



Guida per l'utente dei piani di scalabilità

AWS Auto Scaling



AWS Auto Scaling: Guida per l'utente dei piani di scalabilità

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

I marchi e l'immagine commerciale di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in una qualsiasi modalità che possa causare confusione tra i clienti o in una qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà dei rispettivi proprietari, che possono o meno essere affiliati, collegati o sponsorizzati da Amazon.

Table of Contents

Cos'è un piano di dimensionamento?	1
Risorse supportate	1
Caratteristiche e vantaggi del piano di dimensionamento	1
Come iniziare	2
Come funzionano i piani di dimensionamento	2
Disponibilità regionale	3
Prezzi	3
Come funzionano i piani di dimensionamento	5
Best practice	8
Altre considerazioni	8
Evitare l'errore ActiveWithProblems	10
Nozioni di base	11
Fase 1: individuare le risorse scalabili	12
Prerequisiti	12
Aggiungi il gruppo Auto Scaling al nuovo piano di dimensionamento	12
Ottieni maggiori informazioni su come individuare le tue risorse scalabili	13
Fase 2: specificare la strategia di dimensionamento	15
Fase 3: configurare le impostazioni avanzate (facoltativo)	18
Impostazioni generali	18
Impostazioni dimensionamento dinamico	21
Impostazioni dimensionamento predittivo	22
Fase 4: creare il piano di dimensionamento	23
(Facoltativo) Visualizzare le informazioni di dimensionamento per una risorsa	24
Fase 5: rimozione	26
Eliminare il gruppo Auto Scaling	27
Fase 6: fasi successive	27
Esegui la migrazione del tuo piano di scalabilità	29
Passaggio 1: rivedi la configurazione esistente	29
Differenze tra piani di scalabilità e politiche di scalabilità	30
Fase 2: Creare politiche di scalabilità predittiva	30
Fase 3: esamina le previsioni generate dalle politiche di scalabilità predittiva	36
Fase 4: Prepararsi a eliminare il piano di scalabilità	37
Fase 5: Eliminare il piano di scalabilità	37
Fase 6: Riattivare la scalabilità dinamica	39

Crea politiche di scalabilità per il tracciamento degli obiettivi per i gruppi Auto Scaling	40
Crea politiche di scalabilità di tracciamento degli obiettivi per altre risorse scalabili	41
Fase 7: Riattivare la scalabilità predittiva	43
Riferimento di Amazon EC2 Auto Scaling per la migrazione delle politiche di scalabilità di tracciamento degli obiettivi	44
Riferimento di Application Auto Scaling per la migrazione delle politiche di scalabilità di Target Tracking	46
Informazioni aggiuntive	48
Registrazione delle chiamate API di CloudTrail con	50
AWS Auto Scaling Informazioni in CloudTrail	50
Comprensione delle AWS Auto Scaling voci dei file di registro	51
Sicurezza	54
AWS PrivateLink	54
Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento	55
Crea una policy dell'endpoint VPC per i piani di dimensionamento	55
Migrazione degli endpoint	56
Protezione dei dati	57
Gestione dell'identità e degli accessi	58
Controllo accessi	58
Come funzionano i piani di dimensionamento con IAM	59
Service-linked ruoli	63
Identity-based esempi di politiche	64
Convalida della conformità	71
Sicurezza dell'infrastruttura	71
Quote	73
Cronologia dei documenti	74
.....	lxxvi

Cos'è un piano di dimensionamento?

Utilizza un piano di dimensionamento per configurare la scalabilità automatica per risorse scalabili correlate o associate in pochi minuti. Ad esempio, è possibile utilizzare i tag per raggruppare risorse in categorie come produzione, test o sviluppo. Quindi, è possibile cercare e impostare piani di dimensionamento per risorse scalabili appartenenti a ciascuna categoria. Oppure, se la tua infrastruttura cloud lo include AWS CloudFormation, puoi definire modelli di stack da utilizzare per creare raccolte di risorse. Quindi, creare un piano di dimensionamento per le risorse scalabili appartenenti a ogni stack.

Risorse supportate

AWS Auto Scaling supporta l'uso di piani di scalabilità per i seguenti servizi e risorse:

- Amazon Aurora: consente di aumentare o diminuire il numero di repliche di lettura Aurora assegnate a un cluster database Aurora.
- Amazon EC2 Auto Scaling: consente di avviare o terminare le istanze EC2 aumentando o diminuendo la capacità desiderata di un gruppo Auto Scaling.
- Amazon Elastic Container Service: consente di aumentare o diminuire il numero di attività desiderato in Amazon ECS.
- Amazon DynamoDB: consente di aumentare o diminuire la capacità di lettura e scrittura assegnata alla tabella Dynamo DB o a un indice secondario globale.
- Parco istanze Spot: avvia o termina le istanze EC2 aumentando o diminuendo la capacità target di un parco istanze Spot.

Caratteristiche e vantaggi del piano di dimensionamento

I piani di dimensionamento offrono le seguenti caratteristiche e vantaggi:

- Rilevamento delle risorse: AWS Auto Scaling fornisce l'individuazione automatica delle risorse per aiutare a trovare risorse scalabili nell'applicazione.
- Dimensionamento dinamico: i piani di dimensionamento utilizzano i servizi Amazon EC2 Auto Scaling e Application Auto Scaling per regolare la capacità delle risorse scalabili per gestire le modifiche del traffico o del carico di lavoro. I parametri di dimensionamento dinamico possono essere parametri di utilizzo o throughput standard o personalizzati.

- Consigli di dimensionamento integrati: AWS Auto Scaling fornisce strategie di dimensionamento con suggerimenti che è possibile utilizzare per ottimizzare le prestazioni, i costi o un equilibrio tra i due.
- Dimensionamento predittivo: i piani di dimensionamento supportano anche il dimensionamento predittivo per i gruppi Auto Scaling. Ciò aiuta a scalare la capacità Amazon EC2 più velocemente quando si verificano regolarmente picchi.

Important

Se utilizzi piani di scalabilità solo per la scalabilità predittiva, ti consigliamo vivamente di impostare invece politiche di scalabilità predittiva direttamente sulle tue risorse di Auto Scaling. Questa opzione offre più funzionalità, come l'utilizzo di aggregazioni di metriche per creare nuove metriche personalizzate o conservare i dati storici delle metriche tra le distribuzioni. blue/green Per ulteriori informazioni su Amazon EC2 Auto Scaling, [consulta Predictive scaling for Amazon EC2 Auto Scaling nella Amazon EC2 Auto Scaling User Guide](#). Per ulteriori informazioni su Application Auto Scaling, vedere [Predictive scaling for Application Auto Scaling nella Application Auto Scaling User Guide](#). Per una guida sulla migrazione dai piani di scalabilità alle politiche di scalabilità predittiva di Amazon EC2 Auto Scaling, consulta. [Esegui la migrazione del tuo piano di scalabilità](#)

Come iniziare

Utilizza le risorse seguenti per creare e utilizzare un piano di dimensionamento:

- [Come funzionano i piani di dimensionamento](#)
- [Best practice per i piani di dimensionamento](#)
- [Nozioni di base sui piani di dimensionamento](#)

Come funzionano i piani di dimensionamento

È possibile creare, accedere e gestire i piani di dimensionamento utilizzando una qualsiasi delle seguenti interfacce:

- Console di gestione AWS: fornisce un'interfaccia Web da utilizzare per l'accesso ai piani di dimensionamento. Se ti sei registrato a uno Account AWS, puoi accedere ai tuoi piani di scalabilità

accedendo a Console di gestione AWS, utilizzando la casella di ricerca sulla barra di navigazione per cercare e quindi selezionando. AWS Auto Scaling

- AWS Command Line Interface (AWS CLI) — Fornisce comandi per un ampio set di Servizi AWS ed è supportato su Windows, macOS e Linux. Per iniziare, consulta la [Guida per l'utente di AWS Command Line Interface](#). Per ulteriori informazioni, consulta [piani di dimensionamento automatico](#) nella Documentazione di riferimento sui comandi AWS CLI .
- AWS Tools for Windows PowerShell— Fornisce comandi per un'ampia gamma di AWS prodotti per coloro che eseguono script nell' PowerShell ambiente. Per iniziare, consulta la [Guida per l'utente di AWS Strumenti per PowerShell](#). Per ulteriori informazioni, consulta la [Documentazione di riferimento per Cmdlet AWS Strumenti per PowerShell](#).
- AWS SDKs— Fornisce operazioni API specifiche della lingua e si occupa di molti dettagli di connessione, come il calcolo delle firme, la gestione dei tentativi di richiesta e la gestione degli errori. Per ulteriori informazioni, consulta [AWS SDKs](#).
- API HTTPS: forniscono operazioni API di basso livello accessibili tramite richieste HTTPS. Per ulteriori informazioni, consulta la [documentazione di riferimento dell'API di AWS Auto Scaling](#).
- CloudFormation— Supporta la creazione di piani di scalabilità utilizzando modelli. CloudFormation Per ulteriori informazioni, consulta il [AWS::AutoScalingPlans::ScalingPlan](#) riferimento nella Guida per l'CloudFormation utente.

Disponibilità regionale

L' AWS Auto Scaling API è disponibile in diverse regioni Regioni AWS e fornisce un endpoint per ciascuna di queste regioni. Per un elenco di tutte le regioni e gli endpoint in cui l'API è attualmente disponibile, consulta [AWS Auto Scaling Endpoint e quote negli Endpoint e](#) per Amazon Web ARNs Services in Cina nella Guida per l' l'utente della regione segreta.

Prezzi

Tutte le funzionalità del piano di dimensionamento sono abilitate per l'uso. Le funzionalità vengono fornite senza costi aggiuntivi oltre ai costi del servizio CloudWatch e Cloud AWS alle altre risorse utilizzate.

Note

La funzionalità di scalabilità predittiva si basa sull' CloudWatch [GetMetricData](#) operazione di raccolta di dati metrici storici per la previsione della capacità, il che comporta costi. Tuttavia,

se abiliti la scalabilità predittiva con una policy di Amazon EC2 Auto Scaling anziché con un piano di scalabilità, non ci sono costi per le chiamate verso `GetMetricData`

Come funzionano i piani di dimensionamento

AWS Auto Scaling consente di utilizzare i piani di scalabilità per configurare una serie di istruzioni per scalare le risorse. Se utilizzi CloudFormation o aggiungi tag a risorse scalabili, puoi impostare piani di scalabilità per diversi set di risorse, per applicazione. La AWS Auto Scaling console fornisce consigli per strategie di scalabilità personalizzate per ogni risorsa. Dopo aver creato i piani di dimensionamento, unisce i metodi di dimensionamento dinamici e predittivi a supporto della strategia di dimensionamento.

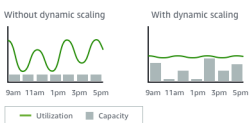
Cos'è una strategia di dimensionamento?

La strategia di scalabilità spiega AWS Auto Scaling come ottimizzare l'utilizzo delle risorse nel piano di scalabilità. È possibile ottimizzare la disponibilità, il costo o un equilibrio di entrambi. In alternativa, puoi creare strategie personalizzate in funzione dei parametri e delle soglie definiti. È possibile impostare delle strategie separate per ciascuna risorsa o per tipo di risorsa.



Che cos'è il dimensionamento dinamico?

Il dimensionamento dinamico crea policy di dimensionamento di monitoraggio obiettivo per le risorse nel piano di dimensionamento. Tali policy di dimensionamento regolano la capacità di risorse in risposta alle variazioni di utilizzo delle risorse. L'obiettivo è fornire capacità sufficiente per mantenere l'utilizzo delle risorse entro il valore obiettivo specificato dalla strategia di dimensionamento. Questa operazione può essere paragonata al modo in cui il termostato regola la temperatura di una casa. Tu selezioni la temperatura, il termostato si occupa del resto.



Ad esempio, è possibile configurare il piano di dimensionamento e mantenere il numero di operazioni che il Amazon Elastic Container Service (Amazon ECS) esegue al 75% di CPU. Quando l'utilizzo della CPU del servizio supera il 75% (ossia più del 75% della CPU riservata per il servizio in uso), questo attiva la policy di dimensionamento e aggiunge un'altra attività al servizio che possa aiutarti con il picco di carico.

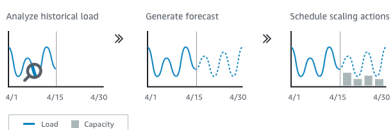
Che cos'è il dimensionamento predittivo?

Il dimensionamento predittivo usa il machine learning per analizzare il carico di lavoro storico di ogni risorsa ed eseguire regolarmente previsioni di carico futuro. Questo è simile a come funzionano le previsioni meteorologiche. Utilizzando la previsione, il dimensionamento predittivo genera operazioni programmate di dimensionamento per assicurare che le capacità delle risorse siano disponibili prima che la tua applicazione ne abbia bisogno. Come nel dimensionamento dinamico, il dimensionamento predittivo lavora per mantenere l'utilizzo al valore target specificato dalla strategia di dimensionamento.

⚠ Important

Se utilizzi piani di scalabilità solo per la scalabilità predittiva, ti consigliamo vivamente di impostare invece politiche di scalabilità predittiva direttamente sulle tue risorse di Auto Scaling. Questa opzione offre più funzionalità, come l'utilizzo di aggregazioni di metriche per creare nuove metriche personalizzate o conservare i dati storici delle metriche tra le distribuzioni. blue/green Per ulteriori informazioni su Amazon EC2 Auto Scaling, [consulta Predictive scaling for Amazon EC2 Auto Scaling nella Amazon EC2 Auto Scaling User Guide](#). Per ulteriori informazioni su Application Auto Scaling, vedere [Predictive scaling for Application Auto Scaling nella Application Auto Scaling User Guide](#).

Per una guida sulla migrazione dai piani di scalabilità alle politiche di scalabilità predittiva di Amazon EC2 Auto Scaling, consulta. [Esegui la migrazione del tuo piano di scalabilità](#)



Ad esempio, puoi abilitare il dimensionamento predittivo e configurare la tua strategia di dimensionamento per mantenere l'utilizzo medio della CPU del gruppo Auto Scaling al 50%. Considera una previsione in base alla quale i picchi di traffico si verificano ogni giorno alle 8 del mattino. Il piano di dimensionamento crea le operazioni di dimensionamento programmate future per assicurare che il gruppo Auto Scaling sia pronto a gestire il traffico in anticipo. Questo aiuta a mantenere le prestazioni dell'applicazione costanti, con l'obiettivo di avere sempre la capacità richiesta per mantenere l'utilizzo delle risorse il più possibile vicino al 50% in qualsiasi momento.

Di seguito sono riportati i concetti chiave per comprendere il dimensionamento predittivo:

- **Previsione del carico:** AWS Auto Scaling analizza fino a 14 giorni di cronologia per una metrica di carico specificata e prevede la domanda futura per i due giorni successivi. Questi dati sono disponibili in intervalli di un'ora e vengono aggiornati quotidianamente.
- **Azioni di scalabilità pianificate:** AWS Auto Scaling pianifica le azioni di ridimensionamento che aumentano e diminuiscono in modo proattivo la capacità in base alla previsione di carico. All'ora pianificata, AWS Auto Scaling aggiorna la capacità minima con il valore specificato dall'azione di scalabilità pianificata. L'obiettivo è mantenere l'utilizzo delle risorse entro il valore obiettivo specificato dalla strategia di dimensionamento. Se la tua applicazione richiede una capacità superiore rispetto a quella prevista, il dimensionamento dinamico è disponibile per aggiungere ulteriore capacità.
- **Comportamento di capacità massima:** i limiti di capacità minima e massima per la scalabilità automatica si applicano a ciascuna risorsa. Tuttavia, è possibile controllare se la tua applicazione è in grado di aumentare la capacità oltre quella massima quando la capacità di previsione è superiore alla massima.

Best practice per i piani di dimensionamento

Le seguenti best practice consentono di sfruttare al massimo i piani di dimensionamento:

- Quando crei un modello di avvio o una configurazione di avvio, abilita il monitoraggio dettagliato per ottenere dati CloudWatch metrici per le istanze EC2 con una frequenza di un minuto, in modo da garantire una risposta più rapida alle modifiche del carico. Il dimensionamento sui parametri a intervalli di 5 minuti potrebbe rallentare il tempo di risposta e causare il dimensionamento su dati di parametro obsoleti. Per impostazione predefinita, le istanze EC2 sono abilitate per il monitoraggio base, il che significa che i dati dei parametri per le istanze sono disponibili a intervalli di 5 minuti. Per un costo aggiuntivo, è possibile abilitare il monitoraggio dettagliato per ottenere i dati dei parametri per le istanze con una frequenza di un minuto. Per ulteriori informazioni, consulta la sezione relativa alla [configurazione del monitoraggio per le istanze di Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- È consigliabile abilitare anche i parametri del gruppo Auto Scaling. In caso contrario, i dati della capacità effettiva non vengono mostrati nei grafici di previsione della capacità disponibili al completamento della procedura guidata per la creazione del piano di dimensionamento. Per ulteriori informazioni, consulta i [CloudWatch parametri di monitoraggio per i gruppi e le istanze di Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- Controlla il tipo di istanza utilizzato dal gruppo Auto Scaling e fai attenzione a usare un tipo di istanza con prestazioni burstable. Le istanze Amazon EC2 espandibili, ossia le istanze T3 e T2, sono progettate per offrire un livello base di prestazioni della CPU con la possibilità di raggiungere un livello più alto quando richiesto dal carico di lavoro. A seconda dell'utilizzo di destinazione specificato dal piano di dimensionamento, è possibile correre il rischio di superare la linea di base e quindi terminare i crediti di CPU, che limitano le prestazioni. Per ulteriori informazioni, consulta [Crediti CPU e prestazioni di base per istanze espandibili](#). Per configurare queste istanze come unlimited, consulta [Using an Auto Scaling group to launch a Burstable Performance Instance come Unlimited](#) nella Guida per l'utente di Amazon EC2.

Altre considerazioni

Important

Se utilizzi piani di scalabilità solo per la scalabilità predittiva, ti consigliamo vivamente di impostare invece politiche di scalabilità predittiva direttamente sulle tue risorse di Auto

Scaling. Questa opzione offre più funzionalità, come l'utilizzo di aggregazioni di metriche per creare nuove metriche personalizzate o conservare i dati storici delle metriche tra le distribuzioni. blue/green Per ulteriori informazioni su Amazon EC2 Auto Scaling, [consulta Predictive scaling for Amazon EC2 Auto Scaling nella Amazon EC2 Auto Scaling User Guide](#). Per ulteriori informazioni su Application Auto Scaling, vedere [Predictive scaling for Application Auto Scaling nella Application Auto Scaling User Guide](#). Per una guida sulla migrazione dai piani di scalabilità alle politiche di scalabilità predittiva di Amazon EC2 Auto Scaling, consulta. [Esegui la migrazione del tuo piano di scalabilità](#)

Tieni presente le seguenti considerazioni aggiuntive:

- Il dimensionamento predittivo impiega previsioni di carico per pianificare la capacità in futuro. La qualità delle previsioni varia in base alla ciclicità del carico e all'applicabilità del modello di previsione qualificato. Il dimensionamento predittivo può essere eseguito in modalità di sola previsione per valutare la qualità delle previsioni e le operazioni di dimensionamento create dalle previsioni. È possibile impostare la modalità del dimensionamento predittivo su Forecast only (Solo previsione) quando si crea il piano di dimensionamento, quindi modificarlo in Forecast and scale (Previsione e dimensionamento) al termine della valutazione della qualità della previsione. Per ulteriori informazioni, consultare [Impostazioni dimensionamento predittivo](#) e [Monitoraggio e valutazione delle previsioni](#).
- Se si sceglie di specificare parametri diversi per il dimensionamento predittivo, è necessario accertarsi che il parametro di dimensionamento e quello di caricamento siano fortemente correlati. Il valore del parametro deve aumentare e diminuire in proporzione al numero di istanze nel gruppo Auto Scaling. In questo modo, i dati dei parametri possono essere utilizzati per scalare proporzionalmente il numero di istanze. Ad esempio, il parametro di carico equivale al conteggio totale di richieste e il parametro di dimensionamento equivale all'utilizzo medio della CPU. Se il conteggio totale delle richieste aumenta del 50%, l'utilizzo medio della CPU deve anche aumentare del 50%, purché la capacità rimanga invariata.
- Prima di creare un piano di scalabilità, è necessario eliminare tutte le azioni di scalabilità pianificate in precedenza che non sono più necessarie accedendo alle console da cui sono state create. AWS Auto Scaling non crea un'azione di ridimensionamento predittiva che si sovrappone a un'azione di ridimensionamento pianificata esistente.
- Le tue impostazioni personalizzate per la capacità minima e massima, insieme ad altre impostazioni utilizzate per il dimensionamento dinamico, vengono visualizzate in altre console. Tuttavia, è consigliabile che, dopo la creazione di un piano di dimensionamento, queste

impostazioni non vengano modificate da altre console, perché il piano di dimensionamento non riceve aggiornamenti da altre console.

- Il piano di dimensionamento può contenere risorse provenienti da più servizi, ma ogni risorsa può trovarsi in un solo piano di dimensionamento alla volta.

Evitare l'errore `ActiveWithProblems`

Un errore `ActiveWithProblems` può verificarsi quando viene creato un piano di scalabilità o vengono aggiunte risorse a un piano di scalabilità. L'errore si verifica quando il piano di dimensionamento è attivo, ma non è stato possibile applicare la configurazione di dimensionamento di una o più risorse.

In genere, ciò accade perché una risorsa dispone già di una policy di dimensionamento o un gruppo Auto Scaling non soddisfa i requisiti minimi per il dimensionamento predittivo.

Se una delle risorse dispone già di policy di dimensionamento da varie console dei servizi, per impostazione predefinita AWS Auto Scaling non sovrascrive queste altre policy di dimensionamento o ne crea di nuove. Facoltativamente, puoi eliminare le politiche di scalabilità esistenti e sostituirle con politiche di scalabilità di Target Tracking create dalla console. AWS Auto Scaling Questa operazione viene effettuata abilitando l'impostazione `Replace external scaling policies` (Sostituisci policy di dimensionamento esterne) per ogni risorsa che dispone di policy di dimensionamento da sovrascrivere.

Con il dimensionamento predittivo, si consiglia di attendere 24 ore dopo la creazione di un nuovo gruppo Auto Scaling per configurare il dimensionamento predittivo. Devono essere presenti almeno 24 ore di dati storici per generare la previsione iniziale. Se il gruppo dispone di meno di 24 ore di dati storici ed è abilitato il dimensionamento predittivo, il piano di dimensionamento non sarà in grado di generare una previsione fino al prossimo periodo di previsione dopo che il gruppo ha raccolto la quantità di dati richiesta. Tuttavia, è anche possibile modificare e salvare il piano di dimensionamento per riavviare il processo di previsione non appena sono disponibili i dati di 24 ore.

Nozioni di base sui piani di dimensionamento

Prima di creare un piano di dimensionamento da usare con l'applicazione, esamina accuratamente l'applicazione quando viene eseguita in Cloud AWS. Prendi nota di quanto segue:

- Se si dispone di regole di dimensionamento esistenti create da altre console. Puoi sostituire le policy di dimensionamento esistenti oppure mantenerle (senza che ti sia consentito apportare modifiche ai relativi valori) quando crei il piano di dimensionamento.
- L'utilizzo di destinazione che ha senso per ogni risorsa scalabile nell'applicazione in base alla risorsa nel suo complesso. Ad esempio, esamina la quantità di CPU che le istanze EC2 in un gruppo Auto Scaling dovrebbero utilizzare rispetto alla CPU disponibile. Oppure, per un servizio come DynamoDB che utilizza un modello di throughput con provisioning, esamina la quantità di attività di lettura e scrittura che una tabella o un indice dovrebbe utilizzare rispetto alla velocità effettiva disponibile; in altre parole, il rapporto tra la capacità utilizzata e la capacità fornita. È possibile modificare l'utilizzo di destinazione in qualsiasi momento dopo aver creato il piano di dimensionamento.
- Quanto tempo occorre per avviare e configurare un server. Queste informazioni ti aiuteranno a configurare una finestra per il riscaldamento di ciascuna istanza EC2 dopo l'avvio, per assicurarti di non avviare un nuovo server mentre il precedente è ancora in fase di avvio.
- Se la cronologia dei parametri è sufficientemente lunga per essere utilizzata con il dimensionamento predittivo (se si utilizzano gruppi Auto Scaling appena creati). In generale, avere 14 giorni completi di dati storici si traduce in previsioni più accurate. Il periodo minimo è di 24 ore.

Più comprendi gli aspetti della tua applicazione, più efficace potrai rendere il tuo piano di dimensionamento.

Le seguenti attività ti consentono di acquisire familiarità con i piani di dimensionamento. Creerai un piano di dimensionamento per un singolo gruppo Auto Scaling e abiliterai il dimensionamento predittivo e il dimensionamento dinamico.

Processi

- [Fase 1: individuare le risorse scalabili](#)
- [Fase 2: specificare la strategia di dimensionamento](#)
- [Fase 3: configurare le impostazioni avanzate \(facoltativo\)](#)
- [Fase 4: creare il piano di dimensionamento](#)

- [Fase 5: rimozione](#)
- [Fase 6: fasi successive](#)

Fase 1: individuare le risorse scalabili

Questa sezione include un'introduzione pratica alla creazione di piani di dimensionamento nella console AWS Auto Scaling . Se si tratta del primo piano di dimensionamento, ti consigliamo di iniziare creando un piano di dimensionamento di esempio utilizzando un gruppo Amazon EC2 Auto Scaling.

Prerequisiti

Per utilizzare un piano di dimensionamento, crea un gruppo Auto Scaling. Avvia almeno un'istanza Amazon EC2 nel gruppo Auto Scaling. Per maggiori informazioni, consulta [Nozioni di base sul Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Utilizza un gruppo Auto Scaling con CloudWatch metriche abilitate per avere dati sulla capacità nei grafici disponibili quando completi la procedura guidata Crea piano di scalabilità. Per ulteriori informazioni, consulta le [CloudWatch metriche di monitoraggio per i gruppi e le istanze di Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Genera un carico di lavoro per alcuni giorni o più per avere a disposizione i dati CloudWatch metrici per la funzionalità di scalabilità predittiva, se possibile.

Verifica di disporre delle autorizzazioni necessarie per l'utilizzo dei piani di dimensionamento. Per ulteriori informazioni, consulta [Gestione delle identità e degli accessi per piani di scalabilità](#).

Aggiungi il gruppo Auto Scaling al nuovo piano di dimensionamento

Quando crei un piano di dimensionamento dalla console, ti aiuta a trovare le risorse scalabili come primo passo. Prima di iniziare, verifica che siano soddisfatti i seguenti requisiti

- Hai creato un gruppo Auto Scaling e avviato almeno un'istanza EC2, come descritto nella sezione precedente.
- Il gruppo Auto Scaling creato esiste da almeno 24 ore.

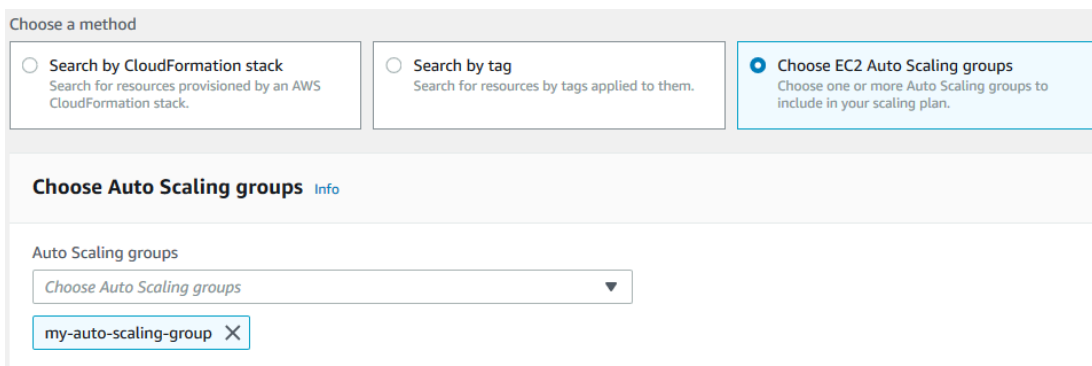
Per iniziare, crea un piano di dimensionamento.

1. Apri la console all' AWS Auto Scaling indirizzo. <https://console.aws.amazon.com/autoscaling/>

2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la stessa Regione utilizzata durante la creazione del gruppo Auto Scaling.
3. Nella pagina di benvenuto, scegliere Get started (Inizia).
4. Nella pagina Find scalable resources (Trova risorse scalabili), esegui una delle seguenti operazioni:
 - Scegli Cerca per CloudFormation pila, quindi scegli lo CloudFormation stack da usare.
 - Scegli Search by tag (Cerca per tag). Quindi, per ciascun tag, scegli una chiave di tag da Key (Chiave) e valori di tag da Value (Valore). Per aggiungere tag, scegliere Add another row (Aggiungi un'altra riga). Per rimuovere i tag, scegliere Remove (Rimuovi).
 - Scegli Choose EC2 Auto Scaling groups (Scegli gruppi EC2 di Auto Scaling), quindi scegli uno o più gruppi Auto Scaling.

Note

Per un tutorial introduttivo, scegli Choose EC2 Auto Scaling groups (Scegli gruppi EC2 Auto Scaling), quindi scegli il gruppo Auto Scaling creato.



Choose a method

Search by CloudFormation stack
Search for resources provisioned by an AWS CloudFormation stack.

Search by tag
Search for resources by tags applied to them.

Choose EC2 Auto Scaling groups
Choose one or more Auto Scaling groups to include in your scaling plan.

Choose Auto Scaling groups [Info](#)

Auto Scaling groups

Choose Auto Scaling groups ▼

my-auto-scaling-group X

5. Scegli Next (Successivo) per continuare con il processo di creazione del piano di dimensionamento.

Ottieni maggiori informazioni su come individuare le tue risorse scalabili

Se hai già creato un piano di scalabilità di esempio e desideri crearne altri, consulta i seguenti scenari per l'utilizzo di uno CloudFormation stack o di un set di tag in modo più dettagliato. Puoi utilizzare

questa sezione per decidere se scegliere l'opzione Search by CloudFormation stack o Search by tag per scoprire le tue risorse scalabili quando usi la console per creare il tuo piano di scalabilità.

Quando scegli l'opzione Cerca per CloudFormation pila o Cerca per tag nel passaggio 1 della procedura guidata Crea piano di scalabilità, le risorse scalabili associate allo stack o al set di tag sono disponibili per il piano di scalabilità. Quando definisci il piano di dimensionamento, puoi specificare quali di queste risorse includere o escludere.

Alla scoperta di risorse scalabili utilizzando uno stack CloudFormation

Quando si utilizza CloudFormation, si utilizzano gli stack per il provisioning delle risorse. Tutte le risorse di uno stack sono definite dal modello dello stack. Il piano di dimensionamento aggiunge un livello di orchestrazione nella parte superiore dello stack che semplifica la configurazione del dimensionamento per più risorse. Senza un piano di dimensionamento, è necessario impostare il dimensionamento individualmente per ogni risorsa scalabile. Ciò significa capire l'ordine per il provisioning delle risorse e le policy di dimensionamento e cogliere le sfumature del funzionamento di queste dipendenze.

Nella AWS Auto Scaling console, puoi selezionare uno stack esistente per scansarlo alla ricerca di risorse che possono essere configurate per il ridimensionamento automatico. AWS Auto Scaling trova solo le risorse definite nello stack selezionato. Non si espande attraverso gli stack nidificati.

Affinché i servizi ECS siano rilevabili in uno CloudFormation stack, la AWS Auto Scaling console deve sapere su quale cluster ECS è in esecuzione il servizio. Ciò richiede che i servizi ECS si trovino nello stesso CloudFormation stack del cluster ECS che esegue il servizio. Altrimenti, devono far parte del cluster predefinito. Per essere identificato correttamente, il nome del servizio ECS deve inoltre essere univoco in ciascuno di questi cluster ECS.

[Per ulteriori informazioni su CloudFormation, consulta *What is? CloudFormation*](#) nella Guida AWS CloudFormation per l'utente.

Individuazione di risorse scalabili tramite tag

I tag forniscono metadati che possono essere utilizzati per scoprire le risorse scalabili correlate nella AWS Auto Scaling console, utilizzando i filtri dei tag.

Utilizza i tag per individuare una qualsiasi delle seguenti risorse:

- Cluster di database Aurora
- Gruppi Auto Scaling
- Tabelle DynamoDB e indici secondari globali

Quando esegui una ricerca in base a più tag, una risorsa viene rilevata solo se dispone di tutti i tag elencati.

Per ulteriori informazioni sull'assegnazione di tag, consulta la documentazione seguente.

- Scopri come assegnare [tag ai cluster di Aurora](#) nella Guida per l'utente di Amazon Aurora.
- Scopri come assegnare [tag ai gruppi Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- Scopri come assegnare [tag alle risorse Dynamo DB](#) nella Guida per gli sviluppatori di Amazon DynamoDB.

Fase 2: specificare la strategia di dimensionamento

Utilizzare la procedura seguente per specificare strategie di dimensionamento per le risorse individuate nel passaggio precedente.

Per ogni tipo di risorsa, AWS Auto Scaling sceglie la metrica più comunemente utilizzata per determinare la quantità di risorsa in uso in un dato momento. Si può scegliere la strategia di dimensionamento più appropriata per ottimizzare le prestazioni dell'applicazione in base a tale parametro. Quando si abilita la funzione di dimensionamento dinamico e quella di dimensionamento predittivo, la strategia di dimensionamento è condivisa tra le due funzioni. Per ulteriori informazioni, consulta [Come funzionano i piani di dimensionamento](#).

Sono disponibili le seguenti strategie di dimensionamento:

- **Ottimizza la disponibilità:** AWS Auto Scaling ridimensiona automaticamente la risorsa verso l'alto e verso l'alto per mantenerne l'utilizzo al 40%. Questa opzione è utile quando l'applicazione ha esigenze di dimensionamento urgenti e imprevedibili.
- **Bilancia disponibilità e costi:** AWS Auto Scaling ridimensiona automaticamente la risorsa verso l'alto e verso l'alto per mantenere l'utilizzo delle risorse al 50%. Questa opzione aiuta a mantenere un'elevata disponibilità riducendo al contempo i costi.
- **Ottimizzazione dei costi:** AWS Auto Scaling ridimensiona automaticamente la risorsa verso l'alto e verso l'alto per mantenere l'utilizzo delle risorse al 70%. Questa opzione è utile per ridurre i costi se l'applicazione è in grado di gestire una ridotta capacità del buffer quando la domanda è soggetta a modifiche impreviste.

Ad esempio, il piano di dimensionamento configura il gruppo Auto Scaling in modo da aggiungere o rimuovere istanze Amazon EC2 in base alla quantità di CPU utilizzata in media per tutte le istanze nel gruppo. Puoi scegliere se ottimizzare l'utilizzo per la disponibilità, per i costi o una combinazione dei due modificando la strategia di dimensionamento.

In alternativa, puoi configurare una strategia personalizzata se una di quelle esistenti non soddisfa le tue esigenze. Con una strategia personalizzata, è possibile modificare il valore di utilizzo target, scegliere un altro parametro o entrambe le opzioni.

Important

Per il tutorial di base, completa solo la prima fase della seguente procedura, quindi scegli Next (Successivo) per continuare.

Specificare strategie di dimensionamento

1. Nella pagina Specify scaling strategy (Specifica strategia di dimensionamento), in Scaling plan details (Dettagli del piano di dimensionamento), Name (Nome), immettere un nome per il piano di dimensionamento. Il nome del piano di dimensionamento deve essere univoco all'interno del set di piani di dimensionamento per la regione. Può avere un massimo di 128 caratteri e non deve contenere barre "|", barre "/" o due punti ":".
2. Tutte le risorse incluse sono elencate per tipo di risorsa. Per Gruppi Auto Scaling, procedi come segue:

Auto Scaling groups (1) Include in scaling plan

Specify a scaling strategy for 1 Auto Scaling group.

Scaling strategy
The strategy defines the scaling metric and target value used to scale your resources.

<input checked="" type="radio"/> Optimize for availability Keep the average CPU utilization of your Auto Scaling groups at 40% to provide high availability and ensure capacity to absorb spikes in demand.	<input type="radio"/> Balance availability and cost Keep the average CPU utilization of your Auto Scaling groups at 50% to provide optimal availability and reduce costs.	<input type="radio"/> Optimize for cost Keep the average CPU utilization of your Auto Scaling groups at 70% to ensure lower costs.	<input type="radio"/> Custom Choose your own scaling metric, target value, and other settings.
---	---	--	--

Enable predictive scaling
Support your scaling strategy by continually forecasting load and proactively scheduling capacity ahead of when you need it. [Info](#)

Enable dynamic scaling
Support your scaling strategy by creating target tracking scaling policies to monitor your scaling metric and increase or decrease capacity as you need it. [Info](#)

► **Configuration details**

- a. Salta questo passaggio per utilizzare la strategia di dimensionamento e i parametri di default. Per utilizzare una strategia o parametri di ridimensionamento diversi, procedi con i seguenti passaggi:
 - i. Per Strategia di dimensionamento, scegli la strategia di dimensionamento desiderata.

Per il tutorial introduttivo, assicurati di scegliere Optimize for availability (Ottimizza per disponibilità). Ciò specifica che l'utilizzo medio della CPU del gruppo Auto Scaling verrà mantenuto al 40%.
 - ii. Se hai scelto Custom (Personalizza), espandi Configuration details (Dettagli di configurazione) per scegliere i parametri e il valore di destinazione desiderati.
 - Per Scaling metric (Parametro di dimensionamento), scegli il parametro di dimensionamento desiderato.
 - Per Target value (Valore di destinazione), scegli il valore di destinazione desiderato, ad esempio l'utilizzo o il throughput di destinazione durante un intervallo di un minuto.
 - Per Load metric (Parametro di carico) [solo gruppi Auto Scaling], scegli il parametro di carico desiderato da utilizzare per il dimensionamento predittivo.
 - Seleziona Sostituisci criteri di scalabilità esterni per specificare che è AWS Auto Scaling possibile eliminare i criteri di scalabilità creati in precedenza dall'esterno del piano di scalabilità (ad esempio da altre console) e sostituirli con nuove politiche di scalabilità di tracciamento degli obiettivi create dal piano di scalabilità.
 - b. (Facoltativo) Per impostazione predefinita, per i gruppi Auto Scaling è attivo il dimensionamento predittivo. Per disabilitare il dimensionamento predittivo per i gruppi Auto Scaling, deseleziona Enable predictive scaling (Abilita dimensionamento predittivo).
 - c. (Facoltativo) Per impostazione predefinita, il dimensionamento dinamico è abilitato per ogni tipo di risorsa. Per disabilitare il dimensionamento dinamico per un tipo di risorsa, deseleziona Enable dynamic scaling (Abilita dimensionamento dinamico).
 - d. (Facoltativo) Per impostazione predefinita, quando si specifica un'origine dell'applicazione da cui vengono individuate più risorse scalabili, tutti i tipi di risorse vengono automaticamente inclusi nel piano di dimensionamento. Per omettere un tipo di risorsa dal piano di dimensionamento, deselezionare Includi in piano di dimensionamento.
3. (Facoltativo) Per specificare una strategia di dimensionamento per un altro tipo di risorsa, ripeti i passaggi precedenti.

4. Una volta terminato, scegli Next (Successivo) per continuare con il processo di creazione del piano di dimensionamento.

Fase 3: configurare le impostazioni avanzate (facoltativo)

Ora che hai specificato la strategia di dimensionamento da usare per ogni tipo di risorsa, puoi scegliere di personalizzare le impostazioni predefinite per ogni risorsa utilizzando la fase Configure advanced settings (Configura impostazioni avanzate). Per ogni tipo di risorsa, sono disponibili per la personalizzazione più gruppi di impostazioni. Nella maggior parte dei casi, tuttavia, le impostazioni predefinite devono essere più efficienti, ad eccezione dei valori per la capacità minima e la capacità massima, che possono essere attentamente regolati.

Ignora questa procedura se desideri mantenere le impostazioni predefinite. Puoi modificare queste impostazioni in qualsiasi momento modificando il piano di dimensionamento.

Important

Per il tutorial introduttivo, apportiamo qualche modifica per aggiornare la capacità massima del gruppo Auto Scaling e per abilitare il dimensionamento predittivo in modalità di sola previsione. Anche se non è necessario personalizzare tutte le impostazioni per il tutorial, esaminiamo brevemente anche le impostazioni in ciascuna sezione.


Impostazioni generali

Utilizza questa procedura per visualizzare e personalizzare le impostazioni specificate nella fase precedente, in base alle risorse. È anche possibile personalizzare la capacità minima e la capacità massima di ciascuna risorsa.

Per visualizzare e personalizzare le impostazioni generali

1. Nella pagina Configure advanced settings (Configura impostazioni avanzate), selezionare la freccia alla sinistra di ogni intestazione della sezione per espanderla. Per il tutorial, espandi la sezione Auto Scaling groups (Gruppi Auto Scaling).
2. Nella tabella visualizzata, scegli il gruppo Auto Scaling gruppo che stai utilizzando in questo tutorial.

3. Lasciare selezionata l'opzione Include in scaling plan (Includi in piano di dimensionamento). Se questa opzione non è selezionata, la risorsa è omessa dal piano di dimensionamento. Se non si include almeno una risorsa, il piano di dimensionamento non può essere creato.
4. Per espandere la visualizzazione e visualizzare i dettagli della sezione Impostazioni generali, selezionare la freccia alla sinistra dell'intestazione della sezione.
5. È possibile effettuare scelte per uno qualsiasi dei seguenti elementi. Per questo tutorial, individuare l'impostazione Maximum capacity (Capacità massima) e inserire il valore 3 al posto di quello attuale.
 - Scaling strategy (Strategia di dimensionamento) - Consente di ottimizzare la disponibilità, il costo o un equilibrio di entrambi, oppure di specificare una strategia personalizzata.
 - Enable dynamic scaling (Abilita il dimensionamento dinamico) - Se questa impostazione è deselezionata, la risorsa selezionata non può essere dimensionata utilizzando una configurazione di dimensionamento del monitoraggio degli obiettivi.
 - Enable predictive scaling (Abilita il dimensionamento predittivo) - [Solo gruppi Auto Scaling] Se questa impostazione è deselezionata, il gruppo selezionato non può dimensionare utilizzando il dimensionamento predittivo.
 - Scaling metric (Parametro di dimensionamento) - Specifica il parametro di dimensionamento da utilizzare. Se si sceglie Custom (Personalizzata), è possibile specificare un parametro di carico personalizzato da utilizzare al posto dei parametri di carico predefiniti disponibili nella console. Per maggiori informazioni, consulta il prossimo argomento in questa sezione.
 - Target value (Valore obiettivo) - Specifica il valore di utilizzo obiettivo da usare.
 - Load metric (Parametro di carico) - [Solo gruppi Auto Scaling] Specifica il parametro di carico da utilizzare. Se si sceglie Custom (Personalizzata), è possibile specificare un parametro di carico personalizzato da utilizzare al posto dei parametri di carico predefiniti disponibili nella console. Per maggiori informazioni, consulta il prossimo argomento in questa sezione.
 - Capacità minima: specifica la capacità minima per la risorsa. AWS Auto Scaling assicura che la risorsa non scenda mai al di sotto di queste dimensioni.
 - Capacità massima: specifica la capacità massima della risorsa. AWS Auto Scaling assicura che la risorsa non superi mai queste dimensioni.

 Note

Quando utilizzi il dimensionamento predittivo, hai la possibilità di scegliere un comportamento diverso per la capacità massima in base alla capacità di previsione.

Questa impostazione si trova nella sezione Predictive scaling settings (Impostazioni dimensionamento predittivo).

parametri personalizzati

AWS Auto Scaling fornisce le metriche più comunemente utilizzate per il ridimensionamento automatico. Tuttavia, in base alle esigenze, è preferibile ottenere i dati provenienti da parametri differenti anziché i parametri nella console. Amazon CloudWatch ha molte metriche diverse tra cui scegliere. CloudWatch ti consente anche di pubblicare le tue metriche.

Si utilizza JSON per specificare una metrica CloudWatch personalizzata. Prima di seguire queste istruzioni, ti consigliamo di acquisire familiarità con la [Amazon CloudWatch User Guide](#).

Per specificare un parametro personalizzato, è necessario creare un payload in formato JSON utilizzando un set di parametri richiesti da un modello. Aggiungi i valori per ogni parametro da CloudWatch. Forniamo il modello nell'ambito delle opzioni personalizzate per Scaling metric (Parametro di dimensionamento) e Load metric (Parametro di carico) nelle impostazioni avanzate del piano di dimensionamento.

JSON rappresenta i dati in due modi:

- Un oggetto, una raccolta non ordinata di coppie nome-valore. Un oggetto viene definito all'interno di parentesi graffe sinistra ({) e destra (}). Ogni coppia nome-valore inizia con il nome, seguita dai due punti e dal valore. Le coppie nome-valore sono separate da virgole.
- Una matrice, che è un insieme ordinato di valori. Una matrice viene definita all'interno di parentesi quadre sinistra ([) e destra (]). Gli elementi nella matrice sono separati da virgole.

Ecco un esempio del modello JSON con valori di esempio per ogni parametro:

```
{
  "MetricName": "MyBackendCPU",
  "Namespace": "MyNamespace",
  "Dimensions": [
    {
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }
  ],
}
```

```
"Statistic": "Sum"  
}
```

Per ulteriori informazioni, consulta [Specifiche di parametri di dimensionamento personalizzati](#) e [Specifiche di parametri di caricamento personalizzati](#) nella Documentazione di riferimento dell'API AWS Auto Scaling .

Impostazioni dimensionamento dinamico

Utilizzate questa procedura per visualizzare e personalizzare le impostazioni per la politica di ridimensionamento del tracciamento del target che AWS Auto Scaling viene creata.

Per visualizzare e personalizzare le impostazioni per il dimensionamento dinamico

1. Per espandere la visualizzazione e visualizzare i dettagli della sezione Impostazioni dimensionamento dinamico, selezionare la freccia alla sinistra dell'intestazione della sezione.
2. È possibile effettuare scelte per i seguenti elementi. Tuttavia, le impostazioni predefinite vanno bene per questo tutorial.
 - Replace external scaling policies (Sostituisci le policy di dimensionamento esterne) - Se questa impostazione è disattivata, conserva le policy di dimensionamento esistenti create dall'esterno del piano di dimensionamento e non ne crea di nuove.
 - Disable scale-in (Disabilita dimensionamento orizzontale) - Se questa impostazione è deselezionata, la scalabilità automatica per ridurre la capacità corrente della risorsa è consentita quando il parametro specificato è inferiore al valore obiettivo.
 - Cooldown (Raffreddamento) - Crea tempi di raffreddamento per il dimensionamento orizzontale. Il tempo di raffreddamento è il tempo che la policy di dimensionamento attende prima di rendere effettiva una precedente attività di dimensionamento. Per ulteriori informazioni, consulta [Tempo di raffreddamento](#) nella Guida per l'utente di Application Auto Scaling. (Questa impostazione non viene utilizzata se la risorsa è un gruppo Auto Scaling.)
 - Riscaldamento dell'istanza — [Solo gruppi di Auto Scaling] Controlla la quantità di tempo che trascorre prima che un'istanza appena lanciata inizi a contribuire alle metriche. CloudWatch Per ulteriori informazioni, consulta [Riscaldamento delle istanze](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Impostazioni dimensionamento predittivo

Se la risorsa è un gruppo Auto Scaling, utilizzate questa procedura per visualizzare e personalizzare le impostazioni AWS Auto Scaling utilizzate per il ridimensionamento predittivo.

Per visualizzare e personalizzare le impostazioni per il dimensionamento predittivo

1. Per espandere la visualizzazione e visualizzare i dettagli della sezione Impostazioni dimensionamento predittivo, selezionare la freccia alla sinistra dell'intestazione della sezione.
2. È possibile effettuare scelte per i seguenti elementi. Per questo tutorial, modificare Predictive scaling mode (Modalità dimensionamento predittivo) in Forecast only (Sola previsione).
 - Predictive scaling mode (Modalità di dimensionamento predittivo) - Specifica la modalità di dimensionamento. L'impostazione predefinita è Forecast and scale (Previsione e dimensionamento). Se imposti il valore su Forecast only (Solo previsione), il piano di dimensionamento effettua la previsione sulla capacità futura ma non applica le operazioni di dimensionamento.
 - Pre-launch instances (Pre-avvia le istanze) - Regola le operazioni di dimensionamento programmate per l'esecuzione prioritaria in caso di dimensionamento orizzontale. Ad esempio, la previsione dichiara di aggiungere capacità alle 10:00 e il tempo relativo al buffer è 5 minuti (300 secondi). Il runtime dell'operazione di dimensionamento corrispondente è quindi 9:55. Questa funzione è utile per i gruppi Auto Scaling, dove può richiedere alcuni minuti dal momento in cui viene avviata un'istanza fino a quando non entra in servizio. Il tempo effettivo può variare in base a diversi fattori, ad esempio la dimensione dell'istanza e la presenza o meno di script di avvio da completare. Il valore predefinito è 300 secondi.
 - Max Capacity Behavior (Comportamento capacità massima) - Controlla se la risorsa selezionata può aumentare oltre la capacità massima quando la capacità di previsione è prossima o superiore alla capacità massima specificata al momento. L'opzione predefinita è Enforce the maximum capacity setting (Applica l'impostazione per la capacità massima).
 - Applica l'impostazione della capacità massima: AWS Auto Scaling non è possibile scalare la capacità delle risorse a un livello superiore alla capacità massima. La capacità massima è applicata come un limite insuperabile.
 - Imposta la capacità massima in modo che corrisponda alla capacità prevista: è AWS Auto Scaling possibile scalare la capacità delle risorse a un livello superiore alla capacità massima in modo da eguagliare ma non superare la capacità prevista.

- Aumenta la capacità massima rispetto alla capacità prevista: AWS Auto Scaling può scalare la capacità delle risorse a un livello superiore alla capacità massima in base a un valore di buffer specificato. L'intenzione è quella di assegnare alla policy di dimensionamento del monitoraggio di target capacità aggiuntiva in caso di traffico imprevisto.
 - Max capacity behavior buffer (Buffer di comportamento di capacità massima) - Se hai scelto Increase maximum capacity above forecast capacity (Aumenta la capacità massima oltre la capacità prevista), seleziona la dimensione del buffer di capacità da utilizzare quando la capacità prevista è prossima o superiore alla capacità massima. Il valore è specificato come una percentuale relativa alla capacità di previsione. Ad esempio, con un buffer del 10%, se la capacità di previsione è 50 e quella massima è 40, la capacità massima effettiva è 55.
3. Quando la definizione delle impostazioni personalizzate è terminata, seleziona Next (Successivo).

Note

Per annullare qualsiasi modifica, seleziona le risorse e scegli Revert to original (Ripristina originale). In questo modo viene ripristinato l'ultimo stato noto delle risorse selezionate all'interno del piano di dimensionamento.

Fase 4: creare il piano di dimensionamento

Nella pagina Review and create (Rivedi e crea) rivedi i dettagli del piano di dimensionamento e scegli Create scaling plan (Crea piano di dimensionamento). Si viene reindirizzati a una pagina che mostra lo stato del piano di dimensionamento. Il completamento della creazione del piano di dimensionamento può richiedere un momento affinché le risorse vengano aggiornate.

Con la scalabilità predittiva, AWS Auto Scaling analizza la cronologia della metrica di carico specificata degli ultimi 14 giorni (sono necessarie almeno 24 ore di dati) per generare una previsione per i due giorni successivi. Pianifica quindi operazioni di dimensionamento programmate per regolare la capacità delle risorse e soddisfare la richiesta prevista per ogni ora del periodo di previsione.

Dopo la creazione del piano di dimensionamento, visualizzare i dettagli del piano di dimensionamento scegliendo il nome dalla schermata Piani di dimensionamento.

(Facoltativo) Visualizzare le informazioni di dimensionamento per una risorsa

Utilizzare questa procedura per visualizzare le informazioni di dimensionamento create per una risorsa.

I dati vengono presentati nei seguenti modi:

- Grafici che mostrano i dati cronologici metrici recenti di CloudWatch
- Grafici di scalabilità predittiva che mostrano le previsioni di carico e le previsioni di capacità basate sui dati di AWS Auto Scaling
- Una tabella che elenca tutte le operazioni di dimensionamento predittivo programmate per la risorsa.

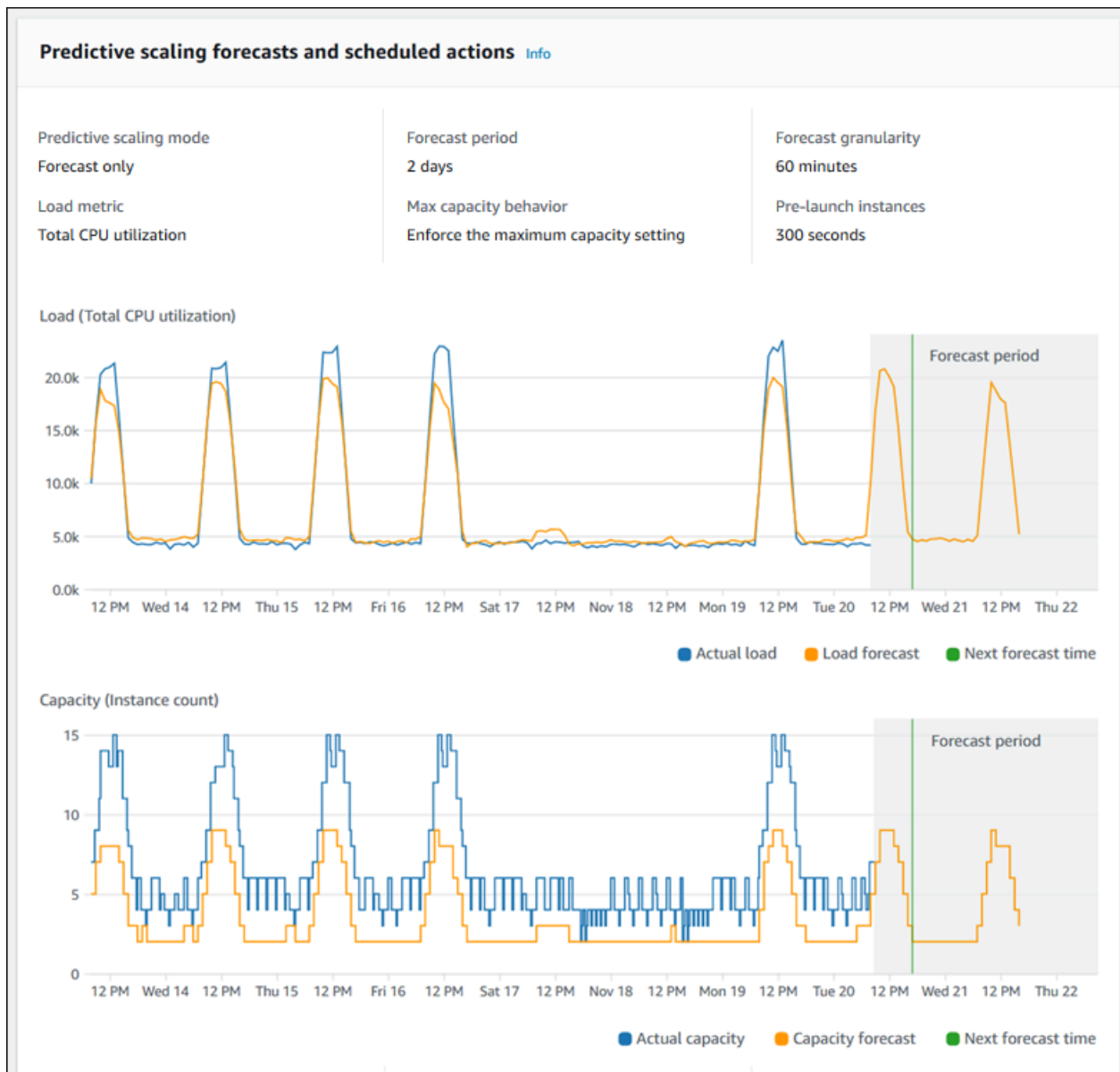
Per visualizzare le informazioni di dimensionamento per una risorsa

1. Apri la console all'indirizzo. AWS Auto Scaling <https://console.aws.amazon.com/autoscaling/>
2. Nella pagina Scaling plans (Piani di dimensionamento), selezionare il piano di dimensionamento.
3. Nella pagina Scaling plan details (Dettagli del piano di dimensionamento), scegliere la risorsa da visualizzare.

Monitoraggio e valutazione delle previsioni

Una volta che il piano di dimensionamento è in esecuzione, è possibile monitorare la previsione del carico, la previsione della capacità e le operazioni di dimensionamento per esaminare le prestazioni del dimensionamento predittivo. Tutti questi dati sono disponibili nella AWS Auto Scaling console per tutti i gruppi di Auto Scaling abilitati per la scalabilità predittiva. Ricorda che il piano di dimensionamento richiede almeno 24 ore di dati storici del carico per effettuare la previsione iniziale.

Nel seguente esempio, il lato sinistro di ogni grafico mostra un modello storico. Il lato destro mostra la previsione generata dal piano di dimensionamento per il periodo di previsione. Vengono tracciati entrambi i valori, attuali e previsti (in blu e in arancione).



AWS Auto Scaling apprende automaticamente dai tuoi dati. In primo luogo, effettua una previsione del carico. Quindi, il calcolo della previsione della capacità determina il numero minimo di istanze necessarie per supportare l'applicazione. In base alla previsione della capacità, AWS Auto Scaling pianifica le operazioni di dimensionamento che dimensionano il gruppo Auto Scaling prima delle modifiche del carico previste. Se il dimensionamento dinamico è abilitato (scelta consigliata), il gruppo Auto Scaling può dimensionare orizzontalmente capacità aggiuntiva (o rimuovere capacità) in base all'utilizzo corrente del gruppo di istanze.

Quando si valuta l'efficacia del dimensionamento predittivo, monitora quanto corrispondono i valori effettivi e quelli di previsione nel corso del tempo. Quando si crea un piano di scalabilità, AWS Auto Scaling fornisce grafici basati sui dati effettivi più recenti. Fornisce inoltre una previsione

iniziale per le successive 48 ore. Tuttavia al momento della creazione del piano di dimensionamento sono disponibili pochissimi dati di previsione per confrontare i dati effettivi. Attendi che il piano di dimensionamento abbia ottenuto i valori di previsione per alcuni periodi prima di confrontare i valori di previsione storici con i valori effettivi. Dopo alcuni giorni di previsioni giornaliere, avrai un campione più ampio di valori di previsione da confrontare con i valori effettivi.

Per i modelli che ricorrono su base giornaliera, l'intervallo di tempo tra la creazione del piano di dimensionamento e la valutazione dell'efficacia della previsione può essere anche di pochi giorni. Tuttavia, questo lasso di tempo non è sufficiente per valutare la previsione in base a un cambiamento di modello recente. Ad esempio, supponiamo che stai osservando le previsioni per un gruppo Auto Scaling che ha iniziato una nuova campagna di marketing la scorsa settimana. La campagna aumenta in modo significativo il traffico Web per gli stessi due giorni ogni settimana. In situazioni come questa, consigliamo di attendere che il gruppo raccolga per una o due settimane intere i nuovi dati prima di valutare l'efficacia della previsione. Lo stesso suggerimento vale per un gruppo Auto Scaling appena creato che ha iniziato a raccogliere solo i dati dei parametri.

Se i valori effettivi e di previsione non corrispondono dopo averli monitorati per un periodo di tempo appropriato, è necessario considerare anche la scelta del parametro di carico. Per essere efficace, il parametro di carico deve rappresentare una misura affidabile e accurata del carico totale in tutte le istanze nel gruppo Auto Scaling. Il parametro di carico è fondamentale per il dimensionamento predittivo. Se scegli un parametro di carico non ottimale, ciò potrebbe impedire al dimensionamento predittivo di effettuare accurate previsioni di carico e di capacità e la pianificazione delle corrette modifiche alla capacità per il gruppo Auto Scaling.

Fase 5: rimozione

Dopo aver completato il tutorial sulle nozioni di base, puoi scegliere di mantenere il tuo piano di dimensionamento. Tuttavia, se non utilizzi questo piano di dimensionamento attivamente, potresti considerare di eliminarlo, in modo che l'account non vada incontro a spese superflue.

L'eliminazione di un piano di scalabilità comporta l'eliminazione delle politiche di scalabilità di Target Tracking, degli CloudWatch allarmi associati e delle azioni di scalabilità predittiva create per conto dell'utente. AWS Auto Scaling

L'eliminazione di un piano di scalabilità non elimina lo CloudFormation stack, il gruppo Auto Scaling o altre risorse scalabili.

Come eliminare un piano di dimensionamento

1. Apri la console all'indirizzo. AWS Auto Scaling <https://console.aws.amazon.com/autoscaling/>
2. Nella pagina Scaling plans (Piani di dimensionamento), selezionare il piano di dimensionamento creato per questo tutorial e selezionare Delete (Elimina).
3. Quando viene richiesta la conferma, seleziona Elimina.

Una volta eliminato il piano di dimensionamento, le tue risorse non sono in grado di ripristinare la propria capacità originale. Ad esempio, se il tuo gruppo Auto Scaling viene dimensionato a 10 istanze quando elimini il piano di dimensionamento, rimarrà dimensionato a 10 istanze una volta eliminato il piano di dimensionamento. È possibile aggiornare la capacità di risorse specifiche mediante l'accesso alla console per ciascun servizio.

Eliminare il gruppo Auto Scaling

Per evitare che l'account accumuli costi Amazon EC2, è necessario eliminare anche il gruppo Auto Scaling creato per questo tutorial.

Per step-by-step istruzioni, consulta [Eliminare il gruppo Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Fase 6: fasi successive

Ora che hai familiarizzato con i piani di dimensionamento e alcune delle sue caratteristiche, prova a creare il tuo modello di piano di dimensionamento utilizzando CloudFormation.

Un CloudFormation modello è un file di testo in formato JSON o YAML che descrive l'infrastruttura Amazon Web Services necessaria per eseguire un'applicazione o un servizio insieme a tutte le interconnessioni tra i componenti dell'infrastruttura. Con CloudFormation, distribuisce e gestisci una raccolta di risorse associata come stack. CloudFormation è disponibile senza costi aggiuntivi e paghi solo per le AWS risorse necessarie per eseguire le applicazioni. Le risorse possono essere costituite da qualsiasi AWS risorsa definita all'interno del modello. Per ulteriori informazioni, consulta [How CloudFormation works](#) in the AWS CloudFormation User Guide.

Nella Guida per l'utente di AWS CloudFormation è riportato un modello semplice per iniziare. Il modello di esempio è disponibile come esempio nella [AWS::AutoScalingPlans::ScalingPlan](#) sezione della documentazione di riferimento del CloudFormation modello. Il modello di esempio crea un piano

di dimensionamento per un singolo gruppo Auto Scaling e consente il dimensionamento predittivo e il dimensionamento dinamico.

Per ulteriori informazioni, consulta [Nozioni di base su CloudFormation](#) nella Guida per l'utente di AWS CloudFormation .

Esegui la migrazione del tuo piano di scalabilità

Puoi migrare da un piano di scalabilità alle politiche di scaling di Amazon EC2 Auto Scaling e Application Auto Scaling.

Processo di migrazione

- [Passaggio 1: rivedi la configurazione esistente](#)
- [Fase 2: Creare politiche di scalabilità predittiva](#)
- [Fase 3: esamina le previsioni generate dalle politiche di scalabilità predittiva](#)
- [Fase 4: Prepararsi a eliminare il piano di scalabilità](#)
- [Fase 5: Eliminare il piano di scalabilità](#)
- [Fase 6: Riattivare la scalabilità dinamica](#)
- [Fase 7: Riattivare la scalabilità predittiva](#)
- [Riferimento di Amazon EC2 Auto Scaling per la migrazione delle politiche di scalabilità di tracciamento degli obiettivi](#)
- [Riferimento di Application Auto Scaling per la migrazione delle politiche di scalabilità di Target Tracking](#)
- [Informazioni aggiuntive](#)

Important

Per migrare un piano di scalabilità, è necessario completare più passaggi nell'ordine esatto. Durante la migrazione del piano di scalabilità, non aggiornarlo, poiché ciò interrompe l'ordine delle operazioni e potrebbe causare comportamenti indesiderati.

Passaggio 1: rivedi la configurazione esistente

Per determinare quali impostazioni di ridimensionamento devi modificare, usa il [describe-scaling-plans](#) comando.

```
aws autoscaling-plans describe-scaling-plans \  
  --scaling-plan-names my-scaling-plan
```

Prendi nota degli elementi che desideri conservare dal piano di ridimensionamento esistente, che può includere quanto segue:

- **MinCapacity**— La capacità minima della risorsa scalabile.
- **MaxCapacity**— La capacità massima della risorsa scalabile.
- **PredefinedLoadMetricType**— Una metrica di carico per la scalabilità predittiva.
- **PredefinedScalingMetricType**— Una metrica di scalabilità per il tracciamento (dinamico) degli obiettivi e il ridimensionamento predittivo.
- **TargetValue**— Il valore target per la metrica di scalabilità.

Differenze tra piani di scalabilità e politiche di scalabilità

Esistono alcune importanti differenze tra i piani di scalabilità e le politiche di scalabilità:

- Una politica di scalabilità può consentire un solo tipo di scalabilità: scalabilità di tracciamento degli obiettivi o scalabilità predittiva. Per utilizzare entrambi i metodi di scalabilità, è necessario creare politiche separate.
- Allo stesso modo, è necessario definire separatamente la metrica di scalabilità per la scalatura predittiva e la metrica di ridimensionamento per il tracciamento degli obiettivi all'interno delle rispettive politiche.

Fase 2: Creare politiche di scalabilità predittiva

Se non utilizzi la scalabilità predittiva, passa a [Fase 4: Prepararsi a eliminare il piano di scalabilità](#)

Per avere il tempo necessario per valutare la previsione, ti consigliamo di creare politiche di scalabilità predittiva prima di altre politiche di scalabilità.

Per qualsiasi gruppo di Auto Scaling con una specifica metrica di carico esistente, procedi come segue per trasformarlo in una politica di scalabilità predittiva basata su Amazon EC2 Auto Scaling.

Per creare politiche di scalabilità predittiva

1. In un file JSON, definisci una `MetricSpecifications` struttura come mostrato nell'esempio seguente:

```
{
```

```

"MetricSpecifications":[
  {
    ...
  }
]
}

```

2. Nella `MetricSpecifications` struttura, per ogni metrica di carico del piano di scalabilità, create `PredefinedLoadMetricSpecification` o `CustomizedLoadMetricSpecification` utilizzate le impostazioni equivalenti del piano di scalabilità.

Di seguito sono riportati alcuni esempi della struttura della sezione metrica del carico.

With predefined metrics

```

{
  "MetricSpecifications":[
    {
      "PredefinedLoadMetricSpecification":{
        "PredefinedMetricType":"ASGTotalCPUUtilization"
      },
      ...
    }
  ]
}

```

Per ulteriori informazioni, consulta il riferimento [PredictiveScalingPredefinedLoadMetric](#) all'API Amazon EC2 Auto Scaling.

With custom metrics

```

{
  "MetricSpecifications":[
    {
      "CustomizedLoadMetricSpecification":{
        "MetricDataQueries":[
          {
            "Id":"load_metric",
            "MetricStat":{
              "Metric":{
                "MetricName":"MyLoadMetric",
                "Namespace":"MyNameSpace",

```

```

        "Dimensions":[
            {
                "Name":"MyOptionalMetricDimensionName",
                "Value":"MyOptionalMetricDimensionValue"
            }
        ],
        "Stat":"Sum"
    }
}
]
}

```

Per ulteriori informazioni, consulta il riferimento

[PredictiveScalingCustomizedLoadMetric](#) all'API Amazon EC2 Auto Scaling.

3. Aggiungi la specifica della metrica di scalabilità a `MetricSpecifications` e definisci un valore target.

Di seguito sono riportati alcuni esempi della struttura delle sezioni relative alla metrica di scala e al valore obiettivo.

With predefined metrics

```

{
  "MetricSpecifications":[
    {
      "PredefinedLoadMetricSpecification":{
        "PredefinedMetricType":"ASGTotalCPUUtilization"
      },
      "PredefinedScalingMetricSpecification":{
        "PredefinedMetricType":"ASGCPUUtilization"
      },
      "TargetValue":50
    }
  ],
  ...
}

```

Per ulteriori informazioni, consulta il riferimento

[PredictiveScalingPredefinedScalingMetric](#) all'API Amazon EC2 Auto Scaling.

With custom metrics

```
{
  "MetricSpecifications": [
    {
      "CustomizedLoadMetricSpecification": {
        "MetricDataQueries": [
          {
            "Id": "load_metric",
            "MetricStat": {
              "Metric": {
                "MetricName": "MyLoadMetric",
                "Namespace": "MyNameSpace",
                "Dimensions": [
                  {
                    "Name": "MyOptionalMetricDimensionName",
                    "Value": "MyOptionalMetricDimensionValue"
                  }
                ]
              },
              "Stat": "Sum"
            }
          }
        ]
      },
      "CustomizedScalingMetricSpecification": {
        "MetricDataQueries": [
          {
            "Id": "scaling_metric",
            "MetricStat": {
              "Metric": {
                "MetricName": "MyUtilizationMetric",
                "Namespace": "MyNameSpace",
                "Dimensions": [
                  {
                    "Name": "MyOptionalMetricDimensionName",
                    "Value": "MyOptionalMetricDimensionValue"
                  }
                ]
              }
            }
          }
        ]
      }
    }
  ]
}
```

```

        "Stat": "Average"
      }
    ]
  },
  "TargetValue": 50
}
],
...
}

```

Per ulteriori informazioni, consulta il riferimento

[PredictiveScalingCustomizedScalingMetric](#) all'API Amazon EC2 Auto Scaling.

4. Solo per la previsione, aggiungi la proprietà `Mode` con un valore di `ForecastOnly`. Dopo aver completato la migrazione della scalabilità predittiva e aver verificato che la previsione sia accurata e affidabile, puoi modificare la modalità per consentire la scalabilità. Per ulteriori informazioni, consulta [Fase 7: Riattivare la scalabilità predittiva](#).

```

{
  "MetricSpecifications": [
    ...
  ],
  "Mode": "ForecastOnly",
  ...
}

```

Per ulteriori informazioni, consulta il riferimento [PredictiveScalingConfiguration](#) all'API Amazon EC2 Auto Scaling.

5. Se la `ScheduledActionBufferTime` proprietà è presente nel tuo piano di scalabilità, copia il suo valore nella `SchedulingBufferTime` proprietà nella tua politica di scalabilità predittiva.

```

{
  "MetricSpecifications": [
    ...
  ],
  "Mode": "ForecastOnly",
  "SchedulingBufferTime": 300,
  ...
}

```

Per ulteriori informazioni, consulta il riferimento [PredictiveScalingConfiguration](#) all'API Amazon EC2 Auto Scaling.

- Se le **PredictiveScalingMaxCapacityBuffer** proprietà **PredictiveScalingMaxCapacityBehavior** and sono presenti nel tuo piano di scalabilità, puoi configurarle nella tua politica di scalabilità predittiva. **MaxCapacityBreachBehavior** **MaxCapacityBuffer** Queste proprietà definiscono cosa dovrebbe succedere se la capacità prevista si avvicina o supera la capacità massima specificata per il gruppo Auto Scaling.

Warning

Se si imposta la **MaxCapacityBreachBehavior** proprietà su **IncreaseMaxCapacity**, potrebbero essere avviate più istanze del previsto, a meno che non si monitori e gestisca l'aumento della capacità massima. L'aumento della capacità massima diventa la nuova capacità massima normale per il gruppo Auto Scaling finché non lo si aggiorna manualmente. La capacità massima non torna automaticamente al massimo originale.

```
{
  "MetricSpecifications": [
    ...
  ],
  "Mode": "ForecastOnly",
  "SchedulingBufferTime": 300,
  "MaxCapacityBreachBehavior": "IncreaseMaxCapacity",
  "MaxCapacityBuffer": 10
}
```

Per ulteriori informazioni, consulta il riferimento [PredictiveScalingConfiguration](#) all'API Amazon EC2 Auto Scaling.

- Salva il file JSON con un nome univoco. Prendi nota del nome del file. Ne hai bisogno nella fase successiva e di nuovo al termine della procedura di migrazione quando riattivi le politiche di scalabilità predittiva. Per ulteriori informazioni, consulta [Fase 7: Riattivare la scalabilità predittiva](#).
- Dopo aver salvato il file JSON, esegui il comando. [put-scaling-policy](#) Nell'esempio seguente, sostituisci ciascuno di essi *user input placeholder* con le tue informazioni.

```
aws autoscaling put-scaling-policy --policy-name my-predictive-scaling-policy \
```

```
--auto-scaling-group-name my-asg --policy-type PredictiveScaling \  
--predictive-scaling-configuration file://my-predictive-scaling-config.json
```

In caso di esito positivo, questo comando restituisce l'Amazon Resource Name (ARN) della policy.

```
{  
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-  
d8a8-4d14-b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/my-predictive-  
scaling-policy",  
  "Alarms": []  
}
```

9. Ripeti questi passaggi per ogni specifica metrica di carico che stai migrando a una politica di scalabilità predittiva basata su Amazon EC2 Auto Scaling.

Fase 3: esamina le previsioni generate dalle politiche di scalabilità predittiva

Se non utilizzi la scalabilità predittiva, salta la procedura seguente.

Una previsione è disponibile subito dopo la creazione di una politica di scalabilità predittiva. Dopo che Amazon EC2 Auto Scaling ha generato la previsione, puoi rivedere la previsione per la policy tramite la console Amazon EC2 Auto Scaling e modificarla se necessario.

Per esaminare la previsione di una politica di scalabilità predittiva

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel riquadro di navigazione, scegli Gruppi Auto Scaling, quindi scegli il nome del tuo gruppo Auto Scaling dall'elenco.
3. Nella scheda Scalabilità automatica, in Politiche di scalabilità predittiva, scegli la tua politica.
4. Nella sezione Monitoraggio, puoi visualizzare le previsioni passate e future della policy in termini di carico e capacità rispetto ai valori effettivi.

Per ulteriori informazioni, [consulta i grafici di monitoraggio della scalabilità predittiva](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

5. Ripeti questi passaggi per ogni policy di scalabilità predittiva che hai creato.

Fase 4: Prepararsi a eliminare il piano di scalabilità

Per tutte le risorse con una configurazione esistente di Target Tracking Scaling, procedi come segue per raccogliere tutte le informazioni aggiuntive di cui hai bisogno dal piano di scalabilità prima di eliminarlo.

Per descrivere le informazioni sulla politica di scalabilità contenute nel piano di scalabilità, utilizza il comando [describe-scaling-plan-resources](#). Nel comando di esempio seguente, sostituisci *my-scaling-plan* con le tue informazioni.

```
aws autoscaling-plans describe-scaling-plan-resources \  
  --scaling-plan-name my-scaling-plan \  
  --scaling-plan-version 1
```

Esamina l'output e conferma di voler migrare le politiche di scalabilità descritte. Utilizza queste informazioni per creare nuove policy di scalabilità di tracciamento degli obiettivi basate su Amazon EC2 Auto Scaling e Application Auto Scaling in [Fase 6: Riattivare la scalabilità dinamica](#)

Fase 5: Eliminare il piano di scalabilità

Prima di creare nuove politiche di scalabilità di Target Tracking, è necessario eliminare il piano di scalabilità per eliminare le politiche di scalabilità da esso create.

Per eliminare il piano di scalabilità, utilizzare il comando [delete-scaling-plan](#). Nel seguente comando di esempio, sostituiscilo *my-scaling-plan* con le tue informazioni.

```
aws autoscaling-plans delete-scaling-plan \  
  --scaling-plan-name my-scaling-plan \  
  --scaling-plan-version 1
```

Dopo aver eliminato il piano di ridimensionamento, il ridimensionamento dinamico viene disattivato. Pertanto, se si verificano improvvisi picchi di traffico o di carico di lavoro, la capacità disponibile per ogni risorsa scalabile non aumenterà da sola. Per precauzione, potresti voler aumentare manualmente la capacità delle tue risorse scalabili a breve termine.

Per aumentare la capacità di un gruppo Auto Scaling

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.

2. Nel riquadro di navigazione, scegli Gruppi Auto Scaling, quindi scegli il nome del tuo gruppo Auto Scaling dall'elenco.
3. Nella scheda Details (Dettagli) scegliere Group details (Dettagli gruppo), Edit (Modifica).
4. Per Capacità desiderata, aumenta la capacità desiderata.
5. Quando hai finito, scegli Aggiorna.

Per aggiungere una replica di Aurora a un cluster di database

1. Aprire la console Amazon RDS all'indirizzo <https://console.aws.amazon.com/rds/>.
2. Nel riquadro di navigazione, scegli Database, quindi seleziona il tuo cluster DB.
3. Assicurarsi che sia il cluster che l'istanza primaria siano nello stato Available (Disponibile) .
4. Scegli Azioni, Aggiungi lettore.
5. Nella pagina Aggiungi lettore, specifica le opzioni per la tua nuova replica Aurora.
6. Scegli Aggiungi lettore.

Per aumentare la capacità di lettura e scrittura fornita di una tabella DynamoDB o di un indice secondario globale

1. Apri la console DynamoDB all'indirizzo. <https://console.aws.amazon.com/dynamodb/>
2. Nel pannello di navigazione, scegli Tabelle, quindi scegli il nome della tabella dall'elenco.
3. Nella scheda Impostazioni aggiuntive, scegli Capacità di lettura/scrittura, Modifica.
4. Nella pagina Modifica read/write capacità, per Capacità di lettura, Unità di capacità assegnata, aumenta la capacità di lettura assegnata della tabella.
5. (Facoltativo) Se desideri che gli indici secondari globali utilizzino le stesse impostazioni di capacità di lettura della tabella di base, seleziona la casella di controllo Usa le stesse impostazioni di capacità di lettura per tutti gli indici secondari globali.
6. Per Capacità di scrittura, Unità di capacità assegnata, aumenta la capacità di scrittura assegnata della tabella.
7. (Facoltativo) Se desideri che gli indici secondari globali utilizzino le stesse impostazioni di capacità di scrittura della tabella di base, seleziona la casella di controllo Usa le stesse impostazioni di capacità di scrittura per tutti gli indici secondari globali.
8. Se non hai selezionato le caselle di controllo nei passaggi 5 o 7, scorri la pagina verso il basso per aggiornare la capacità di lettura e scrittura di qualsiasi indice secondario globale.

9. Scegli Salva modifiche per continuare.

Per aumentare il numero di attività in esecuzione per il tuo servizio Amazon ECS

1. [Apri la console nella versione 2](https://console.aws.amazon.com/ecs/)<https://console.aws.amazon.com/ecs/>.
2. Nel riquadro di navigazione, scegli Cluster, quindi scegli il nome del cluster dall'elenco.
3. Nella sezione Servizi, seleziona la casella di controllo accanto al servizio, quindi scegli Aggiorna.
4. In Attività desiderate, inserisci il numero di attività che desideri eseguire per il servizio.
5. Scegliere Aggiorna.

Per aumentare la capacità di una flotta Spot

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel riquadro di navigazione, scegli Richieste Spot, quindi seleziona la tua richiesta Spot Fleet.
3. Scegliere Actions (Operazioni), quindi Modify target capacity (Modifica capacità target).
4. In Modifica la capacità target, inserisci la nuova parte relativa alla capacità target e all'istanza On-Demand.
5. Seleziona Invia.

Fase 6: Riattivare la scalabilità dinamica

Riattiva il ridimensionamento dinamico creando politiche di scalabilità mirate al tracciamento.

Quando si crea una politica di ridimensionamento del tracciamento degli obiettivi per un gruppo Auto Scaling, la si aggiunge direttamente al gruppo. Quando si crea una politica di ridimensionamento del tracciamento degli obiettivi per altre risorse scalabili, si registra innanzitutto la risorsa come destinazione scalabile e quindi si aggiunge una politica di ridimensionamento del tracciamento degli obiettivi alla destinazione scalabile.

Argomenti

- [Crea politiche di scalabilità per il tracciamento degli obiettivi per i gruppi Auto Scaling](#)
- [Crea politiche di scalabilità di tracciamento degli obiettivi per altre risorse scalabili](#)

Crea politiche di scalabilità per il tracciamento degli obiettivi per i gruppi Auto Scaling

Per creare politiche di ridimensionamento del tracciamento degli obiettivi per i gruppi di Auto Scaling

1. In un file JSON, crea `PredefinedMetricSpecification` o `CustomizedMetricSpecification` utilizza le impostazioni equivalenti del piano di ridimensionamento.

Di seguito sono riportati alcuni esempi di configurazione di tracciamento del target. In questi esempi, sostituisci ciascuno *user input placeholder* con le tue informazioni.

With predefined metrics

```
{
  "TargetValue": 50.0,
  "PredefinedMetricSpecification":
    {
      "PredefinedMetricType": "ASGAverageCPUUtilization"
    }
}
```

Per ulteriori informazioni, consulta il riferimento [PredefinedMetricSpecification](#) all'API Amazon EC2 Auto Scaling.

With custom metrics

```
{
  "TargetValue": 100.0,
  "CustomizedMetricSpecification": {
    "MetricName": "MyBacklogPerInstance",
    "Namespace": "MyNamespace",
    "Dimensions": [{
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }],
    "Statistic": "Average",
    "Unit": "None"
  }
}
```

Per ulteriori informazioni, consulta il riferimento [CustomizedMetricSpecification](#) all'API Amazon EC2 Auto Scaling.

2. Per creare la tua politica di scalabilità, usa il [put-scaling-policy](#) comando, insieme al file JSON creato nel passaggio precedente. Nell'esempio seguente, sostituisci ciascuno di essi *user input placeholder* con le tue informazioni.

```
aws autoscaling put-scaling-policy --policy-name my-target-tracking-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
```

3. Ripeti questo processo per ogni politica di scalabilità basata su un piano di scalabilità che stai migrando a una politica di scalabilità di tracciamento degli obiettivi basata su Amazon EC2 Auto Scaling.

Crea politiche di scalabilità di tracciamento degli obiettivi per altre risorse scalabili

Successivamente, crea politiche di scalabilità di tracciamento degli obiettivi per altre risorse scalabili eseguendo le seguenti attività di configurazione.

- Registra un target scalabile per la scalabilità automatica con il servizio Application Auto Scaling.
- Aggiungi una policy di dimensionamento con monitoraggio della destinazione sulla destinazione dimensionabile.

Per creare politiche di scalabilità di tracciamento degli obiettivi per altre risorse scalabili

1. Utilizzate il [register-scalable-target](#) comando per registrare la risorsa come destinazione scalabile e definire i limiti di scalabilità per la politica di scalabilità.

Nell'esempio seguente, sostituisci ciascuno *user input placeholder* con le tue informazioni. Per le opzioni di comando, fornite le seguenti informazioni:

- `--service-namespace`— Uno spazio dei nomi per il servizio di destinazione (ad esempio, `ecs`). Per ottenere gli spazi dei nomi dei servizi, vedere il riferimento.

[RegisterScalableTarget](#)

- `--scalable-dimension`— Una dimensione scalabile associata alla risorsa di destinazione (ad esempio, `ecs:service:DesiredCount`). Per ottenere dimensioni scalabili, vedere il [RegisterScalableTarget](#) riferimento.
- `--resource-id`— Un ID di risorsa per la risorsa di destinazione (ad esempio, `service/my-cluster/my-service`). Per informazioni sulla sintassi e sugli esempi di risorse specifiche IDs, consultate il [RegisterScalableTarget](#) riferimento.

```
aws application-autoscaling register-scalable-target --service-namespace namespace \
  --scalable-dimension dimension \
  --resource-id identifier \
  --min-capacity 1 --max-capacity 10
```

In caso di esito positivo, il comando restituisce l'ARN dell'obiettivo scalabile.

```
{
  "ScalableTargetARN": "arn:aws:application-autoscaling:region:account-id:scalable-target/1234abcd56ab78cd901ef1234567890ab123"
}
```

2. In un file JSON, create un `PredefinedMetricSpecification` o `CustomizedMetricSpecification` utilizzando le impostazioni equivalenti del piano di scala.

Di seguito sono riportati alcuni esempi di configurazione di tracciamento del target.

With predefined metrics

```
{
  "TargetValue": 70.0,
  "PredefinedMetricSpecification":
    {
      "PredefinedMetricType": "ECSServiceAverageCPUUtilization"
    }
}
```

Per ulteriori informazioni, consulta [PredefinedMetricSpecification](#) l'Application Auto Scaling API Reference.

With custom metrics

```
{
  "TargetValue": 70.0,
  "CustomizedMetricSpecification": {
    "MetricName": "MyUtilizationMetric",
    "Namespace": "MyNamespace",
    "Dimensions": [{
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }],
    "Statistic": "Average",
    "Unit": "Percent"
  }
}
```

Per ulteriori informazioni, consulta [CustomizedMetricSpecification](#) nell'Application Auto Scaling API Reference.

3. Per creare la tua politica di scalabilità, usa il [put-scaling-policy](#) comando insieme al file JSON creato nel passaggio precedente.

```
aws application-autoscaling put-scaling-policy --service-namespace namespace \
  --scalable-dimension dimension \
  --resource-id identifier \
  --policy-name my-target-tracking-scaling-policy --policy-
type TargetTrackingScaling \
  --target-tracking-scaling-policy-configuration file://config.json
```

4. Ripeti questo processo per ogni policy di scalabilità basata su un piano di scalabilità che stai migrando a una politica di scalabilità di tracciamento del target basata su Application Auto Scaling.

Fase 7: Riattivare la scalabilità predittiva

Se non utilizzi la scalabilità predittiva, salta questo passaggio.

Riattiva la scalabilità predittiva passando alla scalabilità predittiva alla previsione e alla scalabilità.

Per apportare questa modifica, aggiorna i file JSON che hai creato [Fase 2: Creare politiche di scalabilità predittiva](#) e modifica il valore dell'opzione come nell'esempio seguente:

ForecastAndScale

```
"Mode": "ForecastAndScale"
```

Quindi, aggiorna ogni politica di scalabilità predittiva con il comando. [put-scaling-policy](#) In questo esempio, sostituisci ognuna *user input placeholder* con le tue informazioni.

```
aws autoscaling put-scaling-policy --policy-name my-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://my-predictive-scaling-config.json
```

In alternativa, puoi apportare questa modifica dalla console Amazon EC2 Auto Scaling attivando l'impostazione Scale based on forecast. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Riferimento di Amazon EC2 Auto Scaling per la migrazione delle politiche di scalabilità di tracciamento degli obiettivi

A scopo di riferimento, la tabella seguente elenca tutte le proprietà di configurazione di tracciamento del target nel piano di scalabilità con le proprietà corrispondenti nell'operazione dell'API Amazon EC2 Auto PutScalingPolicy Scaling.

Proprietà di origine del piano di scalabilità	Proprietà di destinazione di Amazon EC2 Auto Scaling
PolicyName	PolicyName
PolicyType	PolicyType
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Name	TargetTrackingConfiguration.CustomizedMetricSpecification.Dimensions.Name

Proprietà di origine del piano di scalabilità	Proprietà di destinazione di Amazon EC2 Auto Scaling
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Value	TargetTrackingConfiguration.CustomizedMetricSpecification.Dimensions.Value
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.MetricName	TargetTrackingConfiguration.CustomizedMetricSpecification.MetricName
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Namespace	TargetTrackingConfiguration.CustomizedMetricSpecification.Namespace
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Statistic	TargetTrackingConfiguration.CustomizedMetricSpecification.Statistic
TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Unit	TargetTrackingConfiguration.CustomizedMetricSpecification.Unit
TargetTrackingConfiguration.DisableScaleIn	TargetTrackingConfiguration.DisableScaleIn
TargetTrackingConfiguration.EstimatedInstanceWarmup	TargetTrackingConfiguration.EstimatedInstanceWarmup ¹
TargetTrackingConfiguration.PredefinedScalingMetricSpecification.PredefinedScalingMetricType	TargetTrackingConfiguration.PredefinedMetricSpecification.PredefinedMetricType
TargetTrackingConfiguration.PredefinedScalingMetricSpecification.ResourceLabel	TargetTrackingConfiguration.PredefinedMetricSpecification.ResourceLabel

Proprietà di origine del piano di scalabilità	Proprietà di destinazione di Amazon EC2 Auto Scaling
TargetTrackingConfiguration .ScaleInCooldown	Not available
TargetTrackingConfiguration .ScaleOutCooldown	Not available
TargetTrackingConfiguration .TargetValue	TargetTrackingConfiguration .TargetValue

¹ Instance Warmup è una funzionalità per i gruppi di Auto Scaling che aiuta a garantire che le istanze appena lanciate siano pronte a ricevere traffico prima di contribuire con i dati di utilizzo alla metrica di scalabilità. Mentre le istanze sono ancora in fase di riscaldamento, Amazon EC2 Auto Scaling rallenta il processo di aggiunta o rimozione di istanze al gruppo. Invece di specificare un tempo di riscaldamento per una politica di scalabilità, ti consigliamo di utilizzare l'impostazione predefinita di riscaldamento dell'istanza del tuo gruppo Auto Scaling per assicurarti che tutti gli avvii delle istanze utilizzino lo stesso tempo di riscaldamento dell'istanza. Per ulteriori informazioni, consulta [Impostazione del riscaldamento predefinito dell'istanza per un gruppo con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Riferimento di Application Auto Scaling per la migrazione delle politiche di scalabilità di Target Tracking

A scopo di riferimento, la tabella seguente elenca tutte le proprietà di configurazione di tracciamento del target nel piano di scalabilità con le proprietà corrispondenti nell'operazione API Application Auto PutScalingPolicy Scaling.

Proprietà di origine del piano di scalabilità	Proprietà di destinazione Application Auto Scaling
PolicyName	PolicyName
PolicyType	PolicyType

Proprietà di origine del piano di scalabilità	Proprietà di destinazione Application Auto Scaling
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.Dimensions.Name	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.Dimensions .Name
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.Dimensions.Value	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.Dimensions .Value
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.MetricName	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.MetricName
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.Namespace	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.Namespace
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.Statistic	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.Statistic
TargetTrackingConfiguration .CustomizedScalingMetricSpecification.Unit	TargetTrackingScalingPolicy Configuration.CustomizedMetricSpecification.Unit
TargetTrackingConfiguration .DisableScaleIn	TargetTrackingScalingPolicy Configuration.DisableScaleIn
TargetTrackingConfiguration .EstimatedInstanceWarmup	Not available

Proprietà di origine del piano di scalabilità	Proprietà di destinazione Application Auto Scaling
TargetTrackingConfiguration.PredefinedScalingMetricSpecification.PredefinedScalingMetricType	TargetTrackingScalingPolicyConfiguration.PredefinedMetricSpecification.PredefinedMetricType
TargetTrackingConfiguration.PredefinedScalingMetricSpecification.ResourceLabel	TargetTrackingScalingPolicyConfiguration.PredefinedMetricSpecification.ResourceLabel
TargetTrackingConfiguration.ScaleInCooldown ¹	TargetTrackingScalingPolicyConfiguration.ScaleInCooldown
TargetTrackingConfiguration.ScaleOutCooldown ¹	TargetTrackingScalingPolicyConfiguration.ScaleOutCooldown
TargetTrackingConfiguration.TargetValue	TargetTrackingScalingPolicyConfiguration.TargetValue

¹ Application Auto Scaling utilizza i periodi di cooldown per rallentare la scalabilità quando la risorsa scalabile viene scalata verso l'esterno (aumento della capacità) e scalabile verso l'interno (riduzione della capacità). Per ulteriori informazioni, consulta [Definizione dei tempi di raffreddamento](#) nella Guida per l'utente di Application Auto Scaling.

Informazioni aggiuntive

Per informazioni su come creare nuove policy di scalabilità predittiva dalla console, consulta il seguente argomento:

- Amazon EC2 Auto [Scaling: crea una politica di scalabilità predittiva](#) nella Amazon EC2 Auto Scaling User Guide.

Per informazioni su come creare nuove politiche di scalabilità di tracciamento degli obiettivi utilizzando la console, consulta i seguenti argomenti:

- Amazon Aurora: [utilizzo di Amazon Aurora Auto Scaling con repliche Aurora](#) nella Guida per l'utente di Amazon RDS.
- DynamoDB : [utilizzo della scalabilità automatica con Console di gestione AWS DynamoDB nella Amazon DynamoDB Developer Guide.](#)
- Amazon EC2 Auto [Scaling: crea una politica di scalabilità mirata con tracciamento](#) nella Amazon EC2 Auto Scaling User Guide.
- Amazon ECS: [aggiornamento di un servizio utilizzando la console](#) nell'Amazon Elastic Container Service Developer Guide.
- Flotta Spot: [ridimensiona la flotta Spot utilizzando una politica di tracciamento degli obiettivi](#) nella Guida per l'utente di Amazon EC2.

Registrazione delle chiamate AWS Auto Scaling API con AWS CloudTrail

AWS Auto Scaling è integrato con AWS CloudTrail, un servizio che fornisce una registrazione delle azioni intraprese da un utente, ruolo o AWS servizio in AWS Auto Scaling. CloudTrail acquisisce tutte le chiamate API AWS Auto Scaling come eventi. Le chiamate acquisite includono chiamate dalla AWS Auto Scaling console e chiamate di codice all' AWS Auto Scaling API. Se crei un trail, puoi abilitare la distribuzione continua di CloudTrail eventi a un bucket Amazon S3, inclusi gli eventi per. AWS Auto Scaling Se non configuri un percorso, puoi comunque visualizzare gli eventi più recenti nella CloudTrail console nella cronologia degli eventi. Utilizzando le informazioni raccolte da CloudTrail, puoi determinare a quale richiesta è stata inviata AWS Auto Scaling, l'indirizzo IP da cui è stata effettuata la richiesta, chi ha effettuato la richiesta, quando è stata effettuata e dettagli aggiuntivi.

Per ulteriori informazioni CloudTrail, consulta la [Guida AWS CloudTrail per l'utente](#).

AWS Auto Scaling Informazioni in CloudTrail

CloudTrail è abilitato sul tuo AWS account al momento della creazione dell'account. Quando si verifica AWS Auto Scaling un'attività, tale attività viene registrata in un CloudTrail evento insieme ad altri eventi AWS di servizio nella cronologia degli eventi. Puoi visualizzare, cercare e scaricare gli eventi recenti nel tuo AWS account. Per ulteriori informazioni, consulta [Visualizzazione degli eventi con la cronologia degli CloudTrail eventi](#).

Per una registrazione continua degli eventi nel tuo AWS account, inclusi gli eventi di AWS Auto Scaling, crea un percorso. Un trail consente di CloudTrail inviare file di log a un bucket Amazon S3. Per impostazione predefinita, quando si crea un percorso nella console, questo sarà valido in tutte le Regioni AWS. Il trail registra gli eventi di tutte le regioni della AWS partizione e consegna i file di log al bucket Amazon S3 specificato. Inoltre, puoi configurare altri Amazon Web Services per analizzare ulteriormente e agire in base ai dati sugli eventi raccolti nei CloudTrail log. Per ulteriori informazioni, consulta gli argomenti seguenti:

- [Panoramica della creazione di un trail](#)
- [CloudTrail Servizi e integrazioni supportati](#)
- [Configurazione delle notifiche Amazon SNS per CloudTrail](#)

- [Ricezione di file di CloudTrail registro da più regioni](#) e [ricezione di file di CloudTrail registro da più account](#)

Tutte AWS Auto Scaling le azioni vengono registrate CloudTrail e documentate nell'[AWS Auto Scaling API Reference](#). Ad esempio, le chiamate a `CreateScalingPlanDeleteScalingPlan`, e `DescribeScalingPlans` le azioni generano voci nei file di CloudTrail registro.

Ogni evento o voce di log contiene informazioni sull'utente che ha generato la richiesta. Le informazioni di identità consentono di determinare quanto segue:

- Se la richiesta è stata effettuata con credenziali utente root o AWS Identity and Access Management (IAM).
- Se la richiesta è stata effettuata con le credenziali di sicurezza temporanee per un ruolo o un utente federato.
- Se la richiesta è stata effettuata da un altro AWS servizio.

Per ulteriori informazioni, consulta [Elemento CloudTrail userIdentity](#).

Comprensione delle AWS Auto Scaling voci dei file di registro

Un trail è una configurazione che consente la distribuzione di eventi come file di log in un bucket Amazon S3 specificato dall'utente. CloudTrail i file di registro contengono una o più voci di registro. Un evento rappresenta una singola richiesta proveniente da qualsiasi fonte e include informazioni sull'azione richiesta, la data e l'ora dell'azione, i parametri della richiesta e così via. CloudTrail i file di registro non sono una traccia ordinata dello stack delle chiamate API pubbliche, quindi non vengono visualizzati in un ordine specifico.

L'esempio seguente mostra una voce di CloudTrail registro che illustra l'`CreateScalingPlan`azione.

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "Root",
    "principalId": "123456789012",
    "arn": "arn:aws:iam::123456789012:root",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
```

```
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2018-08-21T17:05:42Z"
      }
    }
  },
  "eventTime": "2018-08-01T23:17:19Z",
  "eventSource": "autoscaling.amazonaws.com",
  "eventName": "CreateScalingPlan",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "72.21.196.68",
  "userAgent": "aws-internal/3",
  "requestParameters": {
    "applicationSource": {
      "tagFilters": [
        {
          "key": "TagText",
          "values": [
            "MyApplication"
          ]
        }
      ]
    }
  },
  "scalingInstructions": [
    {
      "resourceId": "autoScalingGroup/MyAutoScalingGroup",
      "targetTrackingConfigurations": [
        {
          "predefinedScalingMetricSpecification": {
            "predefinedScalingMetricType": "ASGAverageCPUUtilization"
          },
          "targetValue": 40
        }
      ],
      "maxCapacity": 10,
      "serviceNamespace": "autoscaling",
      "scalableDimension": "autoscaling:autoScalingGroup:DesiredCapacity",
      "minCapacity": 1
    }
  ],
  "scalingPlanName": "MyScalingPlan"
},
"responseElements": {
```

```
    "scalingPlanVersion": 1
  },
  "additionalEventData": {
    "service": "autoscaling-plans"
  },
  "requestID": "0737e2ea-fb2d-11e3-bfd8-99133058e7bb",
  "eventID": "3fcfb182-98f8-4744-bd45-b38835ab61cb",
  "eventType": "AwsApiCall",
  "recipientAccountId": "123456789012"
}
```

Sicurezza del piano di dimensionamento

La sicurezza del cloud AWS è la massima priorità. In qualità di AWS cliente, puoi beneficiare di un data center e di un'architettura di rete progettati per soddisfare i requisiti delle organizzazioni più sensibili alla sicurezza.

La sicurezza è una responsabilità condivisa tra AWS te e te. Il [modello di responsabilità condivisa](#) descrive questo aspetto come sicurezza del cloud e sicurezza nel cloud:

- **Sicurezza del cloud:** AWS è responsabile della protezione dell'infrastruttura che gestisce AWS i servizi nel AWS cloud. AWS ti fornisce anche servizi che puoi utilizzare in modo sicuro. Third-party Per ulteriori informazioni sui programmi di conformità applicabili AWS Auto Scaling, consulta la sezione [AWS Servizi rientranti nell'ambito dei programmi di conformità](#) ().
- **Sicurezza nel cloud:** la tua responsabilità è determinata dal AWS servizio che utilizzi. Inoltre, sei responsabile anche di altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda e le leggi e le normative applicabili.

Questa documentazione aiuta a comprendere come applicare il modello di responsabilità condivisa quando si utilizzano piani di dimensionamento e aiuta a comprendere come gestirne l'accesso.

Argomenti

- [Accedi ai piani di scalabilità utilizzando gli endpoint VPC dell'interfaccia](#)
- [Protezione dei dati per piani di scalabilità](#)
- [Gestione delle identità e degli accessi per piani di scalabilità](#)
- [Convalida della conformità per i piani di scalabilità](#)
- [Sicurezza dell'infrastruttura per piani di scalabilità](#)

Accedi ai piani di scalabilità utilizzando gli endpoint VPC dell'interfaccia

Puoi usarlo AWS PrivateLink per creare una connessione privata tra il tuo VPC e. AWS Auto Scaling Puoi accedere AWS Auto Scaling come se fosse nel tuo VPC, senza l'uso di un gateway Internet, un dispositivo NAT, una connessione VPN o una connessione. Direct Connect Le istanze del tuo VPC non necessitano di indirizzi IP pubblici per accedervi. AWS Auto Scaling

Stabilisci questa connessione privata creando un endpoint di interfaccia attivato da AWS PrivateLink. In ciascuna sottorete viene creata un'interfaccia di rete endpoint da abilitare per l'endpoint di interfaccia. Queste sono interfacce di rete gestite dal richiedente che fungono da punto di ingresso per il traffico destinato a AWS Auto Scaling.

Per ulteriori informazioni, consulta [Access Servizi AWS through AWS PrivateLink](#) nella AWS PrivateLink Guida.

Argomenti

- [Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento](#)
- [Crea una policy dell'endpoint VPC per i piani di dimensionamento](#)
- [Migrazione degli endpoint](#)

Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento

Crea un endpoint per i piani AWS Auto Scaling di scalabilità utilizzando il seguente nome di servizio:

```
com.amazonaws.region.autoscaling-plans
```

Per ulteriori informazioni, consulta [Accedere a un AWS servizio utilizzando un endpoint VPC di interfaccia nella Guida](#).AWS PrivateLink

Non è necessario modificare altre impostazioni. AWS Auto Scaling Le altre chiamate API Servizi AWS utilizzano endpoint di servizio o endpoint VPC con interfaccia privata, a seconda di quale dei due siano in uso.

Crea una policy dell'endpoint VPC per i piani di dimensionamento

Puoi allegare una policy al tuo endpoint VPC per controllare l'accesso all'API. AWS Auto Scaling La policy specifica:

- Il principale che può eseguire operazioni.
- Le operazioni che possono essere eseguite.
- La risorsa su cui è possibile eseguire le operazioni.

Nell'esempio seguente viene illustrata una policy di endpoint VPC che nega a chiunque l'autorizzazione per eliminare un piano di dimensionamento tramite l'endpoint. Inoltre, la policy di esempio concede a chiunque l'autorizzazione per eseguire tutte le altre operazioni.

```
{
  "Statement": [
    {
      "Action": "*",
      "Effect": "Allow",
      "Resource": "*",
      "Principal": "*"
    },
    {
      "Action": "autoscaling-plans:DeleteScalingPlan",
      "Effect": "Deny",
      "Resource": "*",
      "Principal": "*"
    }
  ]
}
```

Per ulteriori informazioni, consulta [Policy dell'endpoint VPC](#) nella Guida di AWS PrivateLink .

Migrazione degli endpoint

Il 22 novembre 2019, abbiamo introdotto `autoscaling-plans.region.amazonaws.com` come nuovo nome host ed endpoint DNS predefiniti per le chiamate all'API. AWS Auto Scaling Il nuovo endpoint è compatibile con l'ultima versione di and SDK. AWS CLI Se non l'hai già fatto, installa la versione più recente AWS CLI e gli SDK per utilizzare il nuovo endpoint. Per aggiornare il AWS CLI, consulta [Installazione o aggiornamento di AWS CLI nella Guida](#) per l'AWS Command Line Interface utente. Per informazioni sugli AWS SDK, consulta [Tools for Amazon Web Services](#).

Important

Per motivi di compatibilità con le versioni precedenti, l'`autoscaling.region.amazonaws.com` endpoint esistente continuerà a essere supportato per le chiamate all'API. AWS Auto Scaling Per configurare l'endpoint `autoscaling.region.amazonaws.com` come endpoint VPC di interfaccia privato, consulta [Amazon EC2 Auto Scaling ed endpoint VPC di interfaccia](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Endpoint da chiamare quando si utilizza la CLI o l'API AWS Auto Scaling

Nella versione corrente di AWS Auto Scaling, le chiamate all' AWS Auto Scaling API vengono inviate automaticamente all'`autoscaling-plans.region.amazonaws.com` anziché `autoscaling.region.amazonaws.com`

È possibile richiamare il nuovo endpoint nell'interfaccia a riga di comando (CLI) utilizzando il parametro seguente con ogni comando per specificare l'endpoint: `--endpoint-url https://autoscaling-plans.region.amazonaws.com`.

Sebbene non sia consigliato, è inoltre possibile richiamare il vecchio endpoint nell'interfaccia a riga di comando (CLI) utilizzando il parametro seguente per specificare l'endpoint: `--endpoint-url https://autoscaling.region.amazonaws.com`.

Per i vari SDK utilizzati per richiamare le API, consultare la documentazione relativa all'SDK di interesse per informazioni su come indirizzare le richieste a un endpoint specifico. Per ulteriori informazioni, consulta [Strumenti per Amazon Web Services](#).

Protezione dei dati per piani di scalabilità

Il modello di [responsabilità AWS condivisa Il modello](#) di si applica alla protezione dei dati in AWS Auto Scaling. Come descritto in questo modello, AWS è responsabile della protezione dell'infrastruttura globale che gestisce tutti i Cloud AWS. L'utente è responsabile del controllo dei contenuti ospitati su questa infrastruttura. L'utente è inoltre responsabile della configurazione della protezione e delle attività di gestione per i Servizi AWS utilizzati. Per ulteriori informazioni sulla privacy dei dati, consulta [Domande frequenti sulla privacy dei dati](#) . Per informazioni sulla protezione dei dati in Europa, consulta il [General Data Protection Regulation \(GDPR\) Center](#).

Ai fini della protezione dei dati, ti consigliamo di proteggere Account AWS le credenziali e configurare i singoli utenti con AWS IAM Identity Center or AWS Identity and Access Management (IAM). In tal modo, a ogni utente verranno assegnate solo le autorizzazioni necessarie per svolgere i suoi compiti. Suggeriamo, inoltre, di proteggere i dati nei seguenti modi:

- Utilizza l'autenticazione a più fattori (MFA) con ogni account.
- SSL/TLS Da utilizzare per comunicare con AWS le risorse. È richiesto TLS 1.2 ed è consigliato TLS 1.3.
- Configura l'API e la registrazione delle attività degli utenti con AWS CloudTrail. Per informazioni sull'utilizzo dei CloudTrail percorsi per acquisire AWS le attività, consulta [Lavorare con i CloudTrail percorsi](#) nella Guida per l'AWS CloudTrail utente.

- Utilizza soluzioni di AWS crittografia, insieme a tutti i controlli di sicurezza predefiniti all'interno Servizi AWS.
- Utilizza i servizi di sicurezza gestiti avanzati, come Amazon Macie, che aiutano a individuare e proteggere i dati sensibili archiviati in Amazon S3.
- Se hai bisogno di moduli crittografici convalidati FIPS 140-3 per accedere AWS tramite un'interfaccia a riga di comando o un'API, usa un endpoint FIPS. Per ulteriori informazioni sugli endpoint FIPS disponibili, consulta il [Federal Information Processing Standard \(FIPS\) 140-3](#).

Ti consigliamo di non inserire mai informazioni riservate o sensibili, ad esempio gli indirizzi e-mail dei clienti, nei tag o nei campi di testo in formato libero, ad esempio nel campo Nome. Ciò include quando lavori o Servizi AWS utilizzi la console, l'API AWS Auto Scaling o gli SDK. AWS CLI AWS I dati inseriti nei tag o nei campi di testo in formato libero utilizzati per i nomi possono essere utilizzati per i la fatturazione o i log di diagnostica. Quando si fornisce un URL a un server esterno, suggeriamo vivamente di non includere informazioni sulle credenziali nell'URL per convalidare la richiesta al server.

Gestione delle identità e degli accessi per piani di scalabilità

AWS Identity and Access Management (IAM) è uno strumento Servizio AWS che aiuta un amministratore a controllare in modo sicuro l'accesso alle AWS risorse. Gli amministratori IAM controllano chi può essere autenticato (effettuato l'accesso) e autorizzato (disporre delle autorizzazioni) a utilizzare le risorse. AWS Auto Scaling IAM è uno Servizio AWS strumento che puoi utilizzare senza costi aggiuntivi.

Per la documentazione IAM completa, consulta la [Guida per l'utente IAM](#).

Controllo accessi

Per autenticare le richieste, è necessario disporre di credenziali valide, ma a meno che non si disponga delle autorizzazioni non è possibile creare o accedere ai piani di dimensionamento. Ad esempio, è necessario disporre delle autorizzazioni per creare piani di dimensionamento, configurare il dimensionamento predittivo e così via.

Le sezioni seguenti forniscono i dettagli su come un amministratore IAM può utilizzare IAM per proteggere i piani di dimensionamento, controllando chi può lavorarci.

Argomenti

- [Come funzionano i piani di dimensionamento con IAM](#)
- [Ruolo collegato al servizio di dimensionamento predittivo](#)
- [Identity-based esempi di politiche per i piani di scalabilità](#)

Come funzionano i piani di dimensionamento con IAM

Prima di utilizzare IAM per gestire chi può creare, accedere e gestire i piani di AWS Auto Scaling scalabilità, è necessario comprendere quali funzionalità IAM sono disponibili per l'uso con i piani di scalabilità.

Argomenti

- [Identity-based politiche](#)
- [Resource-based politiche](#)
- [Liste di controllo degli accessi \(ACL\)](#)
- [Autorizzazione basata su tag](#)
- [Ruoli IAM](#)

Identity-based politiche

Con le policy IAM basate su identità, puoi specificare operazioni e risorse consentite o rifiutate, nonché le condizioni in base alle quali le operazioni sono consentite o rifiutate. I piani di dimensionamento supportano operazioni, risorse e chiavi di condizione specifiche. Per informazioni su tutti gli elementi utilizzati in una policy JSON, consulta [Documentazione di riferimento agli elementi delle policy JSON IAM](#) nella Guida per l'utente IAM.

Azioni

Gli amministratori possono utilizzare le policy AWS JSON per specificare chi ha accesso a cosa. In altre parole, quale entità principale può eseguire operazioni su quali risorse e in quali condizioni.

L'elemento `Action` di una policy JSON descrive le operazioni che è possibile utilizzare per consentire o negare l'accesso in una policy. Includere le operazioni in una policy per concedere le autorizzazioni di eseguire l'operazione associata.

Le operazioni del piano di dimensionamento nelle istruzioni della policy IAM utilizzano il seguente prefisso prima dell'operazione: `autoscaling-plans:`. Le istruzioni della policy devono includere un

elemento `Action` o `NotAction`. I piani di dimensionamento hanno un proprio set di operazioni che descrivono le attività che è possibile eseguire con questo servizio.

Per specificare più operazioni in una singola istruzione, separale con virgole, come illustrato nell'esempio seguente.

```
"Action": [  
    "autoscaling-plans:DescribeScalingPlans",  
    "autoscaling-plans:DescribeScalingPlanResources"
```

Puoi specificare più operazioni tramite caratteri jolly (*). Ad esempio, per specificare tutte le operazioni che iniziano con la parola `Describe`, includi la seguente operazione.

```
"Action": "autoscaling-plans:Describe*"
```

Per visualizzare un elenco completo delle operazioni del piano di dimensionamento che possono essere utilizzate nelle istruzioni della policy, consulta [Operazioni, risorse e chiavi di condizione per AWS Auto Scaling](#) in Service Authorization Reference.

Resources

L'elemento `Resource` specifica l'oggetto o gli oggetti ai quali si applica l'operazione.

I piani di dimensionamento non includono risorse definite dal servizio che possano essere utilizzate come elemento `Resource` di una dichiarazione di policy IAM. Pertanto, non vi sono Amazon Resource Name (ARN) che è possibile utilizzare in una policy IAM. Per controllare l'accesso alle operazioni del piano di dimensionamento, utilizza sempre un * (asterisco) come risorsa durante la scrittura di una policy IAM.

Chiavi di condizione

L'elemento `Condition` (o blocco `Condition`) consente di specificare le condizioni in cui una dichiarazione è attiva. Ad esempio, potresti decidere che una policy venga applicata solo dopo una data specifica. Per esprimere le condizioni, utilizza chiavi di condizione predefinite.

I piani di dimensionamento non forniscono chiavi di condizione specifiche del servizio, ma supportano l'utilizzo di alcune chiavi di condizione globali. Per visualizzare tutte le chiavi di condizione AWS globali, consulta le chiavi di [contesto delle condizioni AWS globali nella Guida](#) per l'utente IAM.

L'elemento `Condition` è facoltativo.

Esempi

Per visualizzare esempi di policy basate su identità per i piani di dimensionamento, consulta [Identity-based esempi di politiche per i piani di scalabilità](#).

Resource-based politiche

Altri servizi Amazon Web Services, come ad esempio Amazon Simple Storage Service, supportano policy di autorizzazioni basate su risorse. Ad esempio, è possibile allegare una policy di autorizzazione a un bucket S3 per gestire le autorizzazioni di accesso a quel bucket.

I piani di dimensionamento non supportano le policy basate su risorse.

Liste di controllo degli accessi (ACL)

I piani di dimensionamento non supportano le liste di controllo accessi (ACL).

Autorizzazione basata su tag

Non è possibile applicare i tag ai piani di dimensionamento. Non dispongono nemmeno di risorse definite dal servizio che possono essere taggate. Pertanto, non supportano il controllo dell'accesso basato sui tag su una risorsa.

I piani di dimensionamento possono contenere risorse su cui è possibile applicare tag, come i gruppi Auto Scaling, che supportano il controllo dell'accesso in base ai tag. Per ulteriori informazioni, consulta la documentazione per quel Servizio AWS.

Ruoli IAM

Un [ruolo IAM](#) è un'entità all'interno dell' Account AWS che dispone di autorizzazioni specifiche.

Utilizzo di credenziali temporanee

Puoi utilizzare credenziali temporanee per effettuare l'accesso utilizzando la federazione, per assumere un ruolo IAM o per assumere un ruolo multi-account. È possibile ottenere credenziali di sicurezza temporanee chiamando operazioni AWS STS API come [AssumeRole](#) o [GetFederationToken](#).

I piani di dimensionamento supportano l'utilizzo di credenziali temporanee.

Service-linked ruoli per i piani di scalabilità

AWS Auto Scaling utilizza ruoli collegati ai servizi per le autorizzazioni necessarie per chiamare altri AWS servizi per conto dell'utente. Service-linked i ruoli semplificano la configurazione dei piani di scalabilità perché non è necessario aggiungere manualmente le autorizzazioni necessarie. Per ulteriori informazioni, consulta [Utilizzo di ruoli collegati a servizi](#) nella Guida per l'utente di IAM .

AWS Auto Scaling utilizza alcuni tipi di ruoli collegati al servizio per chiamare altre persone per conto dell'utente quando si utilizza un Servizi AWS piano di scalabilità:

- Ruolo collegato al servizio di scalabilità predittiva: consente AWS Auto Scaling di accedere ai dati metrici storici da CloudWatch. Consente inoltre la creazione di azioni pianificate per i gruppi Auto Scaling in base a una previsione di carico e alla previsione della capacità. Per ulteriori informazioni, consulta [Ruolo collegato al servizio di dimensionamento predittivo](#).
- Ruolo collegato al servizio Amazon EC2 Auto Scaling AWS Auto Scaling : consente di accedere e gestire le politiche di scalabilità di tracciamento degli obiettivi per i gruppi Auto Scaling. Per ulteriori informazioni, consulta [Service-linked i ruoli per Amazon EC2 Auto Scaling nella Amazon EC2 Auto Scaling User Guide](#).
- Ruolo collegato al servizio Application Auto Scaling: consente di accedere e gestire AWS Auto Scaling le politiche di scalabilità del tracciamento degli obiettivi per altre risorse scalabili. Esiste un ruolo collegato ai servizi per ciascun servizio. Per ulteriori informazioni, consulta [Service-linked i ruoli per Application Auto Scaling](#) nella Application Auto Scaling User Guide.

È possibile utilizzare la procedura seguente per stabilire se l'account dispone già di un ruolo collegato al servizio.

Per stabilire se esiste già un ruolo collegato al servizio

1. Aprire la console IAM all'indirizzo <https://console.aws.amazon.com/iam/>.
2. Nel riquadro di navigazione, seleziona Ruoli.
3. Cerca nell'elenco `AWSServiceRole` per trovare i ruoli collegati al servizio presenti nel tuo account. Cerca il nome del ruolo collegato al servizio da controllare.

Ruoli di servizio

AWS Auto Scaling non ha ruoli di servizio per i piani di scalabilità.

Ruolo collegato al servizio di dimensionamento predittivo

AWS Auto Scaling utilizza ruoli collegati ai servizi per le autorizzazioni necessarie per chiamare altri utenti per conto dell'utente quando si utilizza un AWS piano di scalabilità. Per ulteriori informazioni, consulta [Service-linked ruoli per i piani di scalabilità](#).

Le seguenti sezioni descrivono come creare e gestire i ruoli collegati al servizio per il dimensionamento predittivo. Per iniziare, configura le autorizzazioni per consentire a un'entità IAM (ad esempio un utente, un gruppo o un ruolo) di creare, modificare o eliminare un ruolo collegato ai servizi.

Autorizzazioni concesse dal ruolo collegato ai servizi

AWS Auto Scaling utilizza il ruolo collegato ai servizi denominato `AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` per chiamare altri AWS servizi per tuo conto quando abiliti la scalabilità predittiva.

`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` ritiene che il `autoscaling-plans.amazonaws.com` servizio assuma il ruolo.

Il ruolo collegato ai servizi utilizza la policy gestita `AWSAutoScalingPlansEC2AutoScalingPolicy`. Per vedere le autorizzazioni per questa policy, consulta [AWSAutoScalingPlansEC2AutoScalingPolicy](#) nella Guida di riferimento sulle policy gestite da AWS .

Creazione di un ruolo collegato ai servizi (automatica)

Non è necessario creare manualmente il `AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` ruolo. AWS crea questo ruolo per te quando crei un piano di scalabilità nel tuo account e abiliti la scalabilità predittiva.

AWS Per creare un ruolo collegato ai servizi per tuo conto, devi disporre delle autorizzazioni richieste. Per ulteriori informazioni, consulta le [autorizzazioni dei Service-linked ruoli nella Guida](#) per l'utente IAM.

Creazione di un ruolo collegato ai servizi (manuale)

Per creare manualmente un ruolo collegato ai servizi, è possibile utilizzare la console IAM, la CLI o l'API IAM. Per ulteriori informazioni, consulta [Create a service linked role](#) nella IAM User Guide.

Come creare un ruolo collegato ai servizi (AWS CLI)

Utilizza il seguente comando [create-service-linked-role](#) per creare il ruolo collegato al servizio.

```
aws iam create-service-linked-role --aws-service-name autoscaling-plans.amazonaws.com
```

Modifica del ruolo collegato ai servizi

Puoi modificare la descrizione dell'utilizzo di IAM.

`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` Per ulteriori informazioni, consulta [Edit a service-linked role description](#) nella Guida per l'utente IAM.

Eliminazione del ruolo collegato ai servizi

Se non hai più bisogno di utilizzare i piani di scalabilità, ti consigliamo di eliminarli `AWSServiceRoleForAutoScalingPlans_EC2AutoScaling`.

È possibile eliminare un ruolo collegato al servizio solo dopo aver eliminato tutti i piani di dimensionamento nel proprio Account AWS con il dimensionamento predittivo abilitato. Questo impedisce di rimuovere involontariamente l'autorizzazione ad accedere ai piani di dimensionamento.

Per eliminare i ruoli collegati ai servizi, puoi utilizzare la console IAM, la CLI o l'API IAM. Per ulteriori dettagli, consulta [Delete a service-linked role](#) nella Guida per l'utente IAM.

Dopo aver eliminato il ruolo `AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` collegato al servizio, AWS Auto Scaling crea nuovamente il ruolo se crei un piano di scalabilità con la scalabilità predittiva abilitata.

Regioni supportate

AWS Auto Scaling supporta l'utilizzo di ruoli collegati ai servizi in tutti i piani di scalabilità disponibili. Regioni AWS Per informazioni sulla disponibilità regionale dei piani di scalabilità, consulta [Endpoint e quote AWS Auto Scaling](#) nella Riferimenti generali di AWS.

Identity-based esempi di politiche per i piani di scalabilità

Per impostazione predefinita, un nuovo utente IAM non ha le autorizzazioni per svolgere alcuna operazione. Un amministratore IAM deve creare e assegnare policy IAM che diano un'autorizzazione di identità IAM (ad esempio un utente o un ruolo) per lavorare con i piani di dimensionamento.

Per informazioni su come creare una policy IAM utilizzando questi documenti di policy JSON di esempio, consulta [Creazione di policy nella scheda JSON](#) nella Guida per l'utente di IAM.

Argomenti

- [Best practice per le policy](#)
- [Consentire agli utenti di creare piani di dimensionamento](#)
- [Consentire agli utenti di abilitare il dimensionamento predittivo](#)
- [Autorizzazioni aggiuntive richieste](#)
- [Autorizzazioni necessarie per creare un ruolo collegato al servizio](#)

Best practice per le policy

Identity-based le politiche determinano se qualcuno può creare, accedere o eliminare AWS Auto Scaling risorse nel tuo account. Queste operazioni possono comportare costi aggiuntivi per l'Account AWS. Quando si creano o modificano policy basate sull'identità, seguire queste linee guida e raccomandazioni:

- Inizia con le policy AWS gestite e passa alle autorizzazioni con privilegi minimi: per iniziare a concedere autorizzazioni a utenti e carichi di lavoro, utilizza le politiche AWS gestite che concedono le autorizzazioni per molti casi d'uso comuni. Sono disponibili nel tuo Account AWS. Ti consigliamo di ridurre ulteriormente le autorizzazioni definendo politiche gestite dai clienti AWS specifiche per i tuoi casi d'uso. Per maggiori informazioni, consulta [Policy gestite da AWS](#) o [Policy gestite da AWS per le funzioni dei processi](#) nella Guida per l'utente di IAM.
- Applicazione delle autorizzazioni con privilegio minimo - Quando si impostano le autorizzazioni con le policy IAM, concedere solo le autorizzazioni richieste per eseguire un'attività. È possibile farlo definendo le azioni che possono essere intraprese su risorse specifiche in condizioni specifiche, note anche come autorizzazioni con privilegio minimo. Per maggiori informazioni sull'utilizzo di IAM per applicare le autorizzazioni, consulta [Policy e autorizzazioni in IAM](#) nella Guida per l'utente di IAM.
- Condizioni d'uso nelle policy IAM per limitare ulteriormente l'accesso - Per limitare l'accesso ad azioni e risorse è possibile aggiungere una condizione alle policy. Ad esempio, è possibile scrivere una condizione di policy per specificare che tutte le richieste devono essere inviate utilizzando SSL. Puoi anche utilizzare le condizioni per concedere l'accesso alle azioni del servizio se vengono utilizzate tramite uno specifico Servizio AWS, ad esempio CloudFormation. Per maggiori informazioni, consultare la sezione [Elementi delle policy JSON di IAM: condizione](#) nella Guida per l'utente di IAM.
- Utilizzo dello strumento di analisi degli accessi IAM per convalidare le policy IAM e garantire autorizzazioni sicure e funzionali - Lo strumento di analisi degli accessi IAM convalida le policy nuove ed esistenti in modo che aderiscano al linguaggio (JSON) della policy IAM e alle best

pratiche di IAM. Lo strumento di analisi degli accessi IAM offre oltre 100 controlli delle policy e consigli utili per creare policy sicure e funzionali. Per maggiori informazioni, consultare [Convalida delle policy per il Sistema di analisi degli accessi IAM](#) nella Guida per l'utente di IAM.

- Richiedi l'autenticazione a più fattori (MFA): se hai uno scenario che richiede utenti IAM o un utente root nel Account AWS tuo, attiva l'MFA per una maggiore sicurezza. Per richiedere la MFA quando vengono chiamate le operazioni API, aggiungere le condizioni MFA alle policy. Per maggiori informazioni, consultare [Protezione dell'accesso API con MFA](#) nella Guida per l'utente di IAM.

Per maggiori informazioni sulle best practice in IAM, consulta [Best practice di sicurezza in IAM](#) nella Guida per l'utente di IAM.

Consentire agli utenti di creare piani di dimensionamento

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per creare piani di dimensionamento.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling-plans:*",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch>DeleteAlarms",
        "cloudwatch:DescribeAlarms",
        "cloudformation:ListStackResources"
      ],
      "Resource": "*"
    }
  ]
}
```

Per utilizzare un piano di dimensionamento, l'utente finale deve disporre di autorizzazioni aggiuntive che gli consentano di lavorare con risorse specifiche nel proprio account. Queste autorizzazioni sono elencate in [Autorizzazioni aggiuntive richieste](#).

Ogni utente della console necessita inoltre di autorizzazioni che gli consentano di scoprire le risorse scalabili nel proprio account e di visualizzare grafici dei CloudWatch dati metrici dalla console. AWS Auto Scaling Il set aggiuntivo di autorizzazioni necessarie per lavorare con la console è elencato di seguito: AWS Auto Scaling

- `cloudformation:ListStacks`: per elencare stack.
- `tag:GetTagKeys`: per trovare risorse scalabili che contengono determinate chiavi tag.
- `tag:GetTagValues`: per trovare le risorse che contengono determinati valori di tag.
- `autoscaling:DescribeTags`: per trovare gruppi Auto Scaling che contengono determinati tag.
- `cloudwatch:GetMetricData`: per visualizzare i dati nei grafici dei parametri.

Consentire agli utenti di abilitare il dimensionamento predittivo

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per abilitare il dimensionamento predittivo. Queste autorizzazioni estendono le funzionalità dei piani di dimensionamento impostati per il dimensionamento dei gruppi Auto Scaling.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "cloudwatch:GetMetricData",
        "autoscaling:DescribeAutoScalingGroups",
        "autoscaling:DescribeScheduledActions",
        "autoscaling:BatchPutScheduledUpdateGroupAction",
        "autoscaling:BatchDeleteScheduledAction"
      ],
      "Resource": "*"
    }
  ]
}
```

Autorizzazioni aggiuntive richieste

Per configurare correttamente i piani di dimensionamento, agli utenti finali devono essere concesse le autorizzazioni per ogni servizio di destinazione per il quale configureranno il dimensionamento. Per concedere le autorizzazioni minime necessarie per lavorare con i servizi di destinazione, leggi le informazioni in questa sezione e specifica le azioni pertinenti nell'elemento `Action` di una dichiarazione di policy IAM.

Gruppi Auto Scaling

Per aggiungere gruppi Auto Scaling a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon EC2 Auto Scaling:

- `autoscaling:UpdateAutoScalingGroup`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:PutScalingPolicy`
- `autoscaling:DescribePolicies`
- `autoscaling>DeletePolicy`

Servizi ECS

Per aggiungere servizi ECS a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon ECS e Application Auto Scaling:

- `ecs:DescribeServices`
- `ecs:UpdateService`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

parco istanze spot

Per aggiungere serie di istanze Spot a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon EC2 e Application Auto Scaling:

- `ec2:DescribeSpotFleetRequests`
- `ec2:ModifySpotFleetRequest`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Tabelle o indici globali DynamoDB

Per aggiungere tabelle o indici globali DynamoDB a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da DynamoDB e Application Auto Scaling:

- `dynamodb:DescribeTable`
- `dynamodb:UpdateTable`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Cluster di database Aurora

Per aggiungere cluster Aurora DB a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon Aurora e Application Auto Scaling:

- `rds:AddTagsToResource`

- `rds:CreateDBInstance`
- `rds>DeleteDBInstance`
- `rds:DescribeDBClusters`
- `rds:DescribeDBInstances`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Autorizzazioni necessarie per creare un ruolo collegato al servizio

AWS Auto Scaling richiede le autorizzazioni per creare un ruolo collegato al servizio la prima volta che un utente Account AWS crea un piano di scalabilità con la scalabilità predittiva abilitata. Se il ruolo collegato al servizio non esiste già, lo crea nel tuo account. AWS Auto Scaling Il ruolo collegato al servizio concede le autorizzazioni in AWS Auto Scaling modo che possa chiamare altri servizi per tuo conto.

Affinché la creazione automatica di un ruolo riesca, gli utenti devono disporre delle autorizzazioni per l'operazione `iam:CreateServiceLinkedRole`.

```
"Action": "iam:CreateServiceLinkedRole"
```

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per creare un ruolo collegato al servizio.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:CreateServiceLinkedRole",
```

```
    "Resource": "arn:aws:iam::*:role/aws-service-role/autoscaling-
plans.amazonaws.com/AWSServiceRoleForAutoScalingPlans_EC2AutoScaling",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": "autoscaling-plans.amazonaws.com"
        }
    }
}
]
```

Per ulteriori informazioni, consulta [Ruolo collegato al servizio di dimensionamento predittivo](#).

Convalida della conformità per i piani di scalabilità

Per sapere se un Servizio AWS programma rientra nell'ambito di specifici programmi di conformità, consulta Servizi AWS la sezione [Scope by Compliance Program Servizi AWS](#) e scegli il programma di conformità che ti interessa. Per informazioni generali, consulta Programmi di [AWS conformità Programmi](#) di di .

È possibile scaricare report di audit di terze parti utilizzando AWS Artifact. Per ulteriori informazioni, consulta [Scaricamento dei report in AWS Artifact](#) .

La vostra responsabilità di conformità durante l'utilizzo Servizi AWS è determinata dalla sensibilità dei dati, dagli obiettivi di conformità dell'azienda e dalle leggi e dai regolamenti applicabili. Per ulteriori informazioni sulla responsabilità di conformità durante l'utilizzo Servizi AWS, consulta [AWS la documentazione sulla sicurezza](#).

Sicurezza dell'infrastruttura per piani di scalabilità

In quanto servizio gestito, AWS Auto Scaling è protetto dalla sicurezza di rete AWS globale. Per informazioni sui servizi AWS di sicurezza e su come AWS protegge l'infrastruttura, consulta [AWS Cloud Security](#). Per progettare il tuo AWS ambiente utilizzando le migliori pratiche per la sicurezza dell'infrastruttura, vedi [Infrastructure Protection](#) in Security Pillar AWS Well-Architected Framework.

Utilizzate chiamate API AWS pubblicate per accedere AWS Auto Scaling attraverso la rete. I client devono supportare quanto segue:

- Transport Layer Security (TLS). È richiesto TLS 1.2 ed è consigliato TLS 1.3.

- Suite di crittografia con Perfect Forward Secrecy (PFS) come DHE (Ephemeral) o ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). Diffie-Hellman La maggior parte dei sistemi moderni, come Java 7 e versioni successive, supporta tali modalità.

Quote per i piani di dimensionamento

Your Account AWS dispone delle quote predefinite (precedentemente denominate limiti) relative ai piani di scalabilità. Salvo diversa indicazione, ogni quota si applica a una regione specifica. Se per alcune quote è possibile richiedere aumenti, altre quote non possono essere modificate.

Per visualizzare le quote per Application Auto Scaling, apri la [Console di Service Quotas](#). Nel riquadro di navigazione, scegli Servizi AWS e seleziona Piani di AWS Auto Scaling.

Per richiedere un aumento delle quote, consultare [Richiesta di aumento delle quote](#) nella Guida dell'utente di Service Quotas.

Hai Account AWS le seguenti quote relative ai piani di scalabilità.

Nome	Predefinita	Adattabile
Risorse scalabili per tipo di risorsa	Amazon DynamoDB: 3.000 Gruppi Amazon Auto EC2 Scaling: 200 Tutti gli altri tipi di risorse: 500	Sì
Piani di dimensionamento	100	Sì
Istruzioni di dimensionamento per piano di dimensionamento	500	No
Configurazioni di monitoraggio obiettivi per istruzione di dimensionamento	10	No

Tieni a mente le quote di servizio mentre dimensiona i carichi di lavoro. Ad esempio, quando raggiungi il numero massimo di unità di capacità consentite da un servizio, la scalabilità orizzontale si interromperà. Se la domanda diminuisce e la capacità attuale diminuisce, è AWS Auto Scaling possibile eseguire nuovamente la scalabilità orizzontale. Per evitare di raggiungere nuovamente questo limite di quota di servizio, puoi richiedere un aumento. Ogni servizio ha le proprie quote predefinite per la capacità massima della risorsa. Per informazioni sulle quote predefinite per altri servizi Amazon Web Services, consulta [Endpoint e quote di servizio](#) nella Riferimenti generali di Amazon Web Services.

Cronologia dei documenti per i piani di ridimensionamento

La tabella seguente descrive importanti aggiunte alla AWS Auto Scaling documentazione. Per ricevere notifiche sugli aggiornamenti della documentazione, puoi sottoscrivere il feed RSS.

Modifica	Descrizione	Data
Nuovi contenuti per la migrazione da opzioni AWS Auto Scaling alternative	Ora puoi AWS Auto Scaling migrare dalla scalabilità predittiva di Amazon EC2 Auto Scaling, che offre più funzionalità. Per ulteriori informazioni, consulta <u>Migrare il piano di scalabilità</u> .	5 aprile 2024
Nuovi contenuti di sicurezza	Abbiamo pubblicato un capitolo sulla sicurezza aggiornato. Come parte di questo aggiornamento, abbiamo sostituito «Autenticazione e controllo degli accessi» con la gestione delle identità e degli accessi per AWS Auto Scaling .	12 marzo 2020
Supporto per gli endpoint Amazon VPC	Ora puoi stabilire una connessione privata tra il tuo VPC e AWS Auto Scaling. Per informazioni e istruzioni sulla migrazione, consulta l'argomento Piani di dimensionamento ed endpoint VPC di interfaccia .	22 novembre 2019
Supporto per aumentare la capacità massima al di sopra della capacità prevista	Aggiunge il supporto per console per consentire al piano di dimensionamento	9 marzo 2019

di aumentare la capacità massima al di sopra di quella di previsione di un valore di buffer specificato. Per ulteriori informazioni, consulta Impostazioni di [ridimensionamento predittivo](#).

[Scalabilità predittiva e miglioramenti](#)

Ora puoi utilizzare il dimensionamento predittivo per dimensionare in modo proattivo i gruppi Amazon EC2 Auto Scaling. Questa release, inoltre, aggiunge il supporto per la sostituzione delle policy di dimensionamento create al di fuori del piano di dimensionamento (ad esempio da altre console) e il controllo se si abilita la funzione di dimensionamento dinamico del piano.

20 novembre 2018

[Supporto delle impostazioni per le risorse personalizzate](#)

Aggiunto il supporto per la personalizzazione di diverse impostazioni per ogni singola risorsa o più risorse nello stesso momento.

9 ottobre 2018

[I tag come origine dell'applicazione](#)

Questa versione aggiunge supporto per specificare un set di tag come origine di un'applicazione.

23 aprile 2018

[Nuovo servizio](#)

Versione iniziale di AWS Auto Scaling

16 gennaio 2018

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.