications de l'Internet des objets (IoT) et le stockage de métriques ou de données de télémétrie</p>
Amazon DocumentDB	Document	<p>Base de données NoSQL qui stocke des données sans structure prescrite ni relations imposées avec d'autres données</p> <p>Souvent utilisé pour les charges de travail intensives en lecture, telles que les catalogues de produits</p>
Amazon Neptune	Graphe	<p>Base de données NoSQL contenant à la fois des données et une représentation des connexions entre les éléments de données</p> <p>Les cas d'utilisation incluent la détection des fraudes, les moteurs de recommandation et les applications sociales</p>

Base de données	Type	Caractéristiques
Amazon Keyspaces	Colonne large	<p>Base de données distribuée haute performance basée sur Apache Cassandra</p> <p>Les cas d'utilisation incluent les applications IoT, le traitement des événements et les applications de jeu</p>

L'un des principaux moteurs de l'adoption de bases de données spécialement conçues peut être attribué à l'élimination des licences commerciales. Cependant, la fonctionnalité d'auto-scaling de bases de données telles que [DynamoDB \(y compris le mode à la demande\)](#), Aurora, [Amazon Neptune](#) et [Amazon Keyspaces](#) vous permet de prévoir de la capacité pour un cas moyen plutôt que pour une utilisation maximale. Les bases de données spécialement conçues, telles que Timestream, sont sans serveur et s'adaptent automatiquement à la demande sans aucun préapprovisionnement.

AWS propose [Babelfish pour Aurora PostgreSQL](#) si vous souhaitez utiliser une base de données relationnelle spécialement conçue et compatible avec le code source ouvert, mais que vous ne pouvez ou ne souhaitez pas apporter de modifications de code importantes à votre application. Dans certains cas, Babelfish vous permet d'utiliser un code d'accès SQL Server existant, pratiquement sans modification.

Lorsque vous choisissez une base de données relationnelle spécialement conçue pour les applications, il est important de conserver les mêmes fonctionnalités (ou fonctionnellement équivalentes) que celles dont vous avez besoin pour vos applications. Cette recommandation concerne les bases de données spécialement conçues en tant que magasin de données principal pour les applications. Des applications spécifiques (telles que la mise en cache) sont abordées dans d'autres recommandations.

Impact sur les coûts

L'adoption de bases de données spécialement conçues pour les charges de travail .NET, bien que peu susceptible d'affecter directement la consommation/coût, peut avoir une incidence directe sur le coût des services de base de données utilisés par les applications .NET. En fait, les économies de coûts peuvent être un objectif secondaire par rapport aux avantages supplémentaires liés à l'agilité, à l'évolutivité, à la résilience et à la durabilité des données.

Il n'entre pas dans le cadre de ce guide d'expliquer le processus complet de sélection d'une base de données spécialement conçue pour les applications et de refonte d'une stratégie de données afin de les utiliser efficacement. Pour plus d'informations, consultez la section [Bases de données spécialement conçues](#) dans le AWS répertoire des didacticiels.

Les tableaux suivants présentent plusieurs exemples de la manière dont le remplacement de SQL Server par une base de données spécialement conçue peut modifier les coûts des applications. Notez qu'il ne s'agit que d'estimations approximatives. Des points de référence et une optimisation des charges de travail réelles sont nécessaires pour calculer le coût de production exact.

Voici quelques estimations de base de données spécialement conçues couramment utilisées, qui incluent le calcul à la demande et un SSD de 100 Go, ainsi que des bases de données à instance unique. us-east-1 Les coûts de licence incluent la licence SQL Server et l'assurance logicielle.

Le tableau suivant indique les coûts estimés pour les exemples de bases de données commerciales.

Moteur de base de données	Modèle de licence	Type/spécifications de l'instance	AWS coût de calcul et de stockage	Coût de licence	Coût mensuel total
Édition standard de SQL Server sur Amazon EC2	Licence incluse	r6i.2xlarge (8 CPU/64 Go de RAM)	1 345,36\$	0,00\$	1 345,36\$
Édition SQL Server Enterprise sur Amazon EC2	Licence incluse	r6i.2xlarge (8 CPU/64 Go de RAM)	2 834,56\$	0,00\$	2 834,56\$
Édition standard de SQL Server sur Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 Go de RAM)	644,56\$	456,00\$	1 100,56\$

Moteur de base de données	Modèle de licence	Type/spécifications de l'instance	AWS coût de calcul et de stockage	Coût de licence	Coût mensuel total
Édition SQL Server Enterprise sur Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 Go de RAM)	644,56\$	1 750,00\$	2 394,56\$
Édition SQL Server Standard sur Amazon RDS		db.r6i.2xlarge (8 CPU/64 Go de RAM)	2 318,30\$	0,00\$	2 318,30\$
Édition SQL Server Enterprise sur Amazon RDS		db.r6i.2xlarge (8 CPU/64 Go de RAM)	3 750,56\$	0,00\$	3 750,56\$

Le tableau suivant indique les coûts estimés pour les exemples spécialement conçus.

Moteur de base de données	Type/spécifications de l'instance	AWS coût de calcul et de stockage	Coût de licence	Coût mensuel total
Amazon Aurora PostgreSQL	r6g.2xlarge (8 CPU/64 Go de RAM)	855,87\$	0,00\$	855,87\$
DynamoDB	Base provisionnée 100 WCU/400 RCU	72,00\$		72,00\$

Moteur de base de données	Type/spécifications de l'instance	AWS coût de calcul et de stockage	Coût de licence	Coût mensuel total
Amazon DocumentDB	db.r6i.2xlarge (8 CPU/64 Go de RAM)	778,60\$		778,60\$

Important

Le tableau est basé sur les coûts de licence estimés pour SQL Server with Software Assurance, au cours des trois premières années suivant l'achat. Pour l'édition standard de SQL Server : 4 100\$, pack de 2 cœurs, 3 ans. Pour l'édition Enterprise de SQL Server : 15 700\$, pack de 2 cœurs, 3 ans.

Nous vous recommandons de prendre en compte les implications financières avant d'adopter des bases de données spécialement conçues. Par exemple, le coût de mise à jour des applications afin d'utiliser une base de données spécialement conçue est lié à la complexité de l'application et de la base de données source. Assurez-vous de prendre en compte le coût total de possession lors de la planification de ce commutateur d'architecture. Cela inclut la refactorisation de vos applications, le renforcement des compétences du personnel en matière de nouvelles technologies et la planification minutieuse des performances et de la consommation prévues pour chaque charge de travail. À partir de là, vous pouvez déterminer si l'investissement vaut les économies réalisées. Dans la plupart des cas, la maintenance d'un end-of-support produit représente un risque pour la sécurité et la conformité, et le coût de sa réparation en vaut la peine et l'investissement initial.

Recommandations en matière d'optimisation des coûts

Pour les applications .NET qui accèdent à SQL Server, il existe des bibliothèques de remplacement pour les bases de données relationnelles spécialement conçues. Vous pouvez implémenter ces bibliothèques dans votre application pour remplacer les fonctionnalités similaires des applications SQL Server.

Le tableau suivant met en évidence certaines bibliothèques qui peuvent être utilisées dans de nombreux scénarios courants.

Bibliothèque	Base de données	Remplacement pour	Compatibilité avec le framework
Fournisseur principal de Npgsql Entity Framework	Amazon Aurora PostgreSQL	Fournisseur SQL Server Entity Framework Core	.NET moderne
Fournisseur Npgsql Entity Framework 6	Amazon Aurora PostgreSQL	Fournisseur SQL Server Entity Framework 6.0	.NET Framework
Npgsql (bibliothèque PostgreSQL compatible avec ADO.NET)	Amazon Aurora PostgreSQL	ADO.NET	.NET Framework/ Modern .NET
Fournisseur principal de MySQL Entity Framework	Amazon Aurora MySQL	Fournisseur SQL Server Entity Framework Core	.NET moderne
Pomelo. EntityFrameworkCore. MySql	Amazon Aurora MySQL	Fournisseur SQL Server Entity Framework Core	.NET moderne

[La connexion à Amazon Aurora PostgreSQL à l'aide de Babelfish](#) ne nécessite aucun codage spécial pour se connecter. Cependant, tout le code doit être soigneusement testé avant utilisation.

D'autres bases de données spécialement conçues disposent de bibliothèques permettant d'accéder à des bibliothèques compatibles .NET qui vous permettent d'accéder à des bases de données spécialement conçues. En voici quelques exemples :

- [Utilisation des bases de données NoSQL Amazon DynamoDB \(documentation\)](#) AWS SDK pour .NET
- [Pilote MongoDB C# \(documentation\)](#) MongoDB)
- [.NET](#) (documentation Timestream)
- [Utilisation d'un pilote client Cassandra .NET Core pour accéder à Amazon Keyspaces par programmation \(documentation Amazon Keyspaces\)](#)

- [Utilisation de .NET pour se connecter à une instance de base de données Neptune \(documentation Neptune\)](#)

Si vous migrez vers des bases de données personnalisées, vous pouvez utiliser les outils suivants AWS pour faciliter le processus de migration :

- [AWS Schema Conversion Tool \(AWS SCT\)](#) peut vous aider à transformer les schémas SQL Server en Amazon Aurora et Amazon DynamoDB.
- [AWS Database Migration Service \(AWS DMS\)](#) peut vous aider à migrer des données, ponctuellement ou de façon continue, de SQL Server vers Aurora ou DynamoDB.
- [Babelfish Compass](#) peut vous aider à vérifier la compatibilité de votre base de données SQL Server pour une utilisation avec Babelfish pour Aurora PostgreSQL.

Ressources supplémentaires

- [Conseils pour la migration de SQL Server vers Amazon Aurora PostgreSQL](#) (blog de base de données)AWS
- [Journée d'immersion dans la modernisation de l'application Babelfish](#) (AWS Workshop Studio)
- [Journée d'immersion .NET](#) (AWS Workshop Studio)
- [Commencer à utiliser Amazon Timestream](#) avec .NET () GitHub
- [Bases de données spécialement conçues pour les applications .NET modernes sur AWS](#) (présentation)AWS

Étapes suivantes

Après avoir lu ce guide, nous vous recommandons de suivre les étapes suivantes pour implémenter MACO :

1. Contactez un expert MACO. Un expert MACO peut vous aider à répondre à vos questions et à vos préoccupations. Si vous travaillez déjà avec une équipe chargée des AWS comptes, contactez l'équipe et demandez l'aide d'un expert MACO. Si vous n'avez pas d'équipe chargée du compte, contactez optimize-microsoft@amazon.com.
2. Appliquez les recommandations. Appliquez les recommandations, les meilleures pratiques et les stratégies que vous avez apprises dans ce guide et en discutant avec un expert MACO.
3. Suivez l'évolution des coûts. Étiquetez vos charges de travail et utilisez des services tels que AWS Cost Explorer et AWS Budgets pour un suivi, une surveillance et un contrôle détaillés des coûts.

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide. Pour être averti des mises à jour à venir, abonnez-vous à un [fil RSS](#).

Modification	Description	Date
Mises à jour de SQL Server	Nous avons mis à jour la section Optimiser le processeur pour les charges de travail SQL Server afin d'ajouter des informations supplémentaires sur la fonctionnalité Optimize CPU pour les EC2 instances Amazon.	22 octobre 2025
Mises à jour de SQL Server	Nous avons mis à jour la section Optimisation du processeur pour les charges de travail SQL Server .	25 octobre 2024
Mises à jour de SQL Server et Container	Nous avons ajouté le dimensionnement Optimize SQL Server en utilisant Compute Optimizer, Trusted Advisor les recommandations relatives aux charges de travail SQL Server et les sections Replatform Windows avec les sections App2Container.	29 juin 2024
Optimisation des licences SQL Server	Nous avons ajouté la licence Optimize SQL Server en utilisant la section Compute Optimizer .	22 mai 2024

Publication initiale

—

21 décembre 2023

AWS Glossaire des directives prescriptives

Les termes suivants sont couramment utilisés dans les stratégies, les guides et les modèles fournis par les directives AWS prescriptives. Pour suggérer des entrées, veuillez utiliser le lien [Faire un commentaire](#) à la fin du glossaire.

Nombres

7 R

Sept politiques de migration courantes pour transférer des applications vers le cloud. Ces politiques s'appuient sur les 5 R identifiés par Gartner en 2011 et sont les suivantes :

- **Refactor/re-architect** — Déplacez une application et modifiez son architecture en tirant pleinement parti des fonctionnalités natives du cloud pour améliorer l'agilité, les performances et l'évolutivité. Cela implique généralement le transfert du système d'exploitation et de la base de données. Exemple : migrez votre base de données Oracle sur site vers l' PostgreSQL-Compatible édition Amazon Aurora.
- **Replateformer (déplacer et remodeler)** : transférez une application vers le cloud et introduisez un certain niveau d'optimisation pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Amazon Relational Database Service (Amazon RDS) pour Oracle dans le. AWS Cloud
- **Racheter (rachat)** : optez pour un autre produit, généralement en passant d'une licence traditionnelle à un modèle SaaS. Exemple : migrez votre système de gestion de la relation client (CRM) vers Salesforce.com.
- **Réhéberger (lift and shift)** : transférez une application vers le cloud sans apporter de modifications pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Oracle sur une instance EC2 dans le. AWS Cloud
- **Relocaliser (lift and shift au niveau de l'hyperviseur)** : transférez l'infrastructure vers le cloud sans acheter de nouveau matériel, réécrire des applications ou modifier vos opérations existantes. Vous migrez des serveurs d'une plateforme sur site vers un service cloud pour la même plateforme. Exemple : migrer une Microsoft Hyper-V application vers AWS.
- **Retenir** : conservez les applications dans votre environnement source. Il peut s'agir d'applications nécessitant une refactorisation majeure, que vous souhaitez retarder, et d'applications existantes que vous souhaitez retenir, car rien ne justifie leur migration sur le plan commercial.

- Retirer : mettez hors service ou supprimez les applications dont vous n'avez plus besoin dans votre environnement source.

A

A2 (1) Agent-to-Agent

Protocole dynamique pour la collaboration agent-agent prenant en charge la délégation de tâches et le transfert d'état.

ABAC

Voir contrôle [d'accès basé sur les attributs](#).

services abstraits

Consultez la section [Services gérés](#).

ACIDE

Voir [atomicité, consistance, isolation, durabilité](#).

migration active-active

Méthode de migration de base de données dans laquelle la synchronisation des bases de données source et cible est maintenue (à l'aide d'un outil de réplication bidirectionnelle ou d'opérations d'écriture double), tandis que les deux bases de données gèrent les transactions provenant de la connexion d'applications pendant la migration. Cette méthode prend en charge la migration par petits lots contrôlés au lieu d'exiger un basculement ponctuel. Elle est plus flexible mais demande plus de travail qu'une migration [active-passive](#).

migration active-passive

Méthode de migration de base de données dans laquelle les bases de données source et cible sont synchronisées, mais seule la base de données source gère les transactions liées à la connexion des applications pendant que les données sont répliquées vers la base de données cible. La base de données cible n'accepte aucune transaction pendant la migration.

Agent

Un système d'IA capable de raisonner, de planifier et de prendre des mesures de manière autonome à l'aide d'outils pour atteindre des objectifs.

Agent Ops

Pratiques opérationnelles pour la création, le test, le déploiement et l'exécution d'agents d'IA en production à grande échelle.

fonction d'agrégation

Fonction SQL qui agit sur un groupe de lignes et calcule une valeur de retour unique pour le groupe. Des exemples de fonctions d'agrégation incluent SUM et MAX.

AI

Voir [intelligence artificielle](#).

AIOps

Voir les [opérations d'intelligence artificielle](#).

anonymisation

Processus de suppression définitive d'informations personnelles dans un ensemble de données. L'anonymisation peut contribuer à protéger la vie privée. Les données anonymisées ne sont plus considérées comme des données personnelles.

anti-motif

Solution fréquemment utilisée pour un problème récurrent lorsque la solution est contre-productive, inefficace ou moins efficace qu'une solution alternative.

contrôle des applications

Une approche de sécurité qui permet d'utiliser uniquement des applications approuvées afin de protéger un système contre les logiciels malveillants.

portefeuille d'applications

Ensemble d'informations détaillées sur chaque application utilisée par une organisation, y compris le coût de génération et de maintenance de l'application, ainsi que sa valeur métier. Ces informations sont essentielles pour [le processus de découverte et d'analyse du portefeuille](#) et permettent d'identifier et de prioriser les applications à migrer, à moderniser et à optimiser.

intelligence artificielle (IA)

Domaine de l'informatique consacré à l'utilisation des technologies de calcul pour exécuter des fonctions cognitives généralement associées aux humains, telles que l'apprentissage, la résolution de problèmes et la reconnaissance de modèles. Pour plus d'informations, veuillez consulter [Qu'est-ce que l'intelligence artificielle ?](#)

opérations d'intelligence artificielle (AIOps)

Processus consistant à utiliser des techniques de machine learning pour résoudre les problèmes opérationnels, réduire les incidents opérationnels et les interventions humaines, mais aussi améliorer la qualité du service. Pour plus d'informations sur la façon dont les AIOps sont utilisées dans la stratégie de migration AWS, veuillez consulter le [guide d'intégration des opérations](#).

chiffrement asymétrique

Algorithme de chiffrement qui utilise une paire de clés, une clé publique pour le chiffrement et une clé privée pour le déchiffrement. Vous pouvez partager la clé publique, car elle n'est pas utilisée pour le déchiffrement, mais l'accès à la clé privée doit être très restreint.

atomicité, cohérence, isolement, durabilité (ACID)

Ensemble de propriétés logicielles garantissant la validité des données et la fiabilité opérationnelle d'une base de données, même en cas d'erreur, de panne de courant ou d'autres problèmes.

contrôle d'accès par attributs (ABAC)

Pratique qui consiste à créer des autorisations détaillées en fonction des attributs de l'utilisateur, tels que le service, le poste et le nom de l'équipe. Pour plus d'informations, consultez [ABAC pour AWS](#) dans la documentation Gestion des identités et des accès AWS (IAM).

source de données faisant autorité

Emplacement où vous stockez la version principale des données, considérée comme la source d'information la plus fiable. Vous pouvez copier les données de la source de données officielle vers d'autres emplacements à des fins de traitement ou de modification des données, par exemple en les anonymisant, en les expurgant ou en les pseudonymisant.

Zone de disponibilité

Un emplacement distinct au sein d'une Région AWS réseau isolé des défaillances dans d'autres zones de disponibilité et fournissant une connectivité réseau peu coûteuse et à faible latence aux autres zones de disponibilité de la même région.

AWS Cadre d'adoption du cloud (AWS CAF)

Un cadre de directives et de meilleures pratiques visant AWS à aider les entreprises à élaborer un plan efficace pour réussir leur migration vers le cloud. AWS La CAF organise ses conseils en six domaines prioritaires appelés perspectives : les affaires, les personnes, la gouvernance, les plateformes, la sécurité et les opérations. Les perspectives d'entreprise, de personnes et de gouvernance mettent l'accent sur les compétences et les processus métier, tandis que les

perspectives relatives à la plateforme, à la sécurité et aux opérations se concentrent sur les compétences et les processus techniques. Par exemple, la perspective liée aux personnes cible les parties prenantes qui s'occupent des ressources humaines (RH), des fonctions de dotation en personnel et de la gestion des personnes. Dans cette perspective, la AWS CAF fournit des conseils pour le développement du personnel, la formation et les communications afin de préparer l'organisation à une adoption réussie du cloud. Pour plus d'informations, veuillez consulter le [site Web AWS CAF](#) et le [livre blanc AWS CAF](#).

AWS Cadre de qualification de la charge de travail (AWS WQF)

Outil qui évalue les charges de travail liées à la migration des bases de données, recommande des stratégies de migration et fournit des estimations de travail. AWS Le WQF est inclus avec AWS Schema Conversion Tool (AWS SCT). Il analyse les schémas de base de données et les objets de code, le code d'application, les dépendances et les caractéristiques de performance, et fournit des rapports d'évaluation.

B

mauvais bot

Un [bot](#) destiné à perturber ou à nuire à des individus ou à des organisations.

BCP

Consultez la section [Planification de la continuité des activités](#).

graphique de comportement

Vue unifiée et interactive des comportements des ressources et des interactions au fil du temps. Vous pouvez utiliser un graphique de comportement avec Amazon Detective pour examiner les tentatives de connexion infructueuses, les appels d'API suspects et les actions similaires. Pour plus d'informations, veuillez consulter [Data in a behavior graph](#) dans la documentation Detective.

système de poids fort

Système qui stocke d'abord l'octet le plus significatif. Voir aussi [endianité](#).

classification binaire

Processus qui prédit un résultat binaire (l'une des deux classes possibles). Par exemple, votre modèle de machine learning peut avoir besoin de prévoir des problèmes tels que « Cet e-mail est-il du spam ou non ? » ou « Ce produit est-il un livre ou une voiture ? ».

filtre de Bloom

Structure de données probabiliste et efficace en termes de mémoire qui est utilisée pour tester si un élément fait partie d'un ensemble.

blue/green déploiement

Stratégie de déploiement dans laquelle vous créez deux environnements distincts mais identiques. Vous exécutez la version actuelle de l'application dans un environnement (bleu) et la nouvelle version de l'application dans l'autre environnement (vert). Cette stratégie vous permet de revenir rapidement en arrière avec un impact minimal.

bot

Application logicielle qui exécute des tâches automatisées sur Internet et simule l'activité ou l'interaction humaine. Certains robots sont utiles ou bénéfiques, comme les robots d'exploration Web qui indexent des informations sur Internet. D'autres robots, connus sous le nom de mauvais robots, sont destinés à perturber ou à nuire à des individus ou à des organisations.

botnet

Réseaux de [robots](#) infectés par des [logiciels malveillants](#) et contrôlés par une seule entité, connue sous le nom d'herder ou d'opérateur de bots. Les botnets sont le mécanisme le plus connu pour faire évoluer les bots et leur impact.

branche

Zone contenue d'un référentiel de code. La première branche créée dans un référentiel est la branche principale. Vous pouvez créer une branche à partir d'une branche existante, puis développer des fonctionnalités ou corriger des bogues dans la nouvelle branche. Une branche que vous créez pour générer une fonctionnalité est communément appelée branche de fonctionnalités. Lorsque la fonctionnalité est prête à être publiée, vous fusionnez à nouveau la branche de fonctionnalités dans la branche principale. Pour plus d'informations, consultez [À propos des branches](#) (GitHub documentation).

accès par brise-vitre

Dans des circonstances exceptionnelles et par le biais d'un processus approuvé, il s'agit d'un moyen rapide pour un utilisateur d'accéder à un accès auquel Compte AWS il n'est généralement pas autorisé. Pour plus d'informations, consultez l'indicateur [Mettre en œuvre des procédures permettant de briser le verre](#) dans le AWS Well-Architected guide.

stratégie existante (brownfield)

L'infrastructure existante de votre environnement. Lorsque vous adoptez une stratégie existante pour une architecture système, vous concevez l'architecture en fonction des contraintes des systèmes et de l'infrastructure actuels. Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et [greenfield](#) (inédites).

cache de tampon

Zone de mémoire dans laquelle sont stockées les données les plus fréquemment consultées.

capacité métier

Ce que fait une entreprise pour générer de la valeur (par exemple, les ventes, le service client ou le marketing). Les architectures de microservices et les décisions de développement peuvent être dictées par les capacités métier. Pour plus d'informations, veuillez consulter la section [Organisation en fonction des capacités métier](#) du livre blanc [Exécution de microservices conteneurisés sur AWS](#).

planification de la continuité des activités (BCP)

Plan qui tient compte de l'impact potentiel d'un événement perturbateur, tel qu'une migration à grande échelle, sur les opérations, et qui permet à une entreprise de reprendre ses activités rapidement.

C

CAF

Voir le [cadre d'adoption du AWS cloud](#).

déploiement de Canary

Diffusion lente et progressive d'une version pour les utilisateurs finaux. Lorsque vous êtes sûr, vous déployez la nouvelle version et remplacez la version actuelle dans son intégralité.

CCoE

Voir [le Centre d'excellence du cloud](#).

CDC

Consultez la section [Capture des données de modification](#).

capture des données de modification (CDC)

Processus de suivi des modifications apportées à une source de données, telle qu'une table de base de données, et d'enregistrement des métadonnées relatives à ces modifications. Vous pouvez utiliser la CDC à diverses fins, telles que l'audit ou la réplication des modifications dans un système cible afin de maintenir la synchronisation.

ingénierie du chaos

Introduire intentionnellement des défaillances ou des événements perturbateurs pour tester la résilience d'un système. Vous pouvez utiliser [AWS Fault Injection Service \(AWS FIS\)](#) pour effectuer des expériences qui stressent vos AWS charges de travail et évaluer leur réponse.

CI/CD

Découvrez [l'intégration continue et la livraison continue](#).

classification

Processus de catégorisation qui permet de générer des prédictions. Les modèles de ML pour les problèmes de classification prédisent une valeur discrète. Les valeurs discrètes se distinguent toujours les unes des autres. Par exemple, un modèle peut avoir besoin d'évaluer la présence ou non d'une voiture sur une image.

Développeur citoyen

Un utilisateur professionnel qui crée des applications d'intelligence artificielle à l'aide de plateformes sans code/low code sans compétences techniques spécialisées.

chiffrement côté client

Chiffrement des données localement, avant que la cible ne les Service AWS reçoive.

Centre d'excellence cloud (CCoE)

Une équipe multidisciplinaire qui dirige les efforts d'adoption du cloud au sein d'une organisation, notamment en développant les bonnes pratiques en matière de cloud, en mobilisant des ressources, en établissant des délais de migration et en guidant l'organisation dans le cadre de transformations à grande échelle. Pour plus d'informations, consultez les [articles du CCoE](#) sur le blog de stratégie AWS Cloud d'entreprise.

cloud computing

Technologie cloud généralement utilisée pour le stockage de données à distance et la gestion des appareils IoT. Le cloud computing est généralement associé à la technologie [informatique de pointe](#).

modèle d'exploitation du cloud

Dans une organisation informatique, modèle d'exploitation utilisé pour créer, faire évoluer et optimiser un ou plusieurs environnements cloud. Pour plus d'informations, consultez la section [Création de votre modèle d'exploitation cloud](#).

étapes d'adoption du cloud

Les quatre phases que les entreprises traversent généralement lorsqu'elles migrent vers AWS Cloud :

- **Projet** : exécution de quelques projets liés au cloud à des fins de preuve de concept et d'apprentissage
- **Base** : réaliser des investissements fondamentaux pour mettre à l'échelle l'adoption du cloud (par exemple, en créant une zone de destination, en définissant un CCoE ou en établissant un modèle opérationnel)
- **Migration** : migration d'applications individuelles
- **Re-invention** — Optimisation des produits et services et innovation dans le cloud

Ces étapes ont été définies par Stephen Orban dans le billet de blog [The Journey Toward Cloud-First & the Stages of Adoption](#) publié sur le blog AWS Cloud Enterprise Strategy. Pour plus d'informations sur leur lien avec la stratégie de AWS migration, consultez le [guide de préparation à la migration](#).

CMDB

Consultez la base de [données de gestion des configurations](#).

référentiel de code

Emplacement où le code source et d'autres ressources, comme la documentation, les exemples et les scripts, sont stockés et mis à jour par le biais de processus de contrôle de version. Les référentiels cloud courants incluent GitHub ou Bitbucket Cloud. Chaque version du code est appelée branche. Dans une structure de microservice, chaque référentiel est consacré à une seule fonctionnalité. Un CI/CD pipeline unique peut utiliser plusieurs référentiels.

cache passif

Cache tampon vide, mal rempli ou contenant des données obsolètes ou non pertinentes. Cela affecte les performances, car l'instance de base de données doit lire à partir de la mémoire principale ou du disque, ce qui est plus lent que la lecture à partir du cache tampon.

données gelées

Données rarement consultées et généralement historiques. Lorsque vous interrogez ce type de données, les requêtes lentes sont généralement acceptables. Le transfert de ces données vers des niveaux ou classes de stockage moins performants et moins coûteux peut réduire les coûts.

vision par ordinateur (CV)

Domaine de l'[IA](#) qui utilise l'apprentissage automatique pour analyser et extraire des informations à partir de formats visuels tels que des images numériques et des vidéos. Par exemple, Amazon SageMaker AI fournit des algorithmes de traitement d'image pour les CV.

dérive de configuration

Pour une charge de travail, une modification de configuration par rapport à l'état attendu. Cela peut entraîner une non-conformité de la charge de travail, et cela est généralement progressif et involontaire.

base de données de gestion des configurations (CMDB)

Référentiel qui stocke et gère les informations relatives à une base de données et à son environnement informatique, y compris les composants matériels et logiciels ainsi que leurs configurations. Vous utilisez généralement les données d'une CMDB lors de la phase de découverte et d'analyse du portefeuille de la migration.

pack de conformité

Ensemble de AWS Config règles et d'actions correctives que vous pouvez assembler pour personnaliser vos contrôles de conformité et de sécurité. Vous pouvez déployer un pack de conformité en tant qu'entité unique dans une région Compte AWS et, ou au sein d'une organisation, à l'aide d'un modèle YAML. Pour plus d'informations, consultez la section [Packs de conformité](#) dans la AWS Config documentation.

intégration continue et livraison continue (CI/CD)

Processus d'automatisation des étapes de source, de construction, de test, de préparation et de production du processus de publication du logiciel. CI/CD est communément décrit comme un pipeline. CI/CD peut vous aider à automatiser les processus, à améliorer la productivité, à améliorer la qualité du code et à accélérer les livraisons. Pour plus d'informations, veuillez consulter [Avantages de la livraison continue](#). CD peut également signifier déploiement continu. Pour plus d'informations, veuillez consulter [Livraison continue et déploiement continu](#).

CV

Voir [vision par ordinateur](#).

D

données au repos

Données stationnaires dans votre réseau, telles que les données stockées.

classification des données

Processus permettant d'identifier et de catégoriser les données de votre réseau en fonction de leur sévérité et de leur sensibilité. Il s'agit d'un élément essentiel de toute stratégie de gestion des risques de cybersécurité, car il vous aide à déterminer les contrôles de protection et de conservation appropriés pour les données. La classification des données est une composante du pilier de sécurité du AWS Well-Architected cadre. Pour plus d'informations, veuillez consulter [Classification des données](#).

dérive des données

Une variation significative entre les données de production et les données utilisées pour entraîner un modèle ML, ou une modification significative des données d'entrée au fil du temps. La dérive des données peut réduire la qualité, la précision et l'équité globales des prédictions des modèles ML.

données en transit

Données qui circulent activement sur votre réseau, par exemple entre les ressources du réseau.

maillage de données

Un cadre architectural qui fournit une propriété des données distribuée et décentralisée avec une gestion et une gouvernance centralisées.

minimisation des données

Le principe de collecte et de traitement des seules données strictement nécessaires. La pratique de la minimisation des données AWS Cloud peut réduire les risques liés à la confidentialité, les coûts et l'empreinte carbone de vos analyses.

périmètre de données

Ensemble de garde-fous préventifs dans votre AWS environnement qui permettent de garantir que seules les identités fiables accèdent aux ressources fiables des réseaux attendus. Pour plus d'informations, voir [Création d'un périmètre de données sur AWS](#).

prétraitement des données

Pour transformer les données brutes en un format facile à analyser par votre modèle de ML. Le prétraitement des données peut impliquer la suppression de certaines colonnes ou lignes et le traitement des valeurs manquantes, incohérentes ou en double.

provenance des données

Le processus de suivi de l'origine et de l'historique des données tout au long de leur cycle de vie, par exemple la manière dont les données ont été générées, transmises et stockées.

sujet des données

Personne dont les données sont collectées et traitées.

entrepôt des données

Un système de gestion des données qui prend en charge les informations commerciales, telles que les analyses. Les entrepôts de données contiennent généralement de grandes quantités de données historiques et sont généralement utilisés pour les requêtes et les analyses.

langage de définition de base de données (DDL)

Instructions ou commandes permettant de créer ou de modifier la structure des tables et des objets dans une base de données.

langage de manipulation de base de données (DML)

Instructions ou commandes permettant de modifier (insérer, mettre à jour et supprimer) des informations dans une base de données.

DDL

Voir [langage de définition de base de données](#) de données.

ensemble profond

Sert à combiner plusieurs modèles de deep learning à des fins de prédiction. Vous pouvez utiliser des ensembles profonds pour obtenir une prévision plus précise ou pour estimer l'incertitude des prédictions.

deep learning

Un sous-champ de ML qui utilise plusieurs couches de réseaux neuronaux artificiels pour identifier le mappage entre les données d'entrée et les variables cibles d'intérêt.

défense en profondeur

Approche de la sécurité de l'information dans laquelle une série de mécanismes et de contrôles de sécurité sont judicieusement répartis sur l'ensemble d'un réseau informatique afin de protéger la confidentialité, l'intégrité et la disponibilité du réseau et des données qu'il contient. Lorsque vous adoptez cette stratégie AWS, vous ajoutez plusieurs contrôles à différentes couches de la AWS Organizations structure afin de sécuriser les ressources. Par exemple, une approche de défense approfondie peut combiner l'authentification multifactorielle, la segmentation du réseau et le chiffrement.

administrateur délégué

Dans AWS Organizations, un service compatible peut enregistrer un compte AWS membre pour administrer les comptes de l'organisation et gérer les autorisations pour ce service. Ce compte est appelé administrateur délégué pour ce service. Pour plus d'informations et une liste des services compatibles, veuillez consulter la rubrique [Services qui fonctionnent avec AWS Organizations](#) dans la documentation AWS Organizations .

déploiement

Processus de mise à disposition d'une application, de nouvelles fonctionnalités ou de corrections de code dans l'environnement cible. Le déploiement implique la mise en œuvre de modifications dans une base de code, puis la génération et l'exécution de cette base de code dans les environnements de l'application.

environnement de développement

Voir [environnement](#).

contrôle de détection

Contrôle de sécurité conçu pour détecter, journaliser et alerter après la survenue d'un événement. Ces contrôles constituent une deuxième ligne de défense et vous alertent en cas d'événements de sécurité qui ont contourné les contrôles préventifs en place. Pour plus d'informations, veuillez consulter la rubrique [Contrôles de détection](#) dans Implementing security controls on AWS.

cartographie de la chaîne de valeur du développement (DVSM)

Processus utilisé pour identifier et hiérarchiser les contraintes qui nuisent à la rapidité et à la qualité du cycle de vie du développement logiciel. DVSM étend le processus de cartographie de la chaîne de valeur initialement conçu pour les pratiques de production allégée. Il met l'accent sur les étapes et les équipes nécessaires pour créer et transférer de la valeur tout au long du processus de développement logiciel.

jumeau numérique

Représentation virtuelle d'un système réel, tel qu'un bâtiment, une usine, un équipement industriel ou une ligne de production. Les jumeaux numériques prennent en charge la maintenance prédictive, la surveillance à distance et l'optimisation de la production.

tableau des dimensions

Dans un [schéma en étoile](#), table plus petite contenant les attributs de données relatifs aux données quantitatives d'une table de faits. Les attributs des tables de dimensions sont généralement des champs de texte ou des nombres discrets qui se comportent comme du texte. Ces attributs sont couramment utilisés pour la contrainte des requêtes, le filtrage et l'étiquetage des ensembles de résultats.

catastrophe

Un événement qui empêche une charge de travail ou un système d'atteindre ses objectifs commerciaux sur son site de déploiement principal. Ces événements peuvent être des catastrophes naturelles, des défaillances techniques ou le résultat d'actions humaines, telles qu'une mauvaise configuration involontaire ou une attaque de logiciel malveillant.

reprise après sinistre (DR)

La stratégie et le processus que vous utilisez pour minimiser les temps d'arrêt et les pertes de données causés par un [sinistre](#). Pour plus d'informations, consultez la section [Reprise après sinistre des charges de travail sur AWS : Restauration dans le cloud](#) dans le AWS Well-Architected Framework.

DML

Voir [langage de manipulation de base](#) de données.

conception axée sur le domaine

Approche visant à développer un système logiciel complexe en connectant ses composants à des domaines évolutifs, ou objectifs métier essentiels, que sert chaque composant. Ce concept a été introduit par Eric Evans dans son livre, *Domain-Driven Design : Tackling Complexity in the Heart of Software* (Boston : Addison-Wesley Professional, 2003). Pour plus d'informations sur la manière dont vous pouvez utiliser la conception axée sur le domaine avec le modèle Strangler Fig, consultez la section [Modernisation incrémentielle des anciens services Web ASP.NET Microsoft \(ASMX\) à l'aide de conteneurs et d'Amazon API Gateway](#).

DR

Consultez la section [Reprise après sinistre](#).

détection de dérive

Suivi des écarts par rapport à une configuration de référence. Par exemple, vous pouvez l'utiliser AWS CloudFormation pour [détecter la dérive des ressources du système](#) ou AWS Control Tower pour [détecter les modifications de votre zone d'atterrissage](#) susceptibles d'affecter le respect des exigences de gouvernance.

DVSM

Voir la [cartographie de la chaîne de valeur du développement](#).

E

EDA

Voir [analyse exploratoire des données](#).

EDI

Voir échange [de données informatisé](#).

informatique de périphérie

Technologie qui augmente la puissance de calcul des appareils intelligents en périphérie d'un réseau IoT. Comparé au [cloud computing, l'informatique](#) de pointe peut réduire la latence des communications et améliorer le temps de réponse.

échange de données informatisé (EDI)

L'échange automatique de documents commerciaux entre les organisations. Pour plus d'informations, voir [Qu'est-ce que l'échange de données informatisé ?](#)

chiffrement

Processus informatique qui transforme des données en texte clair, lisibles par l'homme, en texte chiffré.

clé de chiffrement

Chaîne cryptographique de bits aléatoires générée par un algorithme cryptographique. La longueur des clés peut varier, et chaque clé est conçue pour être imprévisible et unique.

endianisme

Ordre dans lequel les octets sont stockés dans la mémoire de l'ordinateur. Big-endian les systèmes stockent d'abord l'octet le plus significatif. Little-endian les systèmes stockent d'abord l'octet le moins significatif.

point de terminaison

Voir [point de terminaison de service](#).

service de point de terminaison

Service que vous pouvez héberger sur un cloud privé virtuel (VPC) pour le partager avec d'autres utilisateurs. Vous pouvez créer un service de point de terminaison avec AWS PrivateLink et accorder des autorisations à d'autres principaux Comptes AWS ou à Gestion des identités et des accès AWS (IAM) principaux. Ces comptes ou principaux peuvent se connecter à votre service de point de terminaison de manière privée en créant des points de terminaison d'un VPC d'interface. Pour plus d'informations, veuillez consulter [Création d'un service de point de terminaison](#) dans la documentation Amazon Virtual Private Cloud (Amazon VPC).

planification des ressources d'entreprise (ERP)

Système qui automatise et gère les principaux processus métier (tels que la comptabilité, le [MES](#) et la gestion de projet) pour une entreprise.

chiffrement d'enveloppe

Processus de chiffrement d'une clé de chiffrement à l'aide d'une autre clé de chiffrement. Pour plus d'informations, consultez la section [Chiffrement des enveloppes](#) dans la documentation AWS Key Management Service (AWS KMS).

environnement

Instance d'une application en cours d'exécution. Les types d'environnement les plus courants dans le cloud computing sont les suivants :

- Environnement de développement : instance d'une application en cours d'exécution à laquelle seule l'équipe principale chargée de la maintenance de l'application peut accéder. Les environnements de développement sont utilisés pour tester les modifications avant de les promouvoir dans les environnements supérieurs. Ce type d'environnement est parfois appelé environnement de test.
- Environnements inférieurs : tous les environnements de développement d'une application, tels que ceux utilisés pour les générations et les tests initiaux.

- Environnement de production : instance d'une application en cours d'exécution à laquelle les utilisateurs finaux peuvent accéder. Dans un CI/CD pipeline, l'environnement de production est le dernier environnement de déploiement.
- Environnements supérieurs : tous les environnements accessibles aux utilisateurs autres que l'équipe de développement principale. Ils peuvent inclure un environnement de production, des environnements de préproduction et des environnements pour les tests d'acceptation par les utilisateurs.

épopée

Dans les méthodologies agiles, catégories fonctionnelles qui aident à organiser et à prioriser votre travail. Les épopées fournissent une description détaillée des exigences et des tâches d'implémentation. Par exemple, les points forts de la AWS CAF en matière de sécurité incluent la gestion des identités et des accès, les contrôles de détection, la sécurité des infrastructures, la protection des données et la réponse aux incidents. Pour plus d'informations sur les épopées dans la stratégie de migration AWS , veuillez consulter le [guide d'implémentation du programme](#).

ERP

Voir [Planification des ressources d'entreprise](#).

analyse exploratoire des données (EDA)

Processus d'analyse d'un jeu de données pour comprendre ses principales caractéristiques. Vous collectez ou agrégez des données, puis vous effectuez des enquêtes initiales pour trouver des modèles, détecter des anomalies et vérifier les hypothèses. L'EDA est réalisée en calculant des statistiques récapitulatives et en créant des visualisations de données.

F

tableau des faits

La table centrale dans un [schéma en étoile](#). Il stocke des données quantitatives sur les opérations commerciales. Généralement, une table de faits contient deux types de colonnes : celles qui contiennent des mesures et celles qui contiennent une clé étrangère pour une table de dimensions.

échouer rapidement

Une philosophie qui utilise des tests fréquents et progressifs pour réduire le cycle de vie du développement. C'est un élément essentiel d'une approche agile.

limite d'isolation des défauts

Dans le AWS Cloud, une limite telle qu'une zone de disponibilité Région AWS, un plan de contrôle ou un plan de données qui limite l'effet d'une panne et contribue à améliorer la résilience des charges de travail. Pour plus d'informations, consultez la section [Limites d'isolation des AWS pannes](#).

branche de fonctionnalités

Voir [la succursale](#).

fonctionnalités

Les données d'entrée que vous utilisez pour faire une prédiction. Par exemple, dans un contexte de fabrication, les fonctionnalités peuvent être des images capturées périodiquement à partir de la ligne de fabrication.

importance des fonctionnalités

Le niveau d'importance d'une fonctionnalité pour les prédictions d'un modèle. Il s'exprime généralement sous la forme d'un score numérique qui peut être calculé à l'aide de différentes techniques, telles que la méthode Shapley Additive Explanations (SHAP) et les gradients intégrés. Pour plus d'informations, voir [Interprétabilité du modèle d'apprentissage automatique avec AWS](#).

transformation de fonctionnalité

Optimiser les données pour le processus de ML, notamment en enrichissant les données avec des sources supplémentaires, en mettant à l'échelle les valeurs ou en extrayant plusieurs ensembles d'informations à partir d'un seul champ de données. Cela permet au modèle de ML de tirer parti des données. Par exemple, si vous décomposez la date « 2021-05-27 00:15:37 » en « 2021 », « mai », « jeudi » et « 15 », vous pouvez aider l'algorithme d'apprentissage à apprendre des modèles nuancés associés à différents composants de données.

invitation en quelques coups

Fournir à un [LLM](#) un petit nombre d'exemples illustrant la tâche et le résultat souhaité avant de lui demander d'effectuer une tâche similaire. Cette technique est une application de l'apprentissage contextuel, dans le cadre de laquelle les modèles apprennent à partir d'exemples (prises de vue) intégrés dans des instructions. Few-shot l'envoi d'instructions peut être efficace pour les tâches qui nécessitent un formatage, un raisonnement ou une connaissance du domaine spécifiques. Voir également l'[invite Zero-Shot](#).

FGAC

Découvrez le [contrôle d'accès détaillé](#).

contrôle d'accès détaillé (FGAC)

Utilisation de plusieurs conditions pour autoriser ou refuser une demande d'accès.

migration instantanée (flash-cut)

Méthode de migration de base de données qui utilise la réplication continue des données par [le biais de la capture des données de modification](#) afin de migrer les données dans les plus brefs délais, au lieu d'utiliser une approche progressive. L'objectif est de réduire au maximum les temps d'arrêt.

FM

Voir le [modèle de fondation](#).

modèle de fondation (FM)

Un vaste réseau neuronal d'apprentissage profond qui s'entraîne sur des ensembles de données massifs de données généralisées et non étiquetées. Les FM sont capables d'effectuer une grande variété de tâches générales, telles que la compréhension du langage, la génération de texte et d'images et la conversation en langage naturel. Pour plus d'informations, voir [Que sont les modèles de base ?](#)

Passerelle FM

Un intermédiaire centralisé qui contrôle et normalise l'accès aux [modèles de base](#). Également connue sous le nom de passerelle LLM.

G

IA générative

Sous-ensemble de modèles d'[IA](#) qui ont été entraînés sur de grandes quantités de données et qui peuvent utiliser une simple invite textuelle pour créer de nouveaux contenus et artefacts, tels que des images, des vidéos, du texte et du son. Pour plus d'informations, consultez [Qu'est-ce que l'IA générative](#).

blocage géographique

Voir les [restrictions géographiques](#).

restrictions géographiques (blocage géographique)

Sur Amazon CloudFront, option permettant d'empêcher les utilisateurs de certains pays d'accéder aux distributions de contenu. Vous pouvez utiliser une liste d'autorisation ou une liste de blocage

pour spécifier les pays approuvés et interdits. Pour plus d'informations, consultez [la section Restreindre la distribution géographique de votre contenu](#) dans la CloudFront documentation.

Flux de travail Gitflow

Approche dans laquelle les environnements inférieurs et supérieurs utilisent différentes branches dans un référentiel de code source. Le flux de travail Gitflow est considéré comme existant, et le [flux de travail basé sur les troncs](#) est l'approche moderne préférée.

image dorée

Un instantané d'un système ou d'un logiciel utilisé comme modèle pour déployer de nouvelles instances de ce système ou logiciel. Par exemple, dans le secteur de la fabrication, une image dorée peut être utilisée pour fournir des logiciels sur plusieurs appareils et contribue à améliorer la vitesse, l'évolutivité et la productivité des opérations de fabrication des appareils.

stratégie inédite

L'absence d'infrastructures existantes dans un nouvel environnement. Lorsque vous adoptez une stratégie inédite pour une architecture système, vous pouvez sélectionner toutes les nouvelles technologies sans restriction de compatibilité avec l'infrastructure existante, également appelée [brownfield](#). Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et greenfield (inédites).

barrière de protection

Règle de haut niveau qui permet de régir les ressources, les politiques et la conformité au sein des unités d'organisation (UO). Les barrières de protection préventives appliquent des politiques pour garantir l'alignement sur les normes de conformité. Elles sont mises en œuvre à l'aide de politiques de contrôle des services et de limites des autorisations IAM. Les barrières de protection de détection détectent les violations des politiques et les problèmes de conformité, et génèrent des alertes pour y remédier. Ils sont implémentés à l'aide d'Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, d'Amazon Inspector et de AWS Lambda contrôles personnalisés.

rambardes (AI)

Des mécanismes de sécurité qui filtrent, valident et limitent les entrées et sorties des [agents](#) afin de garantir un comportement responsable et sûr de l'IA.

H

HA

Découvrez [la haute disponibilité](#).

migration de base de données hétérogène

Migration de votre base de données source vers une base de données cible qui utilise un moteur de base de données différent (par exemple, Oracle vers Amazon Aurora). La migration hétérogène fait généralement partie d'un effort de réarchitecture, et la conversion du schéma peut s'avérer une tâche complexe. [AWS propose AWS SCT](#) qui facilite les conversions de schémas.

haute disponibilité (HA)

Capacité d'une charge de travail à fonctionner en continu, sans intervention, en cas de difficultés ou de catastrophes. Les systèmes HA sont conçus pour basculer automatiquement, fournir constamment des performances de haute qualité et gérer différentes charges et défaillances avec un impact minimal sur les performances.

modernisation de l'historien

Approche utilisée pour moderniser et mettre à niveau les systèmes de technologie opérationnelle (OT) afin de mieux répondre aux besoins de l'industrie manufacturière. Un historien est un type de base de données utilisé pour collecter et stocker des données provenant de diverses sources dans une usine.

données de rétention

Partie de données historiques étiquetées qui n'est pas divulguée dans un ensemble de données utilisé pour entraîner un modèle d'[apprentissage automatique](#). Vous pouvez utiliser les données de blocage pour évaluer les performances du modèle en comparant les prévisions du modèle aux données de blocage.

humain dans la boucle (HiTL)

Un modèle de flux de travail dans lequel l'exécution des [agents](#) s'arrête pour examen et approbation par l'homme aux points de décision critiques.

migration de base de données homogène

Migration de votre base de données source vers une base de données cible qui partage le même moteur de base de données (par exemple, Microsoft SQL Server vers Amazon RDS for SQL Server). La migration homogène s'inscrit généralement dans le cadre d'un effort de

réhébergement ou de replateforme. Vous pouvez utiliser les utilitaires de base de données natifs pour migrer le schéma.

données chaudes

Données fréquemment consultées, telles que les données en temps réel ou les données translationnelles récentes. Ces données nécessitent généralement un niveau ou une classe de stockage à hautes performances pour fournir des réponses rapides aux requêtes.

correctif

Solution d'urgence à un problème critique dans un environnement de production. En raison de son urgence, un correctif est généralement créé en dehors du flux de travail de DevOps publication habituel.

période de soins intensifs

Immédiatement après le basculement, période pendant laquelle une équipe de migration gère et surveille les applications migrées dans le cloud afin de résoudre les problèmes éventuels. En règle générale, cette période dure de 1 à 4 jours. À la fin de la période de soins intensifs, l'équipe de migration transfère généralement la responsabilité des applications à l'équipe des opérations cloud.

I

laC

Considérez [l'infrastructure comme un code](#).

politique basée sur l'identité

Politique attachée à un ou plusieurs principaux IAM qui définit leurs autorisations au sein de l'AWS Cloud environnement.

application inactive

Application dont l'utilisation moyenne du processeur et de la mémoire se situe entre 5 et 20 % sur une période de 90 jours. Dans un projet de migration, il est courant de retirer ces applications ou de les retenir sur site.

IIoT

Voir [Internet industriel des objets](#).

infrastructure immuable

Modèle qui déploie une nouvelle infrastructure pour les charges de travail de production au lieu de mettre à jour, d'appliquer des correctifs ou de modifier l'infrastructure existante. Les infrastructures immuables sont intrinsèquement plus cohérentes, fiables et prévisibles que les infrastructures [mutables](#). Pour plus d'informations, consultez les meilleures pratiques de [déploiement à l'aide d'une infrastructure immuable](#) dans le AWS Well-Architected Framework.

VPC entrant (d'entrée)

Dans une architecture AWS multi-comptes, un VPC qui accepte, inspecte et achemine les connexions réseau depuis l'extérieur d'une application. L'[architecture de référence de sécurité AWS](#) recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

migration incrémentielle

Stratégie de basculement dans le cadre de laquelle vous migrez votre application par petites parties au lieu d'effectuer un basculement complet unique. Par exemple, il se peut que vous ne transfériez que quelques microservices ou utilisateurs vers le nouveau système dans un premier temps. Après avoir vérifié que tout fonctionne correctement, vous pouvez transférer progressivement des microservices ou des utilisateurs supplémentaires jusqu'à ce que vous puissiez mettre hors service votre système hérité. Cette stratégie réduit les risques associés aux migrations de grande ampleur.

Industry 4.0

Terme introduit par [Klaus Schwab](#) en 2016 pour désigner la modernisation des processus de fabrication grâce aux avancées en matière de connectivité, de données en temps réel, d'automatisation, d'analyse et. AI/ML

infrastructure

Ensemble des ressources et des actifs contenus dans l'environnement d'une application.

infrastructure en tant que code (IaC)

Processus de mise en service et de gestion de l'infrastructure d'une application via un ensemble de fichiers de configuration. IaC est conçue pour vous aider à centraliser la gestion de l'infrastructure, à normaliser les ressources et à mettre à l'échelle rapidement afin que les nouveaux environnements soient reproductibles, fiables et cohérents.

internet industriel des objets (IIoT)

L'utilisation de capteurs et d'appareils connectés à Internet dans les secteurs industriels tels que la fabrication, l'énergie, l'automobile, les soins de santé, les sciences de la vie et l'agriculture. Pour plus d'informations, veuillez consulter [Building an industrial Internet of Things \(IIoT\) digital transformation strategy](#).

VPC d'inspection

Dans une architecture AWS multi-comptes, un VPC centralisé qui gère les inspections du trafic réseau entre les VPC (identiques ou Régions AWS différents), Internet et les réseaux sur site. L'[architecture de référence de sécurité AWS](#) recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

Internet des objets (IoT)

Réseau d'objets physiques connectés dotés de capteurs ou de processeurs intégrés qui communiquent avec d'autres appareils et systèmes via Internet ou via un réseau de communication local. Pour plus d'informations, veuillez consulter la section [Qu'est-ce que l'IoT ?](#).

interprétabilité

Caractéristique d'un modèle de machine learning qui décrit dans quelle mesure un être humain peut comprendre comment les prédictions du modèle dépendent de ses entrées. Pour plus d'informations, voir [Interprétabilité du modèle d'apprentissage automatique avec AWS](#).

IoT

Voir [Internet des objets](#).

Bibliothèque d'informations informatiques (ITIL)

Ensemble de bonnes pratiques pour proposer des services informatiques et les aligner sur les exigences métier. L'ITIL constitue la base de l'ITSM.

gestion des services informatiques (ITSM)

Activités associées à la conception, à la mise en œuvre, à la gestion et à la prise en charge de services informatiques d'une organisation. Pour plus d'informations sur l'intégration des opérations cloud aux outils ITSM, veuillez consulter le [guide d'intégration des opérations](#).

ITIL

Consultez la [bibliothèque d'informations informatiques](#).

ITSM

Voir [Gestion des services informatiques](#).

L

contrôle d'accès basé sur des étiquettes (LBAC)

Une implémentation du contrôle d'accès obligatoire (MAC) dans laquelle une valeur d'étiquette de sécurité est explicitement attribuée aux utilisateurs et aux données elles-mêmes. L'intersection entre l'étiquette de sécurité utilisateur et l'étiquette de sécurité des données détermine les lignes et les colonnes visibles par l'utilisateur.

zone de destination

Une zone d'atterrissage est un AWS environnement multi-comptes bien conçu, évolutif et sécurisé. Il s'agit d'un point de départ à partir duquel vos entreprises peuvent rapidement lancer et déployer des charges de travail et des applications en toute confiance dans leur environnement de sécurité et d'infrastructure. Pour plus d'informations sur les zones de destination, veuillez consulter [Setting up a secure and scalable multi-account AWS environment](#).

grand modèle de langage (LLM)

Un modèle d'[intelligence artificielle basé](#) sur le deep learning qui est préentraîné sur une grande quantité de données. Un LLM peut effectuer plusieurs tâches, telles que répondre à des questions, résumer des documents, traduire du texte dans d'autres langues et compléter des phrases. Pour plus d'informations, voir [Que sont les LLM](#).

migration de grande envergure

Migration de 300 serveurs ou plus.

LBAC

Voir contrôle d'[accès basé sur des étiquettes](#).

principe de moindre privilège

Bonne pratique de sécurité qui consiste à accorder les autorisations minimales nécessaires à l'exécution d'une tâche. Pour plus d'informations, veuillez consulter la rubrique [Accorder les autorisations de moindre privilège](#) dans la documentation IAM.

lift and shift

Voir [7 Rs](#).

système de poids faible

Système qui stocke d'abord l'octet le moins significatif. Voir aussi [endianité](#).

LLM

Voir le [grand modèle de langage](#).

environnements inférieurs

Voir [environnement](#).

M

machine learning (ML)

Type d'intelligence artificielle qui utilise des algorithmes et des techniques pour la reconnaissance et l'apprentissage de modèles. Le ML analyse et apprend à partir de données enregistrées, telles que les données de l'Internet des objets (IoT), pour générer un modèle statistique basé sur des modèles. Pour plus d'informations, veuillez consulter [Machine Learning](#).

branche principale

Voir [la succursale](#).

malware

Logiciel conçu pour compromettre la sécurité ou la confidentialité de l'ordinateur. Les logiciels malveillants peuvent perturber les systèmes informatiques, divulguer des informations sensibles ou obtenir un accès non autorisé. Parmi les malwares, on peut citer les virus, les vers, les rançongiciels, les chevaux de Troie, les logiciels espions et les enregistreurs de frappe.

services gérés

Services AWS pour lequel AWS fonctionnent la couche d'infrastructure, le système d'exploitation et les plateformes, et vous accédez aux points de terminaison pour stocker et récupérer des données. Amazon Simple Storage Service (Amazon S3) et Amazon DynamoDB sont des exemples de services gérés. Ils sont également connus sous le nom de services abstraits.

système d'exécution de la fabrication (MES)

Un système logiciel pour le suivi, la surveillance, la documentation et le contrôle des processus de production qui convertissent les matières premières en produits finis dans l'atelier.

MAP

Voir [Migration Acceleration Program](#).

MCP

Voir [Model Context Protocol](#).

Protocole de contexte du modèle (MCP)

Protocole sans état pour la communication entre [un agent](#) et un [outil](#).

serveur MCP

Service qui expose un ou plusieurs [outils](#) via le [protocole Model Context](#).

mécanisme

Processus complet au cours duquel vous créez un outil, favorisez son adoption, puis inspectez les résultats afin de procéder aux ajustements nécessaires. Un mécanisme est un cycle qui se renforce et s'améliore au fur et à mesure de son fonctionnement. Pour plus d'informations, voir [Création de mécanismes](#) dans le AWS Well-Architected cadre.

compte membre

Tous, à l'exception des Comptes AWS exception du compte de gestion, qui font partie d'une organisation dans AWS Organizations. Un compte ne peut être membre que d'une seule organisation à la fois.

MAILLES

Voir le [système d'exécution de la fabrication](#).

Transport télémétrique en file d'attente de messages (MQTT)

[Un protocole de communication léger de machine à machine \(M2M\), basé sur le publish/subscribe modèle, pour les appareils IoT aux ressources limitées.](#)

microservice

Petit service indépendant qui communique via des API bien définies et qui est généralement détenu par de petites équipes autonomes. Par exemple, un système d'assurance peut inclure des microservices qui mappent à des capacités métier, telles que les ventes ou le marketing, ou à des sous-domaines, tels que les achats, les réclamations ou l'analytique. Les avantages des microservices incluent l'agilité, la flexibilité de la mise à l'échelle, la facilité de déploiement, la réutilisation du code et la résilience. Pour plus d'informations, consultez la section [Intégration de microservices à l'aide de services AWS sans serveur](#).

architecture de microservices

Approche de création d'une application avec des composants indépendants qui exécutent chaque processus d'application en tant que microservice. Ces microservices communiquent via une interface bien définie à l'aide d'API légères. Chaque microservice de cette architecture peut être mis à jour, déployé et mis à l'échelle pour répondre à la demande de fonctions spécifiques d'une application. Pour plus d'informations, consultez la section [Implémentation de microservices sur AWS](#).

Programme d'accélération des migrations (MAP)

Un AWS programme qui fournit un support de conseil, des formations et des services pour aider les entreprises à établir une base opérationnelle solide pour passer au cloud, et pour aider à compenser le coût initial des migrations. MAP inclut une méthodologie de migration pour exécuter les migrations héritées de manière méthodique, ainsi qu'un ensemble d'outils pour automatiser et accélérer les scénarios de migration courants.

migration à grande échelle

Processus consistant à transférer la majeure partie du portefeuille d'applications vers le cloud par vagues, un plus grand nombre d'applications étant déplacées plus rapidement à chaque vague. Cette phase utilise les bonnes pratiques et les enseignements tirés des phases précédentes pour implémenter une usine de migration d'équipes, d'outils et de processus en vue de rationaliser la migration des charges de travail grâce à l'automatisation et à la livraison agile. Il s'agit de la troisième phase de la [stratégie de migration AWS](#).

usine de migration

Cross-functional des équipes qui rationalisent la migration des charges de travail grâce à des approches automatisées et agiles. Les équipes de Migration Factory comprennent généralement des responsables des opérations, des analystes commerciaux et des propriétaires, des ingénieurs de migration, des développeurs et DevOps des professionnels travaillant dans le cadre de sprints. Entre 20 et 50 % du portefeuille d'applications d'entreprise est constitué de modèles répétés qui peuvent être optimisés par une approche d'usine. Pour plus d'informations, veuillez consulter la rubrique [discussion of migration factories](#) et le [guide Cloud Migration Factory](#) dans cet ensemble de contenus.

métadonnées de migration

Informations relatives à l'application et au serveur nécessaires pour finaliser la migration. Chaque modèle de migration nécessite un ensemble de métadonnées de migration différent. Les

exemples de métadonnées de migration incluent le sous-réseau cible, le groupe de sécurité et le AWS compte.

modèle de migration

Tâche de migration reproductible qui détaille la stratégie de migration, la destination de la migration et l'application ou le service de migration utilisé. Exemple : réorganisez la migration vers Amazon EC2 AWS avec le service de migration d'applications.

Évaluation du portefeuille de migration (MPA)

Outil en ligne qui fournit des informations pour valider l'analyse de rentabilisation en faveur de la migration vers le. AWS Cloud La MPA propose une évaluation détaillée du portefeuille (dimensionnement approprié des serveurs, tarification, comparaison du coût total de possession, analyse des coûts de migration), ainsi que la planification de la migration (analyse et collecte des données d'applications, regroupement des applications, priorisation des migrations et planification des vagues). L'[outil MPA](#) (connexion requise) est disponible gratuitement pour tous les AWS consultants et consultants APN Partner.

Évaluation de la préparation à la migration (MRA)

Processus qui consiste à obtenir des informations sur l'état de préparation d'une organisation au cloud, à identifier les forces et les faiblesses et à élaborer un plan d'action pour combler les lacunes identifiées, à l'aide du AWS CAF. Pour plus d'informations, veuillez consulter le [guide de préparation à la migration](#). La MRA est la première phase de la [stratégie de migration AWS](#).

stratégie de migration

L'approche utilisée pour migrer une charge de travail vers le AWS Cloud. Pour plus d'informations, reportez-vous aux [7 R](#) de ce glossaire et à [Mobiliser votre organisation pour accélérer les migrations à grande échelle](#).

ML

Voir [apprentissage automatique](#).

modernisation

Transformation d'une application obsolète (héritée ou monolithique) et de son infrastructure en un système agile, élastique et hautement disponible dans le cloud afin de réduire les coûts, de gagner en efficacité et de tirer parti des innovations. Pour plus d'informations, consultez [la section Stratégie de modernisation des applications dans le AWS Cloud](#).

évaluation de la préparation à la modernisation

Évaluation qui permet de déterminer si les applications d'une organisation sont prêtes à être modernisées, d'identifier les avantages, les risques et les dépendances, et qui détermine dans quelle mesure l'organisation peut prendre en charge l'état futur de ces applications. Le résultat de l'évaluation est un plan de l'architecture cible, une feuille de route détaillant les phases de développement et les étapes du processus de modernisation, ainsi qu'un plan d'action pour combler les lacunes identifiées. Pour plus d'informations, consultez la section [Évaluation de l'état de préparation à la modernisation des applications dans le AWS Cloud](#).

applications monolithiques (monolithes)

Applications qui s'exécutent en tant que service unique avec des processus étroitement couplés. Les applications monolithiques ont plusieurs inconvénients. Si une fonctionnalité de l'application connaît un pic de demande, l'architecture entière doit être mise à l'échelle. L'ajout ou l'amélioration des fonctionnalités d'une application monolithique devient également plus complexe lorsque la base de code s'élargit. Pour résoudre ces problèmes, vous pouvez utiliser une architecture de microservices. Pour plus d'informations, veuillez consulter [Decomposing monoliths into microservices](#).

MPA

Voir [Évaluation du portefeuille de migration](#).

MQTT

Voir [Message Queuing Telemetry Transport](#).

classification multi-classes

Processus qui permet de générer des prédictions pour plusieurs classes (prédiction d'un résultat parmi plus de deux). Par exemple, un modèle de ML peut demander « Ce produit est-il un livre, une voiture ou un téléphone ? » ou « Quelle catégorie de produits intéresse le plus ce client ? ».

infrastructure mutable

Modèle qui met à jour et modifie l'infrastructure existante pour les charges de travail de production. Pour améliorer la cohérence, la fiabilité et la prévisibilité, le AWS Well-Architected Framework recommande l'utilisation d'une [infrastructure immuable](#) comme meilleure pratique.

O

OAC

Voir [Contrôle d'accès à l'origine](#).

OAI

Voir [l'identité d'accès à l'origine](#).

OCM

Voir [gestion du changement organisationnel](#).

migration hors ligne

Méthode de migration dans laquelle la charge de travail source est supprimée au cours du processus de migration. Cette méthode implique un temps d'arrêt prolongé et est généralement utilisée pour de petites charges de travail non critiques.

OI

Consultez la section [Intégration des opérations](#).

OLA

Voir l'accord [au niveau opérationnel](#).

migration en ligne

Méthode de migration dans laquelle la charge de travail source est copiée sur le système cible sans être mise hors ligne. Les applications connectées à la charge de travail peuvent continuer à fonctionner pendant la migration. Cette méthode implique un temps d'arrêt nul ou minimal et est généralement utilisée pour les charges de travail de production critiques.

OPC-UA

Voir [Open Process Communications - Architecture unifiée](#).

Communications par processus ouvert - Architecture unifiée (OPC-UA)

Protocole de communication machine à machine (M2M) pour l'automatisation industrielle. OPC-UA fournit une norme d'interopérabilité avec des schémas de chiffrement, d'authentification et d'autorisation des données.

accord au niveau opérationnel (OLA)

Accord qui précise ce que les groupes informatiques fonctionnels s'engagent à fournir les uns aux autres, afin de prendre en charge un contrat de niveau de service (SLA).

examen de l'état de préparation opérationnelle (ORR)

Une liste de questions et de bonnes pratiques associées qui vous aident à comprendre, à évaluer, à prévenir ou à réduire l'ampleur des incidents et des défaillances possibles. Pour plus d'informations, voir [Examens de l'état de préparation opérationnelle \(ORR\)](#) dans le AWS Well-Architected cadre.

technologie opérationnelle (OT)

Systèmes matériels et logiciels qui fonctionnent avec l'environnement physique pour contrôler les opérations, les équipements et les infrastructures industriels. Dans le secteur manufacturier, l'intégration des systèmes OT et des technologies de l'information (IT) est au cœur des transformations de [l'industrie 4.0](#).

intégration des opérations (OI)

Processus de modernisation des opérations dans le cloud, qui implique la planification de la préparation, l'automatisation et l'intégration. Pour en savoir plus, veuillez consulter le [guide d'intégration des opérations](#).

journal de suivi d'organisation

Un parcours créé par AWS CloudTrail qui enregistre tous les événements pour tous les membres Comptes AWS d'une organisation dans AWS Organizations. Ce journal de suivi est créé dans chaque Compte AWS qui fait partie de l'organisation et suit l'activité de chaque compte. Pour plus d'informations, consultez [la section Création d'un suivi pour une organisation](#) dans la CloudTrail documentation.

gestion du changement organisationnel (OCM)

Cadre pour gérer les transformations métier majeures et perturbatrices du point de vue des personnes, de la culture et du leadership. L'OCM aide les organisations à se préparer et à effectuer la transition vers de nouveaux systèmes et de nouvelles politiques en accélérant l'adoption des changements, en abordant les problèmes de transition et en favorisant des changements culturels et organisationnels. Dans la stratégie de AWS migration, ce cadre est appelé accélération du personnel, en raison de la rapidité du changement requise dans les projets d'adoption du cloud. Pour plus d'informations, veuillez consulter le [guide OCM](#).

contrôle d'accès d'origine (OAC)

Dans CloudFront, une option améliorée pour restreindre l'accès afin de sécuriser votre contenu Amazon Simple Storage Service (Amazon S3). OAC prend en charge tous les compartiments S3 dans leur ensemble Régions AWS, le chiffrement côté serveur avec AWS KMS (SSE-KMS) et les DELETE requêtes dynamiques PUT adressées au compartiment S3.

identité d'accès d'origine (OAI)

Dans CloudFront, une option permettant de restreindre l'accès afin de sécuriser votre contenu Amazon S3. Lorsque vous utilisez OAI, il CloudFront crée un principal auprès duquel Amazon S3 peut s'authentifier. Les principaux authentifiés ne peuvent accéder au contenu d'un compartiment S3 que par le biais d'une distribution spécifique CloudFront . Voir également [OAC](#), qui fournit un contrôle d'accès plus précis et amélioré.

ORR

Voir l'[examen de l'état de préparation opérationnelle](#).

DE

Voir [technologie opérationnelle](#).

VPC sortant (de sortie)

Dans une architecture AWS multi-comptes, un VPC qui gère les connexions réseau initiées depuis une application. L'[architecture de référence de sécuritéAWS](#) recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

P

limite des autorisations

Politique de gestion IAM attachée aux principaux IAM pour définir les autorisations maximales que peut avoir l'utilisateur ou le rôle. Pour plus d'informations, veuillez consulter la rubrique [Limites des autorisations](#) dans la documentation IAM.

informations personnelles identifiables (PII)

Informations qui, lorsqu'elles sont consultées directement ou associées à d'autres données connexes, peuvent être utilisées pour déduire raisonnablement l'identité d'une personne. Les

exemples d'informations personnelles incluent les noms, les adresses et les informations de contact.

PII

Voir les [informations personnelles identifiables](#).

manuel stratégique

Ensemble d'étapes prédéfinies qui capturent le travail associé aux migrations, comme la fourniture de fonctions d'opérations de base dans le cloud. Un manuel stratégique peut revêtir la forme de scripts, de runbooks automatisés ou d'un résumé des processus ou des étapes nécessaires au fonctionnement de votre environnement modernisé.

PLC

Voir [contrôleur logique programmable](#).

PLM

Consultez la section [Gestion du cycle de vie des produits](#).

policy

Objet capable de définir les autorisations (voir la [politique basée sur l'identité](#)), de spécifier les conditions d'accès (voir la [politique basée sur les ressources](#)) ou de définir les autorisations maximales pour tous les comptes d'une organisation dans AWS Organizations (voir la politique de contrôle des [services](#)).

persistance polyglotte

Choix indépendant de la technologie de stockage de données d'un microservice en fonction des modèles d'accès aux données et d'autres exigences. Si vos microservices utilisent la même technologie de stockage de données, ils peuvent rencontrer des difficultés d'implémentation ou présenter des performances médiocres. Les microservices sont plus faciles à mettre en œuvre, atteignent de meilleures performances, ainsi qu'une meilleure capacité de mise à l'échelle s'ils utilisent l'entrepôt de données le mieux adapté à leurs besoins.

évaluation du portefeuille

Processus de découverte, d'analyse et de priorisation du portefeuille d'applications afin de planifier la migration. Pour plus d'informations, veuillez consulter [Evaluating migration readiness](#).

predicate

Une condition de requête qui renvoie `true` ou `false`, généralement située dans une `WHERE` clause.

prédicat pushdown

Technique d'optimisation des requêtes de base de données qui filtre les données de la requête avant le transfert. Cela réduit la quantité de données qui doivent être extraites et traitées à partir de la base de données relationnelle et améliore les performances des requêtes.

contrôle préventif

Contrôle de sécurité conçu pour empêcher qu'un événement ne se produise. Ces contrôles constituent une première ligne de défense pour empêcher tout accès non autorisé ou toute modification indésirable de votre réseau. Pour plus d'informations, veuillez consulter [Preventative controls](#) dans *Implementing security controls on AWS*.

principal

Entité capable d'effectuer AWS des actions et d'accéder à des ressources. Cette entité est généralement un utilisateur root pour un Compte AWS rôle IAM ou un utilisateur. Pour plus d'informations, veuillez consulter la rubrique Principal dans [Termes et concepts relatifs aux rôles](#), dans la documentation IAM.

confidentialité dès la conception

Une approche d'ingénierie système qui prend en compte la confidentialité tout au long du processus de développement.

zones hébergées privées

Conteneur qui contient des informations concernant la façon dont vous souhaitez qu'Amazon Route 53 réponde aux requêtes DNS pour un domaine et ses sous-domaines dans un ou plusieurs VPC. Pour plus d'informations, veuillez consulter [Working with private hosted zones](#) dans la documentation Route 53.

contrôle proactif

[Contrôle de sécurité](#) conçu pour empêcher le déploiement de ressources non conformes. Ces contrôles analysent les ressources avant qu'elles ne soient provisionnées. Si la ressource n'est pas conforme au contrôle, elle n'est pas provisionnée. Pour plus d'informations, consultez le [guide de référence sur les contrôles](#) dans la AWS Control Tower documentation et consultez la section [Contrôles proactifs dans Implémentation](#) des contrôles de sécurité sur AWS.

gestion du cycle de vie des produits (PLM)

Gestion des données et des processus d'un produit tout au long de son cycle de vie, depuis la conception, le développement et le lancement, en passant par la croissance et la maturité, jusqu'au déclin et au retrait.

environnement de production

Voir [environnement](#).

contrôleur logique programmable (PLC)

Dans le secteur manufacturier, un ordinateur hautement fiable et adaptable qui surveille les machines et automatise les processus de fabrication.

chaînage rapide

Utiliser le résultat d'une invite [LLM](#) comme entrée pour l'invite suivante afin de générer de meilleures réponses. Cette technique est utilisée pour décomposer une tâche complexe en sous-tâches ou pour affiner ou développer de manière itérative une réponse préliminaire. Cela permet d'améliorer la précision et la pertinence des réponses d'un modèle et permet d'obtenir des résultats plus précis et personnalisés.

pseudonymisation

Processus de remplacement des identifiants personnels dans un ensemble de données par des valeurs fictives. La pseudonymisation peut contribuer à protéger la vie privée. Les données pseudonymisées sont toujours considérées comme des données personnelles.

publish/subscribe (pub/sub)

Modèle qui permet des communications asynchrones entre les microservices afin d'améliorer l'évolutivité et la réactivité. Par exemple, dans un [MES](#) basé sur des microservices, un microservice peut publier des messages d'événements sur un canal auquel d'autres microservices peuvent s'abonner. Le système peut ajouter de nouveaux microservices sans modifier le service de publication.

Q

plan de requête

Série d'étapes, telles que des instructions, utilisées pour accéder aux données d'un système de base de données relationnelle SQL.

régression du plan de requêtes

Le cas où un optimiseur de service de base de données choisit un plan moins optimal qu'avant une modification donnée de l'environnement de base de données. Cela peut être dû à des

changements en termes de statistiques, de contraintes, de paramètres d'environnement, de liaisons de paramètres de requêtes et de mises à jour du moteur de base de données.

R

Matrice RACI

Voir [responsable, responsable, consulté, informé \(RACI\)](#).

RAG

Voir [Retrieval Augmented Generation](#).

rançongiciel

Logiciel malveillant conçu pour bloquer l'accès à un système informatique ou à des données jusqu'à ce qu'un paiement soit effectué.

Matrice RASCI

Voir [responsable, responsable, consulté, informé \(RACI\)](#).

RCAC

Voir [contrôle d'accès aux lignes et aux colonnes](#).

réplica en lecture

Copie d'une base de données utilisée en lecture seule. Vous pouvez acheminer les requêtes vers le réplica de lecture pour réduire la charge sur votre base de données principale.

réarchitecte

Voir [7 Rs](#).

objectif de point de récupération (RPO)

Durée maximale acceptable depuis le dernier point de récupération des données. Il détermine ce qui est considéré comme étant une perte de données acceptable entre le dernier point de reprise et l'interruption du service.

objectif de temps de récupération (RTO)

Le délai maximum acceptable entre l'interruption du service et le rétablissement du service.

refactoriser

Voir [7 Rs](#).

Région

Un ensemble de AWS ressources dans une zone géographique. Chacun Région AWS est isolé et indépendant des autres pour garantir tolérance aux pannes, stabilité et résilience. Pour plus d'informations, voir [Spécifier ce que Régions AWS votre compte peut utiliser](#).

régression

Technique de ML qui prédit une valeur numérique. Par exemple, pour résoudre le problème « Quel sera le prix de vente de cette maison ? », un modèle de ML pourrait utiliser un modèle de régression linéaire pour prédire le prix de vente d'une maison sur la base de faits connus à son sujet (par exemple, la superficie en mètres carrés).

réhéberger

Voir [7 Rs](#).

version

Dans un processus de déploiement, action visant à promouvoir les modifications apportées à un environnement de production.

déplacer

Voir [7 Rs](#).

replateforme

Voir [7 Rs](#).

rachat

Voir [7 Rs](#).

résilience

La capacité d'une application à résister aux perturbations ou à s'en remettre. [La haute disponibilité et la reprise après sinistre](#) sont des considérations courantes lors de la planification de la résilience dans le AWS Cloud. Pour plus d'informations, consultez la section [AWS Cloud Résilience](#).

politique basée sur les ressources

Politique attachée à une ressource, comme un compartiment Amazon S3, un point de terminaison ou une clé de chiffrement. Ce type de politique précise les principaux auxquels l'accès est autorisé, les actions prises en charge et toutes les autres conditions qui doivent être remplies.

matrice responsable, redevable, consulté et informé (RACI)

Une matrice qui définit les rôles et les responsabilités de toutes les parties impliquées dans les activités de migration et les opérations cloud. Le nom de la matrice est dérivé des types de responsabilité définis dans la matrice : responsable (R), responsable (A), consulté (C) et informé (I). Le type de support (S) est facultatif. Si vous incluez le support, la matrice est appelée matrice RASCI, et si vous l'excluez, elle est appelée matrice RACI.

contrôle réactif

Contrôle de sécurité conçu pour permettre de remédier aux événements indésirables ou aux écarts par rapport à votre référence de sécurité. Pour plus d'informations, veuillez consulter la rubrique [Responsive controls](#) dans *Implementing security controls on AWS*.

retain

Voir [7 Rs](#).

se retirer

Voir [7 Rs](#).

Génération augmentée de récupération (RAG)

Technologie d'[IA générative](#) dans laquelle un [LLM](#) fait référence à une source de données faisant autorité qui se trouve en dehors de ses sources de données de formation avant de générer une réponse. Par exemple, un modèle RAG peut effectuer une recherche sémantique dans la base de connaissances ou dans les données personnalisées d'une organisation. Pour plus d'informations, voir [Qu'est-ce que RAG ?](#)

rotation

Processus de mise à jour périodique d'un [secret](#) pour empêcher un attaquant d'accéder aux informations d'identification.

contrôle d'accès aux lignes et aux colonnes (RCAC)

Utilisation d'expressions SQL simples et flexibles dotées de règles d'accès définies. Le RCAC comprend des autorisations de ligne et des masques de colonnes.

RPO

Voir l'[objectif du point de récupération](#).

RTO

Voir l'[objectif en matière de temps de rétablissement](#).

runbook

Ensemble de procédures manuelles ou automatisées nécessaires à l'exécution d'une tâche spécifique. Elles visent généralement à rationaliser les opérations ou les procédures répétitives présentant des taux d'erreur élevés.

S

SAML 2.0

Un standard ouvert utilisé par de nombreux fournisseurs d'identité (IdPs). Cette fonctionnalité permet l'authentification unique fédérée (SSO), afin que les utilisateurs puissent se connecter AWS Management Console ou appeler les opérations de l' AWS API sans que vous ayez à créer un utilisateur dans IAM pour tous les membres de votre organisation. Pour plus d'informations sur la fédération SAML 2.0, veuillez consulter [À propos de la fédération SAML 2.0](#) dans la documentation IAM.

SCADA

Voir [Contrôle de supervision et acquisition de données](#).

SCP

Voir la [politique de contrôle des services](#).

secret

Dans AWS Secrets Manager des informations confidentielles ou restreintes, telles qu'un mot de passe ou des informations d'identification utilisateur, que vous stockez sous forme cryptée. Il comprend la valeur secrète et ses métadonnées. La valeur secrète peut être binaire, une chaîne unique ou plusieurs chaînes. Pour plus d'informations, voir [Que contient le secret d'un Secrets Manager ?](#) dans la documentation de Secrets Manager.

sécurité dès la conception

Une approche d'ingénierie système qui prend en compte la sécurité tout au long du processus de développement.

contrôle de sécurité

Barrière de protection technique ou administrative qui empêche, détecte ou réduit la capacité d'un assaillant d'exploiter une vulnérabilité de sécurité. Il existe quatre principaux types de contrôles de sécurité : [préventifs](#), [détectifs](#), [réactifs](#) et [proactifs](#).

renforcement de la sécurité

Processus qui consiste à réduire la surface d'attaque pour la rendre plus résistante aux attaques. Cela peut inclure des actions telles que la suppression de ressources qui ne sont plus requises, la mise en œuvre des bonnes pratiques de sécurité consistant à accorder le moindre privilège ou la désactivation de fonctionnalités inutiles dans les fichiers de configuration.

système de gestion des informations et des événements de sécurité (SIEM)

Outils et services qui associent les systèmes de gestion des informations de sécurité (SIM) et de gestion des événements de sécurité (SEM). Un système SIEM collecte, surveille et analyse les données provenant de serveurs, de réseaux, d'appareils et d'autres sources afin de détecter les menaces et les failles de sécurité, mais aussi de générer des alertes.

automatisation des réponses de sécurité

Action prédéfinie et programmée conçue pour répondre automatiquement à un événement de sécurité ou y remédier. Ces automatisations servent de contrôles de sécurité [détectifs ou réactifs](#) qui vous aident à mettre en œuvre les meilleures pratiques en matière AWS de sécurité. Parmi les actions de réponse automatique, citons la modification d'un groupe de sécurité VPC, l'application de correctifs à une instance Amazon EC2 ou la rotation des informations d'identification.

chiffrement côté serveur

Chiffrement des données à destination, par celui Service AWS qui les reçoit.

Politique de contrôle des services (SCP)

Politique qui propose un contrôle centralisé des autorisations pour tous les comptes d'une organisation dans AWS Organizations. Les SCP définissent des barrières de protection ou des limites aux actions qu'un administrateur peut déléguer à des utilisateurs ou à des rôles. Vous pouvez utiliser les SCP comme listes d'autorisation ou de refus, pour indiquer les services ou les actions autorisés ou interdits. Pour plus d'informations, consultez la section [Politiques de contrôle des services](#) dans la AWS Organizations documentation.

point de terminaison du service

URL du point d'entrée pour un Service AWS. Pour vous connecter par programmation au service cible, vous pouvez utiliser un point de terminaison. Pour plus d'informations, veuillez consulter la rubrique [Service AWS endpoints](#) dans Références générales AWS.

contrat de niveau de service (SLA)

Accord qui précise ce qu'une équipe informatique promet de fournir à ses clients, comme le temps de disponibilité et les performances des services.

indicateur de niveau de service (SLI)

Mesure d'un aspect des performances d'un service, tel que son taux d'erreur, sa disponibilité ou son débit.

objectif de niveau de service (SLO)

Mesure cible qui représente l'état d'un service, tel que mesuré par un indicateur de [niveau de service](#).

modèle de responsabilité partagée

Un modèle décrivant la responsabilité que vous partagez en matière AWS de sécurité et de conformité dans le cloud. AWS est responsable de la sécurité du cloud, alors que vous êtes responsable de la sécurité dans le cloud. Pour de plus amples informations, veuillez consulter [Modèle de responsabilité partagée](#).

IA de l'ombre

Applications d'[IA](#) non autorisées créées ou utilisées en dehors des canaux régis au sein d'une organisation.

SIEM

Consultez les [informations de sécurité et le système de gestion des événements](#).

point de défaillance unique (SPOF)

Défaillance d'un seul composant critique d'une application susceptible de perturber le système.

SLA

Voir le contrat [de niveau de service](#).

SLI

Voir l'indicateur de [niveau de service](#).

SLO

Voir l'objectif de [niveau de service](#).

modèle split-and-seed

Modèle permettant de mettre à l'échelle et d'accélérer les projets de modernisation. Au fur et à mesure que les nouvelles fonctionnalités et les nouvelles versions de produits sont définies, l'équipe principale se divise pour créer des équipes de produit. Cela permet de mettre à l'échelle

les capacités et les services de votre organisation, d'améliorer la productivité des développeurs et de favoriser une innovation rapide. Pour plus d'informations, consultez la section [Approche progressive de la modernisation des applications dans](#) le. AWS Cloud

SPOF

Voir [point de défaillance unique](#).

schéma en étoile

Structure organisationnelle de base de données qui utilise une grande table de faits pour stocker les données transactionnelles ou mesurées et utilise une ou plusieurs tables dimensionnelles plus petites pour stocker les attributs des données. Cette structure est conçue pour être utilisée dans un [entrepôt de données](#) ou à des fins de business intelligence.

modèle de figuier étrangleur

Approche de modernisation des systèmes monolithiques en réécrivant et en remplaçant progressivement les fonctionnalités du système jusqu'à ce que le système hérité puisse être mis hors service. Ce modèle utilise l'analogie d'un figuier de vigne qui se développe dans un arbre existant et qui finit par supplanter son hôte. Le schéma a été [présenté par Martin Fowler](#) comme un moyen de gérer les risques lors de la réécriture de systèmes monolithiques. Pour un exemple d'application de ce modèle, consultez la section [Modernisation progressive des anciens services Web Microsoft ASP.NET \(ASMX\) à l'aide de conteneurs et d'Amazon API Gateway](#).

sous-réseau

Plage d'adresses IP dans votre VPC. Un sous-réseau doit se trouver dans une seule zone de disponibilité.

contrôle de supervision et acquisition de données (SCADA)

Dans le secteur manufacturier, un système qui utilise du matériel et des logiciels pour surveiller les actifs physiques et les opérations de production.

chiffrement symétrique

Algorithme de chiffrement qui utilise la même clé pour chiffrer et déchiffrer les données.

tests synthétiques

Tester un système de manière à simuler les interactions des utilisateurs afin de détecter les problèmes potentiels ou de surveiller les performances. Vous pouvez utiliser [Amazon CloudWatch Synthetics](#) pour créer ces tests.

invite du système

Technique permettant de fournir un contexte, des instructions ou des directives à un [LLM](#) afin d'orienter son comportement. Les instructions du système aident à définir le contexte et à établir des règles pour les interactions avec les utilisateurs.

T

tags

Key-value des paires qui agissent comme des métadonnées pour organiser vos AWS ressources. Les balises peuvent vous aider à gérer, identifier, organiser, rechercher et filtrer des ressources. Pour plus d'informations, veuillez consulter la rubrique [Balisage de vos AWS ressources](#).

variable cible

La valeur que vous essayez de prédire dans le cadre du ML supervisé. Elle est également qualifiée de variable de résultat. Par exemple, dans un environnement de fabrication, la variable cible peut être un défaut du produit.

liste de tâches

Outil utilisé pour suivre les progrès dans un runbook. Liste de tâches qui contient une vue d'ensemble du runbook et une liste des tâches générales à effectuer. Pour chaque tâche générale, elle inclut le temps estimé nécessaire, le propriétaire et l'avancement.

environnement de test

Voir [environnement](#).

entraînement

Pour fournir des données à partir desquelles votre modèle de ML peut apprendre. Les données d'entraînement doivent contenir la bonne réponse. L'algorithme d'apprentissage identifie des modèles dans les données d'entraînement, qui mettent en correspondance les attributs des données d'entrée avec la cible (la réponse que vous souhaitez prédire). Il fournit un modèle de ML qui capture ces modèles. Vous pouvez alors utiliser le modèle de ML pour obtenir des prédictions sur de nouvelles données pour lesquelles vous ne connaissez pas la cible.

outil

Fonction ou API qu'un [agent](#) peut invoquer pour effectuer des opérations dans des systèmes externes.

passerelle de transit

Hub de transit de réseau que vous pouvez utiliser pour relier vos VPC et vos réseaux sur site. Pour plus d'informations, voir [Qu'est-ce qu'une passerelle de transit](#) dans la AWS Transit Gateway documentation.

flux de travail basé sur jonction

Approche selon laquelle les développeurs génèrent et testent des fonctionnalités localement dans une branche de fonctionnalités, puis fusionnent ces modifications dans la branche principale. La branche principale est ensuite intégrée aux environnements de développement, de préproduction et de production, de manière séquentielle.

accès sécurisé

Accorder des autorisations à un service que vous spécifiez pour effectuer des tâches au sein de votre organisation AWS Organizations et dans ses comptes en votre nom. Le service de confiance crée un rôle lié au service dans chaque compte, lorsque ce rôle est nécessaire, pour effectuer des tâches de gestion à votre place. Pour plus d'informations, consultez la section [Utilisation AWS Organizations avec d'autres AWS services](#) dans la AWS Organizations documentation.

réglage

Pour modifier certains aspects de votre processus d'entraînement afin d'améliorer la précision du modèle de ML. Par exemple, vous pouvez entraîner le modèle de ML en générant un ensemble d'étiquetage, en ajoutant des étiquettes, puis en répétant ces étapes plusieurs fois avec différents paramètres pour optimiser le modèle.

équipe de deux pizzas

Une petite DevOps équipe que vous pouvez nourrir avec deux pizzas. Une équipe de deux pizzas garantit les meilleures opportunités de collaboration possible dans le développement de logiciels.

U

incertitude

Un concept qui fait référence à des informations imprécises, incomplètes ou inconnues susceptibles de compromettre la fiabilité des modèles de ML prédictifs. Il existe deux types d'incertitude : l'incertitude épistémique est causée par des données limitées et incomplètes, alors que l'incertitude aléatoire est causée par le bruit et le caractère aléatoire inhérents aux données.

tâches indifférenciées

Également connu sous le nom de « levage de charges lourdes », ce travail est nécessaire pour créer et exploiter une application, mais qui n'apporte pas de valeur directe à l'utilisateur final ni d'avantage concurrentiel. Les exemples de tâches indifférenciées incluent l'approvisionnement, la maintenance et la planification des capacités.

environnements supérieurs

Voir [environnement](#).

V

mise à vide

Opération de maintenance de base de données qui implique un nettoyage après des mises à jour incrémentielles afin de récupérer de l'espace de stockage et d'améliorer les performances.

contrôle de version

Processus et outils permettant de suivre les modifications, telles que les modifications apportées au code source dans un référentiel.

Appairage de VPC

Connexion entre deux VPC qui vous permet d'acheminer le trafic à l'aide d'adresses IP privées. Pour plus d'informations, veuillez consulter la rubrique [Qu'est-ce que l'appairage de VPC ?](#) dans la documentation Amazon VPC.

vulnérabilités

Défaut logiciel ou matériel qui compromet la sécurité du système.

W

cache actif

Cache tampon qui contient les données actuelles et pertinentes fréquemment consultées. L'instance de base de données peut lire à partir du cache tampon, ce qui est plus rapide que la lecture à partir de la mémoire principale ou du disque.

données chaudes

Données rarement consultées. Lorsque vous interrogez ce type de données, des requêtes modérément lentes sont généralement acceptables.

fonction de fenêtre

Fonction SQL qui effectue un calcul sur un groupe de lignes liées d'une manière ou d'une autre à l'enregistrement en cours. Les fonctions de fenêtre sont utiles pour traiter des tâches, telles que le calcul d'une moyenne mobile ou l'accès à la valeur des lignes en fonction de la position relative de la ligne en cours.

charge de travail

Ensemble de ressources et de code qui fournit une valeur métier, par exemple une application destinée au client ou un processus de backend.

flux de travail

Groupes fonctionnels d'un projet de migration chargés d'un ensemble de tâches spécifique. Chaque flux de travail est indépendant, mais prend en charge les autres flux de travail du projet. Par exemple, le flux de travail du portefeuille est chargé de prioriser les applications, de planifier les vagues et de collecter les métadonnées de migration. Le flux de travail du portefeuille fournit ces actifs au flux de travail de migration, qui migre ensuite les serveurs et les applications.

VER

Voir [écrire une fois, lire plusieurs](#).

WQF

Voir le [cadre AWS de qualification de la charge](#) de travail.

écrire une fois, lire plusieurs (WORM)

Modèle de stockage qui écrit les données une seule fois et empêche leur suppression ou leur modification. Les utilisateurs autorisés peuvent lire les données autant de fois que nécessaire, mais ils ne peuvent pas les modifier. Cette infrastructure de stockage de données est considérée comme [immuable](#).

Z

exploit Zero-Day

Une attaque, généralement un logiciel malveillant, qui tire parti d'une [vulnérabilité de type « jour zéro »](#).

vulnérabilité de type « jour zéro »

Une faille ou une vulnérabilité non atténuée dans un système de production. Les acteurs malveillants peuvent utiliser ce type de vulnérabilité pour attaquer le système. Les développeurs prennent souvent conscience de la vulnérabilité à la suite de l'attaque.

invite Zero-Shot

Fournir à un [LLM](#) des instructions pour effectuer une tâche, mais aucun exemple (plans) pouvant aider à la guider. Le LLM doit utiliser ses connaissances pré-entraînées pour gérer la tâche. L'efficacité de l'invite zéro dépend de la complexité de la tâche et de la qualité de l'invite. Voir également les instructions [en quelques clics](#).

application zombie

Application dont l'utilisation moyenne du processeur et de la mémoire est inférieure à 5 %. Dans un projet de migration, il est courant de retirer ces applications.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.