



Actividades de mantenimiento de las bases de datos de PostgreSQL en Amazon RDS y Amazon Aurora para evitar problemas de rendimiento

AWS Orientación prescriptiva



AWS Orientación prescriptiva: Actividades de mantenimiento de las bases de datos de PostgreSQL en Amazon RDS y Amazon Aurora para evitar problemas de rendimiento

Table of Contents

Introducción	1
Resultados empresariales específicos	1
Control de simultaneidad multiversión (MVCC)	1
Succión y análisis automáticos de las tablas	3
Parámetros de autovacuum relacionados con la memoria	5
Ajuste de los parámetros de autovacuum	6
A nivel de instancia o clúster	6
Nivel de tabla	6
Uso de una configuración de autovacuum agresiva a nivel de tabla	6
Ventajas y limitaciones	8
Succión y análisis manuales de las tablas	9
Ejecución de operaciones de succión y limpieza en paralelo	10
Reescritura de una tabla completa con VACUUM FULL	14
Eliminación de la sobrecarga con pg_repack	16
Nueva creación de índices	18
Creación de un nuevo índice	22
Nueva creación de un índice	22
Ejemplos	23
Ejemplo: Recuperación de espacio mediante autovacuum y VACUUM FULL	25
Recursos	30
Historial de documentos	31
Glosario	32
#	32
A	33
B	36
C	38
D	42
E	46
F	49
G	51
H	52
I	54
L	56
M	57

O	62
P	65
Q	68
R	68
S	72
T	76
U	77
V	78
W	78
Z	79
.....	lxxi

Actividades de mantenimiento de las bases de datos de PostgreSQL en Amazon RDS y Amazon Aurora para evitar problemas de rendimiento

Anuradha Chintha, Rajesh Madiwale y Srinivas Potlachervoo, Amazon Web Services (AWS)

Agosto de 2025 ([historia del documento](#))

La edición de Amazon Aurora compatible con PostgreSQL y Amazon Relational Database Service (Amazon RDS) para PostgreSQL son servicios de bases de datos relacionales completamente administrados para bases de datos de PostgreSQL. Estos servicios administrados liberan al administrador de la base de datos de muchas tareas de mantenimiento y administración. Sin embargo, algunas tareas de mantenimiento, como VACUUM, requieren una estrecha supervisión y una configuración que se ajuste al uso de la base de datos. En esta guía se describen las actividades de mantenimiento de PostgreSQL en Amazon RDS y Aurora.

Resultados empresariales específicos

El rendimiento de las bases de datos es una medida clave que sustenta el éxito de una empresa. Las actividades de mantenimiento que realiza en las bases de datos de Aurora compatible con PostgreSQL y Amazon RDS para PostgreSQL ofrecen las siguientes ventajas:

- Ayuda a lograr un rendimiento óptimo de las consultas
- Libera el espacio sobrecargado para que lo reutilicen en futuras transacciones
- Evita el reinicio de las transacciones
- Ayuda al optimizador a generar buenos planes
- Garantiza un uso adecuado del índice

Control de simultaneidad multiversión (MVCC)

El mantenimiento de la base de datos de PostgreSQL requiere comprender el control de simultaneidad multiversión (MVCC), que es un mecanismo de PostgreSQL. Cuando se procesan varias transacciones simultáneamente en la base de datos, el MVCC garantiza la atomicidad

y el aislamiento, que son dos características de las transacciones con atomicidad, coherencia, aislamiento y durabilidad (ACID). En el MVCC, cada operación de escritura genera una nueva versión de los datos y almacena la versión anterior. Los lectores y los escritores no se bloquean entre sí. Cuando una transacción lee datos, el sistema elige una de las versiones para aislar la transacción. PostgreSQL y algunas bases de datos relacionales utilizan una adaptación del MVCC llamada aislamiento de instantáneas (SI). Por ejemplo, Oracle implementa SI mediante segmentos de reversión. Durante una operación de escritura, Oracle escribe la versión antigua de los datos en un segmento de reversión y sobrescribe el área de datos con la versión nueva. Las bases de datos de PostgreSQL implementan SI mediante reglas de comprobación de la visibilidad para evaluar las versiones. Cuando se colocan datos nuevos en una página de tabla, PostgreSQL utiliza estas reglas para seleccionar la versión de los datos adecuada para una operación de lectura.

Al modificar los datos de una fila de la tabla, PostgreSQL utiliza el MVCC para mantener varias versiones de la fila. Durante las opciones UPDATE y DELETE que realiza en la tabla, la base de datos conserva las versiones antiguas de las filas para otras transacciones en ejecución que puedan necesitar una vista de los datos coherente. Estas versiones antiguas se llaman filas inactivas (tuplas). Las colecciones de tuplas inactivas producen sobrecarga. Una gran cantidad de sobrecarga en la base de datos puede provocar una serie de problemas, como una creación deficiente de planes de consultas, un rendimiento lento de las consultas y un mayor uso del espacio en disco para almacenar las versiones antiguas.

Para eliminar la sobrecarga y mantener una base de datos en buen estado, se tiene que realizar un mantenimiento periódico con estas actividades, que se analizan en las siguientes secciones:

- [Succión y análisis automáticos de las tablas](#)
- [Succión y análisis manuales de las tablas](#)
- [Eliminación de la sobrecarga con pg_repack](#)
- [Nueva creación de índices](#)

Succión y análisis automáticos de las tablas

Autovacuum es un daemon (es decir, se ejecuta en segundo plano) que succiona (limpia) automáticamente las tuplas inactivas, recupera el almacenamiento y recopila estadísticas. Comprueba si hay tablas hinchadas en la base de datos y borra esa sobrecarga para poder reutilizar el espacio. Supervisa las tablas y los índices de las bases de datos y los agrega a una tarea de succión cuando alcanzan un límite específico de operaciones de actualización o eliminación.

Autovacuum administra la succión mediante la automatización de los comandos VACUUM y ANALYZE de PostgreSQL. VACUUM elimina la sobrecarga de las tablas y recupera espacio, a la vez que ANALYZE actualiza las estadísticas que permiten al optimizador elaborar planes eficientes. VACUUM también lleva a cabo una importante tarea denominada congelación por succión para evitar problemas relacionados con los ID de las transacciones en la base de datos. Cada fila que se actualiza en la base de datos recibe un ID de transacción de parte del mecanismo de control de transacciones de PostgreSQL. Estos IDs controlan la visibilidad de la fila para otras transacciones simultáneas. El ID de la transacción es un número de 32 bits. Dos mil millones IDs siempre se guardan en el pasado visible. Los restantes (unos 2.200 millones) IDs se conservan para transacciones que se realizarán en el futuro y se ocultan de la transacción actual. PostgreSQL necesita que se realicen limpiezas y congelaciones de las filas antiguas de forma ocasional para evitar que las transacciones se acumulen y hagan invisibles las filas antiguas existentes cuando se crean nuevas transacciones. Para obtener más información, consulte [Preventing Transaction ID Wraparound Failures](#) en la documentación de PostgreSQL.

Autovacuum es una funcionalidad recomendada y está activada de manera predeterminada. Acepta los siguientes parámetros:

Parámetro	Descripción	Predeterminado para Amazon RDS	Predeterminado para Aurora
autovacuum_vacuum_threshold	El número mínimo de operaciones de actualización o eliminación de tuplas que deben realizarse en una tabla antes	50 operaciones	50 operaciones

	de que autovacuum la succione.		
autovacuum_analyze_threshold	El número mínimo de inserciones, actualizaciones o eliminaciones de tuplas que deben producirse en una tabla antes de que autovacuum la analice.	50 operaciones	50 operaciones
autovacuum_vacuum_scale_factor	El porcentaje de tuplas que deben modificarse en una tabla antes de que autovacuum la succione.	0.1	0.1
autovacuum_analyze_scale_factor	El porcentaje de tuplas que deben modificarse en una tabla antes de que autovacuum la analice.	0.05	0,05
autovacuum_freeze_max_age	La edad máxima que debe transcurrir IDs antes de que se vacíe una mesa para evitar problemas relacionados con el identificador de la transacción.	200 000 000 de transacciones	200 000 000 de transacciones

Autovacuum crea una lista de tablas para procesarlas en función de fórmulas de umbral específicas, de la siguiente manera.

- Umbral para ejecutar VACUUM en una tabla:

```
vacuum threshold = autovacuum_vacuum_threshold + (autovacuum_vacuum_scale_factor *  
Total row count of table)
```

- Umbral para ejecutar ANALYZE en una tabla:

```
analyze threshold = autovacuum_analyze_threshold + (autovacuum_analyze_scale_factor *  
Total row count of table)
```

En el caso de las tablas pequeñas o medianas, los valores predeterminados pueden ser suficientes. Sin embargo, una tabla grande que tenga modificaciones de datos frecuentes tendrá un mayor número de tuplas inactivas. En este caso, es posible que autovacuum procese la tabla con frecuencia para realizar tareas de mantenimiento y que las tareas de mantenimiento de otras tablas se retrasen o se ignoren hasta que se acabe la tabla grande. Para evitar esta situación, puede ajustar los parámetros de autovacuum que se describen en la siguiente sección.

Parámetros de autovacuum relacionados con la memoria

autovacuum_max_workers

Especifica el número máximo de procesos de autovacuum (aparte del iniciador de autovacuum) que se pueden ejecutar al mismo tiempo. Este parámetro solo se puede configurar al iniciar el servidor. Si el proceso de autovacuum está ocupado con una tabla grande, este parámetro ayuda a realizar la succión en otras tablas.

maintenance_work_mem

Especifica la cantidad máxima de memoria que se utilizará para las operaciones de mantenimiento, como VACUUM, CREATE INDEX y ALTER. En Amazon RDS y Aurora, la memoria se asigna en función de la clase de instancia mediante la fórmula $\text{GREATEST}(\{\text{DBInstanceClassMemory}/63963136*1024\}, 65536)$. Cuando se ejecuta autovacuum, se puede asignar hasta `autovacuum_max_workers` veces el valor calculado, así que procure no establecer un valor demasiado alto. Para controlarlo, puede configurar `autovacuum_work_mem` por separado.

autovacuum_work_mem

Establece la memoria máxima que puede utilizar cada proceso de trabajo de autovacuum. El valor predeterminado de este parámetro es -1, lo que indica que debe utilizar el valor de `maintenance_work_mem` en su lugar.

Para obtener más información sobre los parámetros de memoria de autovacuum, consulte [Asignación de memoria para autovacuum](#) en la documentación de Amazon RDS.

Ajuste de los parámetros de autovacuum

Es posible que los usuarios tengan que ajustar los parámetros de autovacuum en función de sus operaciones de actualización y eliminación. Los parámetros siguientes se pueden configurar por tabla, instancia o clúster.

A nivel de instancia o clúster

Como ejemplo, veamos una base de datos del sector bancario en la que se esperan operaciones continuas con el lenguaje de manipulación de datos (DML). Para mantener el estado de la base de datos, debe ajustar los parámetros de autovacuum a nivel de clúster para Aurora y a nivel de instancia para Amazon RDS, y aplicar también el mismo grupo de parámetros al lector. En caso de que se produzca una conmutación por error, debe aplicar los mismos parámetros al nuevo escritor.

Nivel de tabla

Por ejemplo, en una base de datos relacionada con la entrega de alimentos en la que se esperan operaciones con DML continuas en una sola tabla llamada `orders`, debería valorar si ajustar el parámetro `autovacuum_analyze_threshold` a nivel de tabla mediante el siguiente comando:

```
ALTER TABLE <table_name> SET (autovacuum_analyze_threshold = <threshold rows>)
```

Uso de una configuración de autovacuum agresiva a nivel de tabla

El ejemplo de la tabla `orders` que tiene operaciones de actualización y eliminación continuas se convierte en una buena opción para realizar la succión debido a la configuración de autovacuum predeterminada. Esto provoca que se hagan planes incorrectos y que las consultas sean lentas. Eliminar la sobrecarga y actualizar las estadísticas requiere una configuración de autovacuum agresiva a nivel de tabla.

Para determinar la configuración, lleve un registro de la duración de las consultas que se ejecutan en esta tabla e identifique el porcentaje de operaciones con DML que hacen cambiar los planes. La vista `pg_stat_user_tables` lo ayuda a realizar un seguimiento de las operaciones de inserción, actualización y eliminación.

Ejemplo:

Supongamos que el optimizador hace planes incorrectos cada vez que cambia el 5 % de la tabla `orders`. En este caso, debe cambiar el umbral del factor de escala al 2 %, como se indica a continuación:

```
ALTER TABLE orders SET (autovacuum_vacuum_scale_factor = 0.02)
```

 Tip

Seleccione cuidadosamente la configuración de autovacuum agresiva para evitar un consumo elevado de los recursos.

Para obtener más información, consulte los siguientes temas:

- [Descripción del autovacuum en entornos Amazon RDS para AWS PostgreSQL](#) (entrada del blog)
- [Automatic Vacuuming](#) (documentación de PostgreSQL)
- [Ajuste de los parámetros de PostgreSQL en Amazon RDS y Amazon AWS Aurora](#) (guía prescriptiva)

Para asegurarse de que autovacuum funcione de forma eficaz, supervise las filas inactivas, el uso del disco y la última vez que autovacuum o ANALYZE se ejecutó de forma periódica. La vista `pg_stat_all_tables` proporciona información sobre cada tabla (`relname`) y sobre cuántas tuplas inactivas (`n_dead_tup`) hay en la tabla.

Supervisar la cantidad de tuplas inactivas que hay en cada tabla, especialmente en las que se actualizan con frecuencia, permite determinar si los procesos de autovacuum eliminan periódicamente las tuplas inactivas para reutilizar el espacio en disco y mejorar el rendimiento. Puede utilizar la siguiente consulta para comprobar la cantidad de tuplas muertas y la hora de la última ejecución de autovacuum en las tablas:

```
SELECT
relname AS TableName,n_live_tup AS LiveTuples,n_dead_tup AS DeadTuples,
last_autovacuum AS Autovacuum,last_autoanalyze AS Autoanalyze_FROM
pg_stat_user_tables;
```

Ventajas y limitaciones

Autovacuum ofrece las siguientes ventajas:

- Elimina automáticamente la sobrecarga de las tablas.
- Evita el reinicio de los ID de transacción.
- Mantiene actualizadas las estadísticas de la base de datos.

Limitaciones:

- Si las consultas utilizan el procesamiento paralelo, es posible que la cantidad de procesos de trabajo no sea suficiente para el autovacuum.
- Si autovacuum se ejecuta durante las horas pico, es posible que aumente la utilización de los recursos. Debe ajustar los parámetros para solucionar este problema.
- Si las páginas de la tabla están ocupadas en otra sesión, es posible que autovacuum las omita.
- Autovacuum no puede acceder a las tablas temporales.

Succión y análisis manuales de las tablas

Si la base de datos se succiona mediante el proceso de autovacuum, se recomienda evitar realizar succiones manuales en toda la base de datos con demasiada frecuencia. Una aspiradora manual podría provocar I/O cargas innecesarias o picos de CPU, y también podría no eliminar las tuplas muertas. Utilice las aspiradoras manuales table-by-table únicamente si es realmente necesario, por ejemplo, cuando la proporción de tuplas activas y muertas sea baja o cuando haya espacios largos entre las aspiradoras automáticas. Además, debe realizar succiones manuales cuando la actividad del usuario sea mínima.

Autovacuum también mantiene actualizadas las estadísticas de una tabla. Al ejecutar el comando ANALYZE manualmente, vuelve a crear estas estadísticas en lugar de actualizarlas. La nueva creación de las estadísticas mediante el proceso habitual de autovacuum cuando ya están actualizadas puede provocar la utilización de los recursos del sistema.

Se recomienda que ejecute los comandos [VACUUM](#) y [ANALYZE](#) manualmente en las siguientes situaciones:

- Durante las horas de menor actividad y en las tablas más utilizadas, es posible que autovacuum no sea suficiente.
- Inmediatamente después de cargar datos de forma masiva en la tabla de destino. En este caso, si ejecuta ANALYZE manualmente, las estadísticas se vuelven a crear completamente, lo cual es una mejor opción que esperar a que autovacuum comience.
- Para succionar tablas temporales (autovacuum no puede acceder a ellas).

Para reducir el I/O impacto de la ejecución de los ANALYZE comandos VACUUM y en la actividad simultánea de la base de datos, puede utilizar el parámetro `vacuum_cost_delay`. En muchas situaciones, los comandos de mantenimiento como VACUUM y ANALYZE no tienen por qué finalizarse rápidamente. Sin embargo, estos comandos no deberían interferir con la capacidad del sistema para realizar otras operaciones en la base de datos. Para evitarlo, puede activar los retardos de succión basados en los costos mediante el parámetro `vacuum_cost_delay`. Este parámetro está desactivado de forma predeterminada para los comandos VACUUM que se ejecutan manualmente. Para activarlo, configúrelo en un valor distinto de cero.

Ejecución de operaciones de succión y limpieza en paralelo

La opción [PARALLEL](#) del comando VACUUM utiliza trabajos paralelos para las fases de succión y limpieza de índices y está desactivada de forma predeterminada. El número de trabajos paralelos (el grado de paralelismo) viene determinado por el número de índices de la tabla, y el usuario puede especificar este número. Si ejecuta operaciones VACUUM paralelas sin un argumento entero, el grado de paralelismo se calcula en función del número de índices de la tabla.

Los siguientes parámetros lo ayudan a configurar la succión paralela en Amazon RDS para PostgreSQL y Aurora compatible con PostgreSQL:

- [max_worker_processes](#) establece el número máximo de procesos de trabajo que se ejecutan simultáneamente.
- [min_parallel_index_scan_size](#) establece la cantidad mínima de datos de índice que se deben escanear para poder considerar la posibilidad de realizar un escaneo paralelo.
- [max_parallel_maintenance_workers](#) establece el número máximo de trabajos paralelos que se pueden iniciar con un único comando de utilidad.

Note

La opción PARALLEL se utiliza únicamente para la succión. No afecta al comando ANALYZE.

En el siguiente ejemplo se ilustra el comportamiento de la base de datos cuando se usa VACUUM y ANALYZE manualmente en una base de datos.

Esta es una tabla de ejemplo en la que se ha desactivado autovacuum (solo con fines ilustrativos; no se recomienda desactivarlo):

```
create table t1 ( a int, b int, c int );
alter table t1 set (autovacuum_enabled=false);
```

```
apgl=> \d+ t1
Table "public.t1"
Column | Type | Collation | Nullable | Default | Storage | Stats target | Description
-----+-----+-----+-----+-----+-----+-----+-----
+-----
```

```
a | integer | | | plain | |
b | integer | | | plain | |
c | integer | | | plain | |
Access method: heap
Options: autovacuum_enabled=false
```

Agregue 1 millón de filas a la tabla t1:

```
apgl=> select count(*) from t1;
count
1000000
(1 row)
```

Estadísticas de la tabla t1:

```
select * from pg_stat_all_tables where relname='t1';
-[ RECORD 1 ]-----+-----
relid          | 914744
schemaname     | public
relname        | t1
seq_scan       | 0
seq_tup_read   | 0
idx_scan       |
idx_tup_fetch  |
n_tup_ins      | 1000000
n_tup_upd      | 0
n_tup_del      | 0
n_tup_hot_upd  | 0
n_live_tup     | 1000000
n_dead_tup     | 0
n_mod_since_analyze | 1000000
last_vacuum    |
last_autovacuum |
last_analyze   |
last_autoanalyze |
vacuum_count   | 0
autovacuum_count | 0
analyze_count  | 0
autoanalyze_count | 0
```

Agregue un índice:

```
create index i2 on t1 (b,a);
```

Ejecute el comando EXPLAIN (plan 1):

```
Bitmap Heap Scan on t1 (cost=10521.17..14072.67 rows=5000 width=4)
Recheck Cond: (a = 5)
# Bitmap Index Scan on i2 (cost=0.00..10519.92 rows=5000 width=0)
Index Cond: (a = 5)
(4 rows)
```

Ejecute el comando EXPLAIN ANALYZE (plan 2):

```
explain (analyze, buffers, costs off) select a from t1 where b = 5;
QUERY PLAN
Bitmap Heap Scan on t1 (actual time=0.023..0.024 rows=1 loops=1)
Recheck Cond: (b = 5)
Heap Blocks: exact=1
Buffers: shared hit=4
# Bitmap Index Scan on i2 (actual time=0.016..0.016 rows=1 loops=1)
Index Cond: (b = 5)
Buffers: shared hit=3
Planning Time: 0.054 ms
Execution Time: 0.076 ms
(9 rows)
```

Los comandos EXPLAIN y EXPLAIN ANALYZE muestran planes diferentes, ya que autovacuum estaba desactivado en la tabla y el comando ANALYZE no se ejecutó manualmente. Ahora vamos a actualizar un valor de la tabla y volveremos a crear el plan EXPLAIN ANALYZE:

```
update t1 set a=8 where b=5;
explain (analyze, buffers, costs off) select a from t1 where b = 5;
```

El comando EXPLAIN ANALYZE (plan 3) ahora muestra:

```
apgl=> explain (analyze, buffers, costs off) select a from t1 where b = 5;
QUERY PLAN
Bitmap Heap Scan on t1 (actual time=0.075..0.076 rows=1 loops=1)
Recheck Cond: (b = 5)
Heap Blocks: exact=1
```

```

Buffers: shared hit=5
# Bitmap Index Scan on i2 (actual time=0.017..0.017 rows=2 loops=1)
Index Cond: (b = 5)
Buffers: shared hit=3
Planning Time: 0.053 ms
Execution Time: 0.125 ms

```

Si compara los costos entre el plan 2 y el plan 3, verá las diferencias en el tiempo de planificación y ejecución, ya que aún no hemos recopilado estadísticas.

Ahora ejecutemos ANALYZE manualmente en la tabla y, a continuación, comprobemos las estadísticas y volvamos a crear el plan:

```

apgl=> analyze t1
apgl# ;
ANALYZE
Time: 212.223 ms

apgl=> select * from pg_stat_all_tables where relname='t1';
-[ RECORD 1 ]-----+-----
relid          | 914744
schemaname     | public
relname        | t1
seq_scan       | 3
seq_tup_read   | 1000000
idx_scan       | 3
idx_tup_fetch  | 3
n_tup_ins      | 1000000
n_tup_upd      | 1
n_tup_del      | 0
n_tup_hot_upd  | 0
n_live_tup     | 1000000
n_dead_tup    | 1
n_mod_since_analyze | 0
last_vacuum    |
last_autovacuum |
last_analyze   | 2023-04-15 11:39:02.075089+00
last_autoanalyze |
vacuum_count   | 0
autovacuum_count | 0
analyze_count  | 1
autoanalyze_count | 0

```

```
Time: 148.347 ms
```

Ejecute el comando `EXPLAIN ANALYZE (plan 4)`:

```
apgl=> explain (analyze,buffers,costs off) select a from t1 where b = 5;
QUERY PLAN
Index Only Scan using i2 on t1 (actual time=0.022..0.023 rows=1 loops=1)
Index Cond: (b = 5)
Heap Fetches: 1
Buffers: shared hit=4
Planning Time: 0.056 ms
Execution Time: 0.068 ms
(6 rows)

Time: 138.462 ms
```

Si compara todos los resultados del plan después de analizar manualmente la tabla y recopilar estadísticas, observará que el plan 4 del optimizador es mejor que los demás y, además, reduce el tiempo de ejecución de las consultas. En este ejemplo se muestra la importancia de realizar actividades de mantenimiento en la base de datos.

Reescritura de una tabla completa con `VACUUM FULL`

Al ejecutar el comando `VACUUM` con el parámetro `FULL`, se reescribe todo el contenido de una tabla en un nuevo archivo de disco sin espacio adicional y el espacio no utilizado se devuelve al sistema operativo. Esta operación es mucho más lenta y requiere un bloqueo `ACCESS EXCLUSIVE` en cada tabla. También requiere espacio adicional en el disco, ya que escribe una nueva copia de la tabla y no libera la copia anterior hasta que se completa la operación.

`VACUUM FULL` puede ser de utilidad en los siguientes casos:

- Cuando desee recuperar una cantidad importante de espacio de las tablas.
- Cuando desee recuperar el espacio de sobrecarga en las tablas que no contienen claves principales.

Si la base de datos puede tolerar el tiempo de inactividad, le recomendamos que utilice `VACUUM FULL` cuando tenga tablas sin claves principales.

Como `VACUUM FULL` requiere más bloqueos que otras operaciones, es más caro ejecutarlo en las bases de datos cruciales. Para reemplazar este método, puede usar la extensión `pg_repack`, que se describe en la [siguiente sección](#). Esta opción es similar a `VACUUM FULL`, pero requiere un bloqueo mínimo, y es compatible tanto con Amazon RDS para PostgreSQL como con Aurora compatible con PostgreSQL.

Eliminación de la sobrecarga con pg_repack

Puede usar la extensión `pg_repack` para eliminar la sobrecarga de las tablas y los índices con un bloqueo mínimo de la base de datos. Puede crear esta extensión en la instancia de base de datos y ejecutar el cliente `pg_repack` (si la versión del cliente y la versión de la extensión coinciden) desde Amazon Elastic Compute Cloud (Amazon EC2) o desde un ordenador que se pueda conectar a la base de datos.

A diferencia de `VACUUM FULL`, `pg_repack` no necesita tiempo de inactividad ni un periodo de mantenimiento y no bloqueará otras sesiones.

`pg_repack` es útil en las situaciones en que es posible que `VACUUM FULL`, `CLUSTER` o `REINDEX` no funcionen. Crea una tabla nueva que contiene los datos de la tabla sobrecargada, realiza un seguimiento de los cambios con respecto a la tabla original y, a continuación, reemplaza la tabla original por la nueva. No bloquea la tabla original para operaciones de lectura o escritura mientras crea la nueva tabla.

Puede utilizar `pg_repack` para una tabla completa o para un índice. Para ver una lista de tareas, consulte la [documentación de pg_repack](#).

Limitaciones:

- Para ejecutar `pg_repack`, la tabla debe tener una clave principal o un índice único.
- `pg_repack` no funcionará con las tablas temporales.
- `pg_repack` no funcionará en las tablas que tengan índices globales.
- Mientras `pg_repack` esté en curso, no podrá realizar operaciones de DDL en las tablas.

En la siguiente tabla se describen las diferencias entre `pg_repack` y `VACUUM FULL`.

VACUUM FULL	pg_repack
Comando integrado	Una extensión que ejecuta desde Amazon EC2 o desde el ordenador local
Requiere un bloqueo ACCESS EXCLUSIVE mientras se utiliza en una tabla	Requiere un bloqueo ACCESS EXCLUSIVE solo durante un breve periodo de tiempo

Funciona con todas las tablas

Solo funciona en las tablas que tienen claves principales y únicas

Requiere el doble del espacio de almacenamiento que consumen la tabla y los índices

Requiere el doble del espacio de almacenamiento que consumen la tabla y los índices

Para ejecutar `pg_repack` en una tabla, use el siguiente comando:

```
pg_repack -h <host> -d <dbname> --table <tablename> -k
```

Para ejecutar `pg_repack` en un índice, use el siguiente comando:

```
pg_repack -h <host> -d <dbname> --index <index name>
```

Para obtener más información, consulte la entrada en el blog de AWS [Remove bloat from Amazon Aurora and RDS for PostgreSQL with pg_repack](#).

Advertencia

El error `error-on-invalid-index` de `pg_repack` suele indicar que uno o varios índices de la tabla están dañados o no son válidos. `pg_repack` no puede funcionar de forma segura en las tablas que no tienen índices válidos porque depende de los índices para mantener la coherencia de datos durante el proceso de reempaquetado.

Este error se produce en las siguientes situaciones:

- El índice está marcado como no válido (por ejemplo, debido a un error en una instrucción `CREATE INDEX CONCURRENTLY`).
- El índice está dañado (posiblemente debido a problemas de hardware o a cierres bruscos).

Utilice la siguiente consulta para identificar los índices no válidos y suprimirlos primero si los encuentra.

```
SELECT indexrelid::regclass, indisvalid FROM pg_index WHERE indrelid = 'orders'::regclass AND NOT indisvalid; Drop the invalid index: DROP INDEX index_name;
```

Nueva creación de índices

El comando [REINDEX](#) de PostgreSQL vuelve a crear un índice utilizando los datos que hay almacenados en la tabla del índice y reemplazando la copia antigua del índice. Le recomendamos utilizar REINDEX en las siguientes situaciones:

- Cuando un índice se dañe y deje de contener datos válidos. Esto puede ocurrir a consecuencia de errores de software o hardware.
- Cuando las consultas que anteriormente utilizaban el índice dejen de usarlo.
- Cuando el índice se sobrecargue con un gran número de páginas vacías o casi vacías. Debe ejecutar REINDEX cuando el porcentaje de sobrecarga (`bloat_pct`) sea superior a 20.

La siguiente consulta lo ayuda a encontrar `bloat_pct`:

```
SELECT current_database(), nspname AS schemaname, tblname, idxname,
bs*(relpages)::bigint AS real_size,
bs*(relpages-est_pages)::bigint AS extra_size,
100 * (relpages-est_pages)::float / relpages AS extra_pct,
fillfactor,
CASE WHEN relpages > est_pages_ff
  THEN bs*(relpages-est_pages_ff)
  ELSE 0
END AS bloat_size,
100 * (relpages-est_pages_ff)::float / relpages AS bloat_pct,
is_na
-- , 100-(pst).avg_leaf_density AS pst_avg_bloat, est_pages, index_tuple_hdr_bm,
maxalign, pagehdr, nulldatawidth, nulldatahdrwidth, reltuples, relpages -- (DEBUG
INFO)
FROM (
  SELECT coalesce(1 +
    ceil(reltuples/floor((bs-pageopqdata-pagehdr)/(4+nulldatahdrwidth)::float)), 0
  -- ItemIdData size + computed avg size of a tuple (nulldatahdrwidth)
  ) AS est_pages,
  coalesce(1 +
    ceil(reltuples/floor((bs-pageopqdata-pagehdr)*fillfactor/
(100*(4+nulldatahdrwidth)::float))), 0
  ) AS est_pages_ff,
  bs, nspname, tblname, idxname, relpages, fillfactor, is_na
```

```

-- , pgstatindex(idxoid) AS pst, index_tuple_hdr_bm, maxalign, pagehdr,
nulldatawidth, nulldatahdrwidth, reltuples -- (DEBUG INFO)
FROM (
    SELECT maxalign, bs, nspname, tblname, idxname, reltuples, relpages, idxoid,
fillfactor,
        ( index_tuple_hdr_bm +
            maxalign - CASE -- Add padding to the index tuple header to align on
MAXALIGN
                WHEN index_tuple_hdr_bm%maxalign = 0 THEN maxalign
                ELSE index_tuple_hdr_bm%maxalign
            END
        + nulldatawidth + maxalign - CASE -- Add padding to the data to align on
MAXALIGN
                WHEN nulldatawidth = 0 THEN 0
                WHEN nulldatawidth::integer%maxalign = 0 THEN maxalign
                ELSE nulldatawidth::integer%maxalign
            END
        )::numeric AS nulldatahdrwidth, pagehdr, pageopqdata, is_na
-- , index_tuple_hdr_bm, nulldatawidth -- (DEBUG INFO)
FROM (
    SELECT n.nspname, i.tblname, i.idxname, i.reltuples, i.relpages,
i.idxoid, i.fillfactor, current_setting('block_size')::numeric AS bs,
CASE -- MAXALIGN: 4 on 32bits, 8 on 64bits (and mingw32 ?)
    WHEN version() ~ 'mingw32' OR version() ~ '64-bit|x86_64|ppc64|ia64|
amd64' THEN 8
    ELSE 4
END AS maxalign,
/* per page header, fixed size: 20 for 7.X, 24 for others */
24 AS pagehdr,
/* per page btree opaque data */
16 AS pageopqdata,
/* per tuple header: add IndexAttributeBitMapData if some cols are null-
able */
CASE WHEN max(coalesce(s.null_frac,0)) = 0
    THEN 8 -- IndexTupleData size
    ELSE 8 + (( 32 + 8 - 1 ) / 8) -- IndexTupleData size +
IndexAttributeBitMapData size ( max num filed per index + 8 - 1 /8)
END AS index_tuple_hdr_bm,
/* data len: we remove null values save space using it fractionnal part
from stats */
sum( (1-coalesce(s.null_frac, 0)) * coalesce(s.avg_width, 1024)) AS
nulldatawidth,

```

```

        max( CASE WHEN i.atttypid = 'pg_catalog.name'::regtype THEN 1 ELSE 0
END ) > 0 AS is_na
    FROM (
        SELECT ct.relname AS tblname, ct.relnamespace, ic.idxname, ic.attpos,
ic.indkey, ic.indkey[ic.attpos], ic.reltuples, ic.relpages, ic.tbloid, ic.idxoid,
ic.fillfactor,
            coalesce(a1.attnum, a2.attnum) AS attnum, coalesce(a1.attname,
a2.attname) AS attname, coalesce(a1.atttypid, a2.atttypid) AS atttypid,
            CASE WHEN a1.attnum IS NULL
            THEN ic.idxname
            ELSE ct.relname
            END AS attrelname
    FROM (
        SELECT idxname, reltuples, relpages, tbloid, idxoid, fillfactor,
indkey,
            pg_catalog.generate_series(1,indnatts) AS attpos
    FROM (
        SELECT ci.relname AS idxname, ci.reltuples, ci.relpages,
i.indrelid AS tbloid,
            i.indexrelid AS idxoid,
            coalesce(substring(
                array_to_string(ci.reloptions, ' ')
                from 'fillfactor=([0-9]+)')::smallint, 90) AS fillfactor,
            i.indnatts,
            pg_catalog.string_to_array(pg_catalog.textin(
                pg_catalog.int2vectorout(i.indkey)), ' ')::int[] AS indkey
    FROM pg_catalog.pg_index i
    JOIN pg_catalog.pg_class ci ON ci.oid = i.indexrelid
    WHERE ci.relam=(SELECT oid FROM pg_am WHERE amname = 'btree')
    AND ci.relpages > 0
    ) AS idx_data
    ) AS ic
    JOIN pg_catalog.pg_class ct ON ct.oid = ic.tbloid
    LEFT JOIN pg_catalog.pg_attribute a1 ON
        ic.indkey[ic.attpos] <> 0
        AND a1.attrelid = ic.tbloid
        AND a1.attnum = ic.indkey[ic.attpos]
    LEFT JOIN pg_catalog.pg_attribute a2 ON
        ic.indkey[ic.attpos] = 0
        AND a2.attrelid = ic.idxoid
        AND a2.attnum = ic.attpos
    ) i
    JOIN pg_catalog.pg_namespace n ON n.oid = i.relnamespace

```

```

        JOIN pg_catalog.pg_stats s ON s.schemaname = n.nspname
                                AND s.tablename = i.attrelname
                                AND s.attname = i.attname
        GROUP BY 1,2,3,4,5,6,7,8,9,10,11
    ) AS rows_data_stats
) AS rows_hdr_pdg_stats
) AS relation_stats
ORDER BY nspname, tblname, idxname;

```

Las páginas de índice que están completamente vacías se recuperan para su reutilización. Sin embargo, recomendamos realizar una nueva indexación periódicamente si se han eliminado las claves de índice de una página pero se conserva el espacio asignado.

Volver a crear el índice ayuda a mejorar el rendimiento de las consultas. Puede volver a crear un índice de tres maneras, tal y como se describe en la siguiente tabla.

Método	Descripción	Limitaciones
CREATE INDEX y DROP INDEX con la opción CONCURRENTLY	Crea un índice nuevo y elimina el antiguo. El optimizador genera planes utilizando el índice recién creado en lugar del índice antiguo. Durante las horas de menor actividad , puede eliminar el índice antiguo.	La creación del índice lleva más tiempo cuando se utiliza la opción CONCURRENTLY , ya que tiene que hacer un seguimiento de todos los cambios entrantes. Cuando los cambios se congelan, el proceso se marca como completo.
REINDEX con la opción CONCURRENTLY	Bloquea las operaciones de escritura durante el nuevo proceso de creación. La versión 12 y las versiones posteriores de PostgreSQL ofrecen la opción CONCURRENTLY , lo que evita estos bloqueos.	El uso de CONCURRENTLY requiere más tiempo para volver a crear el índice.

Extensión `pg_repack`

Limpia la sobrecarga de una tabla y vuelve a crear el índice.

Debe ejecutar esta extensión desde una instancia de EC2 o desde una computadora local que esté conectada a la base de datos.

Creación de un nuevo índice

Cuando los comandos `DROP INDEX` y `CREATE INDEX` se utilizan juntos, vuelven a crear un índice:

```
DROP INDEX <index_name>
CREATE INDEX <index_name> ON TABLE <table_name> (<column1>[,<column2>])
```

La desventaja de este enfoque es el bloqueo de exclusividad que se aplica a la tabla, lo que afecta al rendimiento durante esta actividad. El comando `DROP INDEX` adquiere un bloqueo de exclusividad que bloquea las operaciones de lectura y de escritura en la tabla. El comando `CREATE INDEX` bloquea las operaciones de escritura en la tabla. Permite operaciones de lectura, pero tienen un costo elevado durante la creación del índice.

Nueva creación de un índice

El comando `REINDEX` lo ayuda a mantener un rendimiento uniforme de la base de datos. Al realizar un gran número de operaciones con DML en una tabla, se produce una sobrecarga tanto de la tabla como de los índices. Los índices se utilizan para acelerar las búsquedas en las tablas y mejorar el rendimiento de las consultas. La sobrecarga de los índices afecta al rendimiento de las búsquedas y de las consultas. Por lo tanto, es recomendable volver a indexar las tablas que tienen un gran volumen de operaciones con DML para mantener la coherencia en el rendimiento de las consultas.

El comando `REINDEX` bloquea las operaciones de escritura en la tabla subyacente para volver a crear el índice desde cero, pero permite las operaciones de lectura en la tabla. Sin embargo, sí bloquea las operaciones de lectura en el índice. Las consultas que utilizan el índice correspondiente están bloqueadas, pero las demás no.

La versión 12 de PostgreSQL introdujo un nuevo parámetro opcional, `CONCURRENTLY`, que vuelve a crear el índice desde cero, pero no bloquea las operaciones de lectura ni escritura en la tabla ni en

las consultas que utilizan el índice. Sin embargo, se tarda más en completar el proceso cuando se usa esta opción.

Ejemplos

Creación y eliminación de un índice

Cree un índice nuevo con la opción CONCURRENTLY:

```
create index CONCURRENTLY on table(columns) ;
```

Elimine el índice antiguo con la opción CONCURRENTLY:

```
drop index CONCURRENTLY <index name> ;
```

Nueva creación de un índice

Para volver a crear un índice único:

```
reindex index <index name> ;
```

Para volver a crear todos los índices de una tabla:

```
reindex table <table name> ;
```

Para volver a crear todos los índices de un esquema:

```
reindex schema <schema name> ;
```

Nueva creación de un índice de forma simultánea

Para volver a crear un índice único:

```
reindex index CONCURRENTLY <indexname> ;
```

Para volver a crear todos los índices de una tabla:

```
reindex table CONCURRENTLY <tablename> ;
```

Para volver a crear todos los índices de un esquema:

```
reindex schema CONCURRENTLY <schemaname> ;
```

Nueva creación o reubicación únicamente de índices

Para volver a crear un índice único:

```
pg_repack -h <hostname> -d <dbname> -i <indexname> -k
```

Para volver a crear todos los índices:

```
pg_repack -h <hostname> -d <dbname> -x <indexname> -t <tablename> -k
```

Ejemplo: Recuperación de espacio mediante autovacuum y VACUUM FULL

Como ejemplo, vamos a crear una tabla emp con 500 000 filas y, a continuación, vamos a actualizar las filas con nuevos valores. Autovacuum está activado, por lo que ejecutará los comandos VACUUM y ANALYZE de esta tabla para eliminar la sobrecarga y recuperar espacio. El espacio recuperado se puede reutilizar, pero no se devolverá al sistema operativo.

La siguiente consulta determina la sobrecarga de la tabla:

```
-- WARNING: When run with a non-superuser role, the query inspects only indexes on
tables you are granted to read.
-- WARNING: Rows with is_na = 't' are known to have bad statistics ("name" type is not
supported).
-- This query is compatible with PostgreSQL 8.2 and later.
SELECT current_database(), nspname AS schemaname, tblname, idxname,
bs*(relpages)::bigint AS real_size,
bs*(relpages-est_pages)::bigint AS extra_size,
100 * (relpages-est_pages)::float / relpages AS extra_pct,
fillfactor,
CASE WHEN relpages > est_pages_ff
THEN bs*(relpages-est_pages_ff)
ELSE 0
END AS bloat_size,
100 * (relpages-est_pages_ff)::float / relpages AS bloat_pct,
is_na
-- , 100-(pst).avg_leaf_density AS pst_avg_bloat, est_pages, index_tuple_hdr_bm,
maxalign, pagehdr, nulldatawidth, nulldatahdrwidth, reltuples, relpages -- (DEBUG
INFO)
FROM (
SELECT coalesce(1 +
ceil(reltuples/floor((bs-pageopqdata-pagehdr)/(4+nulldatahdrwidth)::float)), 0
-- ItemIdData size + computed avg size of a tuple (nulldatahdrwidth)
) AS est_pages,
coalesce(1 +
ceil(reltuples/floor((bs-pageopqdata-pagehdr)*fillfactor/
(100*(4+nulldatahdrwidth)::float))), 0
) AS est_pages_ff,
bs, nspname, tblname, idxname, relpages, fillfactor, is_na
```

```

-- , pgstatindex(idxoid) AS pst, index_tuple_hdr_bm, maxalign, pagehdr,
nulldatawidth, nulldatahdrwidth, reltuples -- (DEBUG INFO)
FROM (
    SELECT maxalign, bs, nspname, tblname, idxname, reltuples, relpages, idxoid,
fillfactor,
        ( index_tuple_hdr_bm +
            maxalign - CASE -- Add padding to the index tuple header to align on
MAXALIGN
                WHEN index_tuple_hdr_bm%maxalign = 0 THEN maxalign
                ELSE index_tuple_hdr_bm%maxalign
            END
        + nulldatawidth + maxalign - CASE -- Add padding to the data to align on
MAXALIGN
                WHEN nulldatawidth = 0 THEN 0
                WHEN nulldatawidth::integer%maxalign = 0 THEN maxalign
                ELSE nulldatawidth::integer%maxalign
            END
        )::numeric AS nulldatahdrwidth, pagehdr, pageopqdata, is_na
-- , index_tuple_hdr_bm, nulldatawidth -- (DEBUG INFO)
FROM (
    SELECT n.nspname, i.tblname, i.idxname, i.reltuples, i.relpages,
        i.idxoid, i.fillfactor, current_setting('block_size')::numeric AS bs,
        CASE -- MAXALIGN: 4 on 32bits, 8 on 64bits (and mingw32 ?)
            WHEN version() ~ 'mingw32' OR version() ~ '64-bit|x86_64|ppc64|ia64|
amd64' THEN 8
            ELSE 4
        END AS maxalign,
        /* per page header, fixed size: 20 for 7.X, 24 for others */
        24 AS pagehdr,
        /* per page btree opaque data */
        16 AS pageopqdata,
        /* per tuple header: add IndexAttributeBitMapData if some cols are null-
able */
        CASE WHEN max(coalesce(s.null_frac,0)) = 0
            THEN 8 -- IndexTupleData size
            ELSE 8 + (( 32 + 8 - 1 ) / 8) -- IndexTupleData size +
IndexAttributeBitMapData size ( max num filed per index + 8 - 1 /8)
        END AS index_tuple_hdr_bm,
        /* data len: we remove null values save space using it fractionnal part
from stats */
        sum( (1-coalesce(s.null_frac, 0)) * coalesce(s.avg_width, 1024)) AS
nulldatawidth,

```

```

        max( CASE WHEN i.atttypid = 'pg_catalog.name'::regtype THEN 1 ELSE 0
END ) > 0 AS is_na
    FROM (
        SELECT ct.relname AS tblname, ct.relnamespace, ic.idxname, ic.attpos,
ic.indkey, ic.indkey[ic.attpos], ic.reltuples, ic.relpages, ic.tbloid, ic.idxoid,
ic.fillfactor,
            coalesce(a1.attnum, a2.attnum) AS attnum, coalesce(a1.attname,
a2.attname) AS attname, coalesce(a1.atttypid, a2.atttypid) AS atttypid,
            CASE WHEN a1.attnum IS NULL
            THEN ic.idxname
            ELSE ct.relname
            END AS attrelname
        FROM (
            SELECT idxname, reltuples, relpages, tbloid, idxoid, fillfactor,
indkey,
                pg_catalog.generate_series(1,indnatts) AS attpos
            FROM (
                SELECT ci.relname AS idxname, ci.reltuples, ci.relpages,
i.indrelid AS tbloid,
                    i.indexrelid AS idxoid,
                    coalesce(substring(
                        array_to_string(ci.reloptions, ' ')
                        from 'fillfactor=([0-9]+)')::smallint, 90) AS fillfactor,
                    i.indnatts,
                    pg_catalog.string_to_array(pg_catalog.textin(
                        pg_catalog.int2vectorout(i.indkey)), ' ')::int[] AS indkey
                FROM pg_catalog.pg_index i
                JOIN pg_catalog.pg_class ci ON ci.oid = i.indexrelid
                WHERE ci.relam=(SELECT oid FROM pg_am WHERE amname = 'btree')
                AND ci.relpages > 0
            ) AS idx_data
        ) AS ic
    JOIN pg_catalog.pg_class ct ON ct.oid = ic.tbloid
    LEFT JOIN pg_catalog.pg_attribute a1 ON
        ic.indkey[ic.attpos] <> 0
        AND a1.attrelid = ic.tbloid
        AND a1.attnum = ic.indkey[ic.attpos]
    LEFT JOIN pg_catalog.pg_attribute a2 ON
        ic.indkey[ic.attpos] = 0
        AND a2.attrelid = ic.idxoid
        AND a2.attnum = ic.attpos
    ) i
    JOIN pg_catalog.pg_namespace n ON n.oid = i.relnamespace

```



```
autoanalyze_count | 3
```

Observe que autovacuum actualizó las columnas `last_autovacuum` y `last_autoanalyze` después de ejecutarse.

Ahora, insertemos algunas filas en la tabla y comprobemos `extra_size(bloat_size)`, ya que el espacio vacío también se considera sobrecarga.

```
apgl=> select count(*) from emp;
count | 900000

current_database | schemaname | tblname | real_size | extra_size | extra_pct |
fillfactor | bloat_size | bloat_pct | is_na
-----+-----+-----+-----+-----+-----
+-----+-----+-----+-----+-----
apgl | public | emp | 61349888 | 327680 | 0.5341167044999332 | 100 | 327680 |
0.5341167044999332 | f
(1 row)
```

La columna `bloat_pct` de la salida indica que hay nuevas inserciones que han ocupado el espacio limpio. Ejecutemos `VACUUM FULL`:

```
apgl=> vacuum full emp ;
VACUUM

current_database | schemaname | tblname | real_size | extra_size | extra_pct |
fillfactor | bloat_size | bloat_pct | is_na
-----+-----+-----+-----+-----+-----
+-----+-----+-----+-----
apgl | public | emp | 60792832 | -229376 | 0 | 100 | 0 | 0 | f
(1 row)
```

En esta salida, puede ver que se han eliminado el espacio vacío y la sobrecarga y que el espacio se ha devuelto al sistema operativo.

Note

En lugar de `VACUUM FULL`, puede ejecutar `pg_repack` para obtener los mismos resultados.

Recursos

- [Descripción del autovacuum en entornos Amazon RDS para AWS PostgreSQL](#) (entrada del blog)
- [Automatic Vacuuming](#) (documentación de PostgreSQL)
- [Asignación de memoria para autovacuum](#) (documentación de Amazon RDS)
- [Prevención de errores envolventes en los identificadores de transacciones; \(documentación de PostgreSQL\)](#)
- [Remove bloat from Amazon Aurora and RDS for PostgreSQL with pg_repack](#) (entrada en el blog de AWS)

Historial de documentos

En la siguiente tabla, se describen cambios significativos de esta guía. Si quiere recibir notificaciones de futuras actualizaciones, puede suscribirse a las [notificaciones RSS](#).

Cambio	Descripción	Fecha
Actualizaciones	Se corrigieron errores y se agregó información a las secciones Succión y análisis automáticos de las tablas y Eliminación de la sobrecarga con pg_repack .	22 de agosto de 2025
Se corrigió la sintaxis de reindex.	En la sección dedicada a la nueva creación simultánea de un índice , se corrigieron los ejemplos de reindex.	30 de junio de 2025
Publicación inicial	—	22 de diciembre de 2023

AWS Glosario de orientación prescriptiva

Los siguientes son términos de uso común en las estrategias, guías y patrones proporcionados por la Guía AWS prescriptiva. Para sugerir entradas, utilice el enlace [Enviar comentarios](#) al final del glosario.

Números

Las 7 R

Siete estrategias de migración comunes para trasladar aplicaciones a la nube. Estas estrategias se basan en las 5 R que Gartner identificó en 2011 y consisten en lo siguiente:

- **Refactor/re-architect** — Mueva una aplicación y modifique su arquitectura aprovechando al máximo las funciones nativas de la nube para mejorar la agilidad, el rendimiento y la escalabilidad. Por lo general, esto implica trasladar el sistema operativo y la base de datos. Ejemplo: migre su base de datos Oracle local a la PostgreSQL-Compatible edición Amazon Aurora.
- **Redefinir la plataforma (transportar y redefinir)**: traslade una aplicación a la nube e introduzca algún nivel de optimización para aprovechar las capacidades de la nube. Ejemplo: Migrar la base de datos Oracle en las instalaciones a Amazon Relational Database Service (Amazon RDS) para Oracle en la nube de Nube de AWS.
- **Recomprar (readquirir)**: cambie a un producto diferente, lo cual se suele llevar a cabo al pasar de una licencia tradicional a un modelo SaaS. Ejemplo: migre su sistema de gestión de relaciones con los clientes (CRM) a Salesforce.com.
- **Volver a alojar (migrar mediante lift-and-shift)**: traslade una aplicación a la nube sin hacer cambios para aprovechar las funcionalidades de la nube. Ejemplo: Migrar la base de datos de Oracle en las instalaciones a Oracle en una instancia de EC2 en la Nube de AWS.
- **Reubicar**: (migrar el hipervisor mediante lift and shift): traslade la infraestructura a la nube sin comprar equipo nuevo, reescribir aplicaciones o modificar las operaciones actuales. Los servidores se migran de una plataforma en las instalaciones a un servicio en la nube para la misma plataforma. Ejemplo: migrar una Microsoft Hyper-V aplicación a AWS.
- **Retener (revisitar)**: conserve las aplicaciones en el entorno de origen. Estas pueden incluir las aplicaciones que requieren una refactorización importante, que desee posponer para más

adelante, y las aplicaciones heredadas que desee retener, ya que no hay ninguna justificación empresarial para migrarlas.

- Retirar: retire o elimine las aplicaciones que ya no sean necesarias en un entorno de origen.

A

A2A () Agent-to-Agent

Un protocolo completo para la colaboración entre agentes que facilita la delegación de tareas y la transferencia de estados.

ABAC

Consulte [control de acceso basado en atributos](#).

servicios abstractos

Consulte [servicios administrados](#).

ACID

Consulte [atomicidad, consistencia, aislamiento, durabilidad](#).

migración activa-activa

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas (mediante una herramienta de replicación bidireccional o mediante operaciones de escritura doble) y ambas bases de datos gestionan las transacciones de las aplicaciones conectadas durante la migración. Este método permite la migración en lotes pequeños y controlados, en lugar de requerir una transición única. Es más flexible, pero requiere más trabajo que una [migración activa-pasiva](#).

migración activa-pasiva

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas, pero solo la de origen gestiona las transacciones de las aplicaciones conectadas, mientras los datos se replican en la de destino. La base de datos de destino no acepta ninguna transacción durante la migración.

Agente

Un sistema de IA que puede razonar, planificar y tomar medidas de forma autónoma utilizando herramientas para alcanzar los objetivos.

Agent Ops

Prácticas operativas para crear, probar, implementar y ejecutar agentes de IA en producción a escala.

función de agregación

Función SQL que actúa en un grupo de filas y calcula un único valor de devolución para el grupo. Entre los ejemplos de funciones de agregación se incluyen SUM y MAX.

IA

Consulte [inteligencia artificial](#).

AIOps

Consulte [operaciones de inteligencia artificial](#)

anonimización

El proceso de eliminar permanentemente la información personal de un conjunto de datos. La anonimización puede ayudar a proteger la privacidad personal. Los datos anonimizados ya no se consideran datos personales.

antipatronos

Una solución que se utiliza con frecuencia para un problema recurrente en el que la solución es contraproducente, ineficaz o menos eficaz que una alternativa.

control de aplicaciones

Enfoque de seguridad que permite usar de manera exclusiva aplicaciones aprobadas para ayudar a proteger un sistema contra el malware.

cartera de aplicaciones

Recopilación de información detallada sobre cada aplicación que utiliza una organización, incluido el costo de creación y mantenimiento de la aplicación y su valor empresarial. Esta información es clave para [el proceso de detección y análisis de la cartera](#) y ayuda a identificar y priorizar las aplicaciones que se van a migrar, modernizar y optimizar.

inteligencia artificial (IA)

El campo de la informática que se dedica al uso de tecnologías informáticas para realizar funciones cognitivas que suelen estar asociadas a los seres humanos, como el aprendizaje, la

resolución de problemas y el reconocimiento de patrones. Para más información, consulte [¿Qué es la inteligencia artificial?](#)

operaciones de inteligencia artificial (AIOps)

El proceso de utilizar técnicas de machine learning para resolver problemas operativos, reducir los incidentes operativos y la intervención humana, y mejorar la calidad del servicio. Para obtener más información sobre cómo se utiliza AIOps en la estrategia de migración de AWS, consulte la [Guía de integración de operaciones](#).

cifrado asimétrico

Algoritmo de cifrado que utiliza un par de claves, una clave pública para el cifrado y una clave privada para el descifrado. Puede compartir la clave pública porque no se utiliza para el descifrado, pero el acceso a la clave privada debe estar sumamente restringido.

atomicidad, consistencia, aislamiento, durabilidad (ACID)

Conjunto de propiedades de software que garantizan la validez de los datos y la fiabilidad operativa de una base de datos, incluso en caso de errores, cortes de energía u otros problemas.

control de acceso basado en atributos (ABAC)

La práctica de crear permisos detallados basados en los atributos del usuario, como el departamento, el puesto de trabajo y el nombre del equipo. Para obtener más información, consulte [ABAC AWS en la](#) documentación AWS Identity and Access Management (IAM).

origen de datos fidedigno

Ubicación en la que se almacena la versión principal de los datos, que se considera la fuente de información más fiable. Puede copiar los datos del origen de datos autorizado a otras ubicaciones con el fin de procesarlos o modificarlos, por ejemplo, anonimizarlos, redactarlos o seudonimizarlos.

Zona de disponibilidad

Una ubicación distinta dentro de una Región de AWS que está aislada de los fallos en otras zonas de disponibilidad y que proporciona una conectividad de red económica y de baja latencia a otras zonas de disponibilidad de la misma región.

AWS Marco de adopción de la nube (AWS CAF)

Un marco de directrices y mejores prácticas AWS para ayudar a las organizaciones a desarrollar un plan eficiente y eficaz para migrar con éxito a la nube. AWS CAF organiza la orientación en

seis áreas de enfoque denominadas perspectivas: negocios, personas, gobierno, plataforma, seguridad y operaciones. Las perspectivas empresariales, humanas y de gobernanza se centran en las habilidades y los procesos empresariales; las perspectivas de plataforma, seguridad y operaciones se centran en las habilidades y los procesos técnicos. Por ejemplo, la perspectiva humana se dirige a las partes interesadas que se ocupan de los Recursos Humanos (RR. HH.), las funciones del personal y la administración de las personas. Desde esta perspectiva, AWS CAF proporciona orientación para el desarrollo, la formación y la comunicación de las personas a fin de preparar a la organización para una adopción exitosa de la nube. Para obtener más información, consulte la [Página web de AWS CAF](#) y el [Documento técnico de AWS CAF](#).

AWS Marco de calificación de la carga de trabajo (AWS WQF)

Herramienta que evalúa las cargas de trabajo de migración de bases de datos, recomienda estrategias de migración y proporciona estimaciones de trabajo. AWS WQF se incluye con AWS Schema Conversion Tool (). AWS SCT Analiza los esquemas de bases de datos y los objetos de código, el código de las aplicaciones, las dependencias y las características de rendimiento y proporciona informes de evaluación.

B

bot malicioso

[Bot](#) destinado a causar interrupciones o daños a personas u organizaciones.

BCP

Consulte [planificación de la continuidad del negocio](#).

gráfico de comportamiento

Una vista unificada e interactiva del comportamiento de los recursos y de las interacciones a lo largo del tiempo. Puede utilizar un gráfico de comportamiento con Amazon Detective para examinar los intentos de inicio de sesión fallidos, las llamadas sospechosas a la API y acciones similares. Para obtener más información, consulte [Datos en un gráfico de comportamiento](#) en la documentación de Detective.

sistema big-endian

Un sistema que almacena primero el byte más significativo. Consulte también [endianidad](#).

clasificación binaria

Un proceso que predice un resultado binario (una de las dos clases posibles). Por ejemplo, es posible que su modelo de ML necesite predecir problemas como “¿Este correo electrónico es spam o no es spam?” o “¿Este producto es un libro o un automóvil?”.

filtro de floración

Estructura de datos probabilística y eficiente en términos de memoria que se utiliza para comprobar si un elemento es miembro de un conjunto.

blue/green despliegue

Estrategia de implementación en la que se crean dos entornos separados, pero idénticos. La versión actual de la aplicación se ejecuta en un entorno (azul) y la nueva versión de la aplicación se ejecuta en el otro entorno (verde). Esta estrategia lo ayuda a hacer reversiones rápidas con un impacto mínimo.

bot

Aplicación de software que ejecuta tareas automatizadas a través de Internet y simula la actividad o interacción humana. Algunos bots son útiles o beneficiosos, como los rastreadores web que indexan la información de Internet. Otros bots, conocidos como bots maliciosos, tienen como objetivo causar interrupciones o daños a personas u organizaciones.

botnet

Redes de [bots](#) infectadas por [malware](#) y que están bajo el control de una sola parte, conocida como pastor de bots u operador de bots. Las botnets son el mecanismo más conocido para escalar los bots y su impacto.

branch

Área contenida de un repositorio de código. La primera rama que se crea en un repositorio es la rama principal. Puede crear una rama nueva a partir de una rama existente y, a continuación, desarrollar características o corregir errores en la rama nueva. Una rama que se genera para crear una característica se denomina comúnmente rama de característica. Cuando la característica se encuentra lista para su lanzamiento, se vuelve a combinar la rama de característica con la rama principal. Para obtener más información, consulte [Acerca de las sucursales](#) (GitHub documentación).

acceso de emergencia

En circunstancias excepcionales y mediante un proceso aprobado, es una forma rápida de que un usuario pueda acceder a un Cuenta de AWS sitio al que normalmente no tiene permisos de acceso. Para obtener más información, consulte el indicador de [implementación de procedimientos rompe-cristales](#) en la AWS Well-Architected guía.

estrategia de implementación sobre infraestructura existente

La infraestructura existente en su entorno. Al adoptar una estrategia de implementación sobre infraestructura existente para una arquitectura de sistemas, se diseña la arquitectura en función de las limitaciones de los sistemas y la infraestructura actuales. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de [implementación desde cero](#).

caché de búfer

El área de memoria donde se almacenan los datos a los que se accede con más frecuencia.

capacidad empresarial

Lo que hace una empresa para generar valor (por ejemplo, ventas, servicio al cliente o marketing). Las arquitecturas de microservicios y las decisiones de desarrollo pueden estar impulsadas por las capacidades empresariales. Para obtener más información, consulte la sección [Organizado en torno a las capacidades empresariales](#) del documento técnico [Ejecutar microservicios en contenedores en AWS](#).

planificación de la continuidad del negocio (BCP)

Plan que aborda el posible impacto de un evento disruptivo, como una migración a gran escala en las operaciones y permite a la empresa reanudar las operaciones rápidamente.

C

CAF

Consulte [AWS Cloud Adoption Framework](#).

implementación canario

Lanzamiento lento e incremental de una versión para los usuarios finales. Cuando tenga mayor confianza en la nueva versión, la implementa y reemplaza la versión actual en su totalidad.

CCoE

Consulte [Centro de excelencia en la nube](#).

CDC

Consulte [captura de datos de cambios](#).

captura de datos de cambio (CDC)

Proceso de seguimiento de los cambios en un origen de datos, como una tabla de base de datos, y registro de los metadatos relacionados con el cambio. Puede utilizar los CDC para diversos fines, como auditar o replicar los cambios en un sistema de destino para mantener la sincronización.

ingeniería del caos

Introducción intencionada de fallos o eventos disruptivos para poner a prueba la resiliencia de un sistema. Puedes usar [AWS Fault Injection Service \(AWS FIS\)](#) para realizar experimentos que estresen tus AWS cargas de trabajo y evalúen su respuesta.

CI/CD

Consulte [integración continua y entrega continua](#).

clasificación

Un proceso de categorización que permite generar predicciones. Los modelos de ML para problemas de clasificación predicen un valor discreto. Los valores discretos siempre son distintos entre sí. Por ejemplo, es posible que un modelo necesite evaluar si hay o no un automóvil en una imagen.

Desarrollador ciudadano

Un usuario empresarial que crea aplicaciones de IA utilizando plataformas sin code/low código sin conocimientos técnicos especializados.

cifrado del cliente

Cifrado de datos localmente, antes de que el objetivo los Servicio de AWS reciba.

Centro de excelencia en la nube (CCoE)

Equipo multidisciplinario que impulsa los esfuerzos de adopción de la nube en toda la organización, incluido el desarrollo de las prácticas recomendadas en la nube, la movilización de recursos, el establecimiento de plazos de migración y la dirección de la organización durante

las transformaciones a gran escala. Para obtener más información, consulte las [publicaciones de CCoE](#) en el blog de estrategia Nube de AWS empresarial.

computación en la nube

La tecnología en la nube que se utiliza normalmente para la administración de dispositivos de IoT y el almacenamiento de datos de forma remota. La computación en la nube suele estar relacionada con la tecnología de [computación de periferia](#).

modelo operativo en la nube

En una organización de TI, el modelo operativo que se utiliza para crear, madurar y optimizar uno o más entornos de nube. Para obtener más información, consulte [Creación de su modelo operativo de nube](#).

etapas de adopción de la nube

Las siguientes son las cuatro fases por las que suelen pasar las empresas cuando migran a la Nube de AWS:

- Proyecto: ejecución de algunos proyectos relacionados con la nube con fines de prueba de concepto y aprendizaje
- Fundamento: realización de inversiones fundamentales para escalar la adopción de la nube (p. ej., crear una zona de aterrizaje, definir un CCoE, establecer un modelo de operaciones)
- Migración: migración de aplicaciones individuales
- Re-invention — Optimizar los productos y servicios e innovar en la nube

Stephen Orban definió estas etapas en la entrada del blog The [Journey Toward Cloud-First & the Stages of Adoption del](#) blog Nube de AWS Enterprise Strategy. Para obtener información sobre su relación con la estrategia de AWS migración, consulte la [guía de preparación para la migración](#).

CMDB

Consulte [base de datos de administración de configuración](#).

repositorio de código

Una ubicación donde el código fuente y otros activos, como documentación, muestras y scripts, se almacenan y actualizan mediante procesos de control de versiones. Algunos repositorios en la nube comunes son GitHub o Bitbucket Cloud. Cada versión del código se denomina rama. En una estructura de microservicios, cada repositorio se encuentra dedicado a una única funcionalidad. Una sola CI/CD canalización puede utilizar varios repositorios.

caché en frío

Una caché de búfer que está vacía no está bien poblada o contiene datos obsoletos o irrelevantes. Esto afecta al rendimiento, ya que la instancia de la base de datos debe leer desde la memoria principal o el disco, lo que es más lento que leer desde la memoria caché del búfer.

datos fríos

Datos a los que se accede con poca frecuencia y que suelen ser históricos. Al consultar este tipo de datos, normalmente se aceptan consultas lentas. Trasladar estos datos a niveles o clases de almacenamiento de menor rendimiento y menos costosos puede reducir los costos.

visión artificial (CV)

Campo de la [IA](#) que utiliza el machine learning para analizar y extraer información de formatos visuales, como imágenes y videos digitales. Por ejemplo, Amazon SageMaker AI proporciona algoritmos de procesamiento de imágenes para CV.

deriva de configuración

En el caso de una carga de trabajo, un cambio en la configuración con respecto al estado esperado. Podría provocar que la carga de trabajo deje de cumplir las normas y, por lo general, es gradual e involuntaria.

base de datos de administración de configuración (CMDB)

Repositorio que almacena y administra información sobre una base de datos y su entorno de TI, incluidos los componentes de hardware y software y sus configuraciones. Por lo general, los datos de una CMDB se utilizan en la etapa de detección y análisis de la cartera de productos durante la migración.

paquete de conformidad

Un conjunto de AWS Config reglas y medidas correctivas que puede reunir para personalizar sus controles de conformidad y seguridad. Puede implementar un paquete de conformidad como una entidad única en una región Cuenta de AWS y, o en una organización, mediante una plantilla YAML. Para obtener más información, consulta los [paquetes de conformidad](#) en la documentación. AWS Config

integración y entrega continuas (I) CI/CD

El proceso de automatización de las etapas de origen, creación, prueba, puesta en escena y producción del proceso de publicación del software. CI/CD se describe comúnmente como una

canalización. CI/CD puede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar la calidad del código y entregar más rápido. Para obtener más información, consulte [Beneficios de la entrega continua](#). CD también puede significar implementación continua. Para obtener más información, consulte [Entrega continua frente a implementación continua](#).

CV

Consulte [visión artificial](#).

D

datos en reposo

Datos que están estacionarios en la red, como los datos que se encuentran almacenados.

clasificación de datos

Un proceso para identificar y clasificar los datos de su red en función de su importancia y sensibilidad. Es un componente fundamental de cualquier estrategia de administración de riesgos de ciberseguridad porque lo ayuda a determinar los controles de protección y retención adecuados para los datos. La clasificación de los datos es un componente del pilar de seguridad del AWS Well-Architected Framework. Para obtener más información, consulte [Clasificación de datos](#).

deriva de datos

Una variación significativa entre los datos de producción y los datos que se utilizaron para entrenar un modelo de machine learning, o un cambio significativo en los datos de entrada a lo largo del tiempo. La deriva de datos puede reducir la calidad, la precisión y la imparcialidad generales de las predicciones de los modelos de machine learning.

datos en tránsito

Datos que se mueven de forma activa por la red, por ejemplo, entre los recursos de la red.

mallado de datos

Marco de arquitectura que proporciona una propiedad de datos distribuida y descentralizada con una administración y una gobernanza centralizadas.

minimización de datos

El principio de recopilar y procesar solo los datos estrictamente necesarios. Practicar la minimización de los datos Nube de AWS puede reducir los riesgos de privacidad, los costos y la huella de carbono de la analítica.

perímetro de datos

Un conjunto de barreras preventivas en su AWS entorno que ayudan a garantizar que solo las identidades confiables accedan a los recursos confiables desde las redes esperadas. Para obtener más información, consulte [Crear un perímetro de datos sobre](#). AWS

preprocesamiento de datos

Transformar los datos sin procesar en un formato que su modelo de ML pueda analizar fácilmente. El preprocesamiento de datos puede implicar eliminar determinadas columnas o filas y corregir los valores faltantes, incoherentes o duplicados.

procedencia de los datos

El proceso de rastrear el origen y el historial de los datos a lo largo de su ciclo de vida, por ejemplo, la forma en que se generaron, transmitieron y almacenaron los datos.

titular de los datos

Persona cuyos datos se recopilan y procesan.

almacenamiento de datos

Sistema de administración de datos que respalda la inteligencia empresarial, como los análisis. Los almacenes de datos suelen contener grandes cantidades de datos históricos y, por lo general, se utilizan para las consultas y los análisis.

lenguaje de definición de datos (DDL)

Instrucciones o comandos para crear o modificar la estructura de tablas y objetos de una base de datos.

lenguaje de manipulación de datos (DML)

Instrucciones o comandos para modificar (insertar, actualizar y eliminar) la información de una base de datos.

DDL

Consulte [lenguaje de definición de bases de datos](#).

conjunto profundo

Combinar varios modelos de aprendizaje profundo para la predicción. Puede utilizar conjuntos profundos para obtener una predicción más precisa o para estimar la incertidumbre de las predicciones.

aprendizaje profundo

Un subcampo del ML que utiliza múltiples capas de redes neuronales artificiales para identificar el mapeo entre los datos de entrada y las variables objetivo de interés.

defensa en profundidad

Un enfoque de seguridad de la información en el que se distribuyen cuidadosamente una serie de mecanismos y controles de seguridad en una red informática para proteger la confidencialidad, la integridad y la disponibilidad de la red y de los datos que contiene. Al adoptar esta estrategia AWS, se añaden varios controles en diferentes capas de la AWS Organizations estructura para ayudar a proteger los recursos. Por ejemplo, un enfoque de defensa en profundidad podría combinar la autenticación multifactor, la segmentación de la red y el cifrado.

administrador delegado

En AWS Organizations, un servicio compatible puede registrar una cuenta de AWS miembro para administrar las cuentas de la organización y gestionar los permisos de ese servicio. Esta cuenta se denomina administrador delegado para ese servicio. Para obtener más información y una lista de servicios compatibles, consulte [Servicios que funcionan con AWS Organizations](#) en la documentación de AWS Organizations .

Implementación

El proceso de hacer que una aplicación, características nuevas o correcciones de código se encuentren disponibles en el entorno de destino. La implementación abarca implementar cambios en una base de código y, a continuación, crear y ejecutar esa base en los entornos de la aplicación.

entorno de desarrollo

Consulte [entorno](#).

control de detección

Un control de seguridad que se ha diseñado para detectar, registrar y alertar después de que se produzca un evento. Estos controles son una segunda línea de defensa, ya que lo advierten sobre los eventos de seguridad que han eludido los controles preventivos establecidos. Para

obtener más información, consulte [Controles de detección](#) en Implementación de controles de seguridad en AWS.

asignación de flujos de valor para el desarrollo (DVSM)

Proceso que se utiliza para identificar y priorizar las restricciones que afectan negativamente a la velocidad y la calidad en el ciclo de vida del desarrollo de software. DVSM amplía el proceso de asignación del flujo de valor diseñado originalmente para las prácticas de fabricación ajustada. Se centra en los pasos y los equipos necesarios para crear y transferir valor a través del proceso de desarrollo de software.

gemelo digital

Representación virtual de un sistema del mundo real, como un edificio, una fábrica, un equipo industrial o una línea de producción. Los gemelos digitales son compatibles con el mantenimiento predictivo, la supervisión remota y la optimización de la producción.

tabla de dimensiones

En un [esquema en estrella](#), tabla más pequeña que contiene los atributos de datos sobre los datos cuantitativos en una tabla de hechos. Los atributos de la tabla de dimensiones suelen ser campos de texto o números discretos que se comportan como texto. Estos atributos se suelen utilizar para restringir consultas, filtrarlas y etiquetar los conjuntos de resultados.

desastre

Un evento que impide que una carga de trabajo o un sistema cumplan sus objetivos empresariales en su ubicación principal de implementación. Estos eventos pueden ser desastres naturales, fallos técnicos o el resultado de acciones humanas, como una configuración incorrecta involuntaria o un ataque de malware.

recuperación de desastres (DR)

Estrategia y proceso que utiliza para minimizar el tiempo de inactividad y la pérdida de datos a causa de un [desastre](#). Para obtener más información, consulte [Recuperación de cargas de trabajo ante desastres en AWS: Recuperación en la nube](#) en el AWS Well-Architected marco.

DML

Consulte [lenguaje de manipulación de bases de datos](#).

diseño basado en el dominio

Un enfoque para desarrollar un sistema de software complejo mediante la conexión de sus componentes a dominios en evolución, o a los objetivos empresariales principales, a los que

sirve cada componente. Eric Evans introdujo este concepto en su libro *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003). Para obtener información sobre cómo utilizar el diseño basado en dominios con el patrón de higos estranguladores, consulte [Modernización gradual de los servicios web antiguos de ASP.NET Microsoft \(ASMX\) mediante contenedores y Amazon API Gateway](#).

DR

Consulte [recuperación ante desastres](#).

Detección de desviaciones

Seguimiento de las desviaciones con respecto a una configuración con línea de base. Por ejemplo, puedes usarlo AWS CloudFormation para [detectar desviaciones en los recursos del sistema](#) o puedes usarlo AWS Control Tower para [detectar cambios en tu landing zone](#) que puedan afectar al cumplimiento de los requisitos de gobierno.

DVSM

Consulte [asignación de flujos de valor para el desarrollo](#).

E

EDA

Consulte [análisis de datos de tipo exploratorio](#).

EDI

Consulte [intercambio electrónico de datos](#).

computación en la periferia

La tecnología que aumenta la potencia de cálculo de los dispositivos inteligentes en la periferia de una red de IoT. En comparación con la [computación en la nube](#), la computación de periferia puede reducir la latencia de la comunicación y mejorar el tiempo de respuesta.

intercambio electrónico de datos (EDI)

Intercambio automatizado de documentos comerciales entre organizaciones. Para más información, consulte [¿Qué es el intercambio electrónico de datos?](#)

cifrado

Proceso de computación que transforma datos de texto plano, que son legibles por humanos, en texto cifrado.

clave de cifrado

Cadena criptográfica de bits aleatorios que se genera mediante un algoritmo de cifrado. Las claves pueden variar en longitud y cada una se ha diseñado para ser impredecible y única.

endianidad

El orden en el que se almacenan los bytes en la memoria del ordenador. Big-endian los sistemas almacenan primero el byte más significativo. Little-endian los sistemas almacenan primero el byte menos significativo.

punto de conexión

Consulte [punto de conexión de servicio](#).

servicio de punto de conexión

Servicio que puede alojar en una nube privada virtual (VPC) para compartir con otros usuarios. Puede crear un servicio de punto final con AWS PrivateLink entidades principales Cuentas de AWS o AWS Identity and Access Management (de IAM) y conceder permisos a ellas. Estas cuentas o entidades principales pueden conectarse a su servicio de punto de conexión de forma privada mediante la creación de puntos de conexión de VPC de interfaz. Para obtener más información, consulte [Creación de un servicio de punto de conexión](#) en la documentación de Amazon Virtual Private Cloud (Amazon VPC).

planificación de recursos empresariales (ERP)

Sistema que automatiza y administra los procesos empresariales clave (como la contabilidad, [MES](#) y la administración de proyectos) de una empresa.

cifrado de sobre

El proceso de cifrar una clave de cifrado con otra clave de cifrado. Para obtener más información, consulte el [cifrado de sobres](#) en la documentación de AWS Key Management Service (AWS KMS).

entorno

Una instancia de una aplicación en ejecución. Los siguientes son los tipos de entornos más comunes en la computación en la nube:

- entorno de desarrollo: instancia de una aplicación en ejecución que solo se encuentra disponible para el equipo principal responsable del mantenimiento de la aplicación. Los entornos de desarrollo se utilizan para probar los cambios antes de promocionarlos a los entornos superiores. Este tipo de entorno a veces se denomina entorno de prueba.
- entornos inferiores: todos los entornos de desarrollo de una aplicación, como los que se utilizan para las compilaciones y pruebas iniciales.
- entorno de producción: instancia de una aplicación en ejecución a la que pueden acceder los usuarios finales. En un CI/CD proceso, el entorno de producción es el último entorno de implementación.
- entornos superiores: todos los entornos a los que pueden acceder usuarios que no sean del equipo de desarrollo principal. Esto puede incluir un entorno de producción, entornos de preproducción y entornos para las pruebas de aceptación por parte de los usuarios.

epopeya

En las metodologías ágiles, son categorías funcionales que ayudan a organizar y priorizar el trabajo. Las epopeyas brindan una descripción detallada de los requisitos y las tareas de implementación. Por ejemplo, las epopeyas AWS de seguridad de CAF incluyen la gestión de identidades y accesos, los controles de detección, la seguridad de la infraestructura, la protección de datos y la respuesta a incidentes. Para obtener más información sobre las epopeyas en la estrategia de migración de AWS , consulte la [Guía de implementación del programa](#).

ERP

Consulte [planificación de recursos empresariales](#).

análisis de datos de tipo exploratorio (EDA)

El proceso de analizar un conjunto de datos para comprender sus características principales. Se recopilan o agregan datos y, a continuación, se realizan las investigaciones iniciales para encontrar patrones, detectar anomalías y comprobar las suposiciones. El EDA se realiza mediante el cálculo de estadísticas resumidas y la creación de visualizaciones de datos.

F

tabla de hechos

Tabla central de un [esquema en estrella](#). Almacena datos cuantitativos sobre operaciones empresariales. Por lo general, una tabla de hechos contiene dos tipos de columnas: las que contienen medidas y las que contienen una clave externa para una tabla de dimensiones.

Fail Fast

Filosofía que utiliza pruebas frecuentes e incrementales para reducir el ciclo de vida del desarrollo. Es una parte fundamental de los enfoques ágiles.

límite de aislamiento de errores

En el Nube de AWS, un límite, como una zona de disponibilidad Región de AWS, un plano de control o un plano de datos, que limita el efecto de una falla y ayuda a mejorar la resiliencia de las cargas de trabajo. Para más información, consulte [AWS Fault Isolation Boundaries](#).

rama de característica

Consulte [rama](#).

características

Los datos de entrada que se utilizan para hacer una predicción. Por ejemplo, en un contexto de fabricación, las características pueden ser imágenes que se capturan periódicamente desde la línea de fabricación.

importancia de las características

La importancia que tiene una característica para las predicciones de un modelo. Por lo general, esto se expresa como una puntuación numérica que se puede calcular mediante diversas técnicas, como las explicaciones aditivas de Shapley (SHAP) y los gradientes integrados. Para obtener más información, consulte [Interpretabilidad del modelo de aprendizaje automático](#) con AWS

transformación de funciones

Optimizar los datos para el proceso de ML, lo que incluye enriquecer los datos con fuentes adicionales, escalar los valores o extraer varios conjuntos de información de un solo campo de datos. Esto permite que el modelo de ML se beneficie de los datos. Por ejemplo, si divide la fecha del “27 de mayo de 2021 00:15:37” en “jueves”, “mayo”, “2021” y “15”, puede ayudar al algoritmo

de aprendizaje a aprender patrones matizados asociados a los diferentes componentes de los datos.

peticiones con pocos pasos

Proporcionar a un [LLM](#) una pequeña cantidad de ejemplos que demuestren la tarea y el resultado deseado antes de pedirle que lleve a cabo una tarea similar. Esta técnica es una aplicación del aprendizaje contextual, en el que los modelos aprenden a partir de ejemplos (tomas) integrados en las instrucciones. Few-shot Las indicaciones pueden ser eficaces para tareas que requieren un formato, un razonamiento o un conocimiento del dominio específicos. Consulte también [peticiones desde cero](#).

FGAC

Consulte [control de acceso detallado](#).

control de acceso preciso (FGAC)

El uso de varias condiciones que tienen por objetivo permitir o denegar una solicitud de acceso.

migración relámpago

Método de migración de bases de datos que utiliza la replicación continua de datos mediante la [captura de datos de cambio](#) para migrar los datos en el menor tiempo posible, en lugar de utilizar un enfoque gradual. El objetivo es reducir al mínimo el tiempo de inactividad.

FM

Consulte [modelo fundacional](#).

Modelo fundacional (FM)

Gran red neuronal de aprendizaje profundo que se entrenó con conjuntos de datos masivos de datos generalizados y no etiquetados. Los FM pueden hacer una amplia variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural. Para más información, consulte [¿Qué son los modelos fundacionales?](#)

Puerta de enlace FM

Un intermediario centralizado que controla y normaliza el acceso a los modelos básicos. También se conoce como puerta de enlace LLM.

G

IA generativa

Subconjunto de modelos de [IA](#) que se entrenaron con grandes cantidades de datos y que pueden utilizar una simple petición de texto para crear contenido y artefactos nuevos, como imágenes, videos, texto y audio. Para más información, consulte [¿Qué es la IA generativa?](#)

bloqueo geográfico

Consulte [restricciones geográficas](#).

restricciones geográficas (bloqueo geográfico)

En Amazon CloudFront, una opción para impedir que los usuarios de países específicos accedan a las distribuciones de contenido. Puede utilizar una lista de permitidos o bloqueados para especificar los países aprobados y prohibidos. Para obtener más información, consulta [Restringir la distribución geográfica del contenido](#) en la CloudFront documentación.

Flujo de trabajo de Gitflow

Un enfoque en el que los entornos inferiores y superiores utilizan diferentes ramas en un repositorio de código fuente. El flujo de trabajo de Gitflow se considera heredado, mientras que el [flujo de trabajo basado en enlaces troncales](#) es el enfoque moderno preferido.

imagen dorada

Instantánea de un sistema o software que se usa como plantilla para implementar nuevas instancias de ese sistema o software. Por ejemplo, en la fabricación, una imagen dorada se puede utilizar para aprovisionar software en varios dispositivos y ayuda a mejorar la velocidad, la escalabilidad y la productividad de las operaciones de fabricación de dispositivos.

estrategia de implementación desde cero

La ausencia de infraestructura existente en un entorno nuevo. Al adoptar una estrategia de implementación desde cero para una arquitectura de sistemas, puede seleccionar todas las tecnologías nuevas sin que estas deban ser compatibles con una infraestructura existente, lo que también se conoce como [implementación sobre infraestructura existente](#). Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

barrera de protección

Una regla de alto nivel que ayuda a regular los recursos, las políticas y la conformidad en todas las unidades organizativas (OU). Las barreras de protección preventivas aplican políticas para garantizar la alineación con los estándares de conformidad. Se implementan mediante políticas de control de servicios y límites de permisos de IAM. Las barreras de protección de detección detectan las vulneraciones de las políticas y los problemas de conformidad, y generan alertas para su corrección. Se implementan mediante Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

barandas (AI)

Mecanismos de seguridad que filtran, validan y restringen las entradas y salidas de los [agentes](#) para ayudar a garantizar un comportamiento responsable y seguro de la IA.

H

HA

Consulte [alta disponibilidad](#).

migración heterogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que utilice un motor de base de datos diferente (por ejemplo, de Oracle a Amazon Aurora). La migración heterogénea suele ser parte de un esfuerzo de rediseño de la arquitectura y convertir el esquema puede ser una tarea compleja. [AWS ofrece AWS SCT](#), lo cual ayuda con las conversiones de esquemas.

alta disponibilidad (HA)

La capacidad de una carga de trabajo para funcionar de forma continua, sin intervención, en caso de desafíos o desastres. Los sistemas de alta disponibilidad están diseñados para realizar una conmutación por error automática, ofrecer un rendimiento de alta calidad de forma constante y gestionar diferentes cargas y fallos con un impacto mínimo en el rendimiento.

modernización histórica

Un enfoque utilizado para modernizar y actualizar los sistemas de tecnología operativa (TO) a fin de satisfacer mejor las necesidades de la industria manufacturera. Un histórico es un tipo de base de datos que se utiliza para recopilar y almacenar datos de diversas fuentes en una fábrica.

datos de reserva

Parte de los datos históricos etiquetados que se ocultan de un conjunto de datos que se utiliza para entrenar un modelo de [machine learning](#). Puede utilizar los datos de reserva para evaluar el rendimiento del modelo mediante la comparación de las predicciones del modelo con los datos de reserva.

human-in-the-loop (HiTL)

Un patrón de flujo de trabajo en el que la ejecución de los [agentes](#) se detiene para su revisión y aprobación por parte de una persona en los puntos de decisión críticos.

migración homogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que comparte el mismo motor de base de datos (por ejemplo, Microsoft SQL Server a Amazon RDS para SQL Server). La migración homogénea suele formar parte de un esfuerzo para volver a alojar o redefinir la plataforma. Puede utilizar las utilidades de bases de datos nativas para migrar el esquema.

datos recientes

Datos a los que se accede con frecuencia, como datos en tiempo real o datos traslacionales recientes. Por lo general, estos datos requieren un nivel o una clase de almacenamiento de alto rendimiento para proporcionar respuestas rápidas a las consultas.

hotfix

Una solución urgente para un problema crítico en un entorno de producción. Debido a su urgencia, una revisión suele realizarse fuera del flujo de trabajo habitual de las DevOps versiones.

periodo de hiperatención

Periodo, inmediatamente después de la transición, durante el cual un equipo de migración administra y monitorea las aplicaciones migradas en la nube para solucionar cualquier problema. Por lo general, este periodo dura de 1 a 4 días. Al final del periodo de hiperatención, el equipo de migración suele transferir la responsabilidad de las aplicaciones al equipo de operaciones en la nube.

|

IaC

Consulte [infraestructura como código](#).

políticas basadas en identidades

Política asociada a uno o más directores de IAM que define sus permisos en el entorno. Nube de AWS

aplicación inactiva

Aplicación que utiliza un promedio de CPU y memoria de entre 5 y 20 por ciento durante un periodo de 90 días. En un proyecto de migración, es habitual retirar estas aplicaciones o mantenerlas en las instalaciones.

IIoT

Consulte [Internet de las cosas industrial](#).

infraestructura inmutable

Modelo que implementa una nueva infraestructura para las cargas de trabajo de producción en lugar de actualizar o modificar la infraestructura existente o aplicarle revisiones. Las infraestructuras inmutables son de manera intrínseca más coherentes, fiables y predecibles que las [infraestructuras mutables](#). Para obtener más información, consulte las mejores prácticas del [Framework para implementar con una infraestructura inmutable](#). AWS Well-Architected

VPC entrante (de entrada)

En una arquitectura de AWS cuentas múltiples, una VPC que acepta, inspecciona y enruta las conexiones de red desde fuera de una aplicación. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

migración gradual

Estrategia de transición en la que se migra la aplicación en partes pequeñas en lugar de realizar una transición única y completa. Por ejemplo, puede trasladar inicialmente solo unos pocos microservicios o usuarios al nuevo sistema. Tras comprobar que todo funciona correctamente, puede trasladar microservicios o usuarios adicionales de forma gradual hasta que pueda retirar su sistema heredado. Esta estrategia reduce los riesgos asociados a las grandes migraciones.

|

Industria 4.0

Un término que [Klaus Schwab](#) introdujo en 2016 para referirse a la modernización de los procesos de fabricación mediante avances en la conectividad, los datos en tiempo real, la automatización, el análisis y. AI/ML

infraestructura

Todos los recursos y activos que se encuentran en el entorno de una aplicación.

infraestructura como código (IaC)

Proceso de aprovisionamiento y administración de la infraestructura de una aplicación mediante un conjunto de archivos de configuración. La IaC se ha diseñado para ayudarlo a centralizar la administración de la infraestructura, estandarizar los recursos y escalar con rapidez a fin de que los entornos nuevos sean repetibles, fiables y consistentes.

Internet de las cosas industrial (IIoT)

El uso de sensores y dispositivos conectados a Internet en los sectores industriales, como el productivo, el eléctrico, el automotriz, el sanitario, el de las ciencias de la vida y el de la agricultura. Para obtener más información, consulte [Creación de una estrategia de transformación digital del Internet de las cosas industrial \(IIoT\)](#).

VPC de inspección

En una arquitectura de AWS cuentas múltiples, una VPC centralizada que gestiona las inspecciones del tráfico de red entre las VPC (iguales o Regiones de AWS diferentes), Internet y las redes locales. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

Internet de las cosas (IoT)

Red de objetos físicos conectados con sensores o procesadores integrados que se comunican con otros dispositivos y sistemas a través de Internet o de una red de comunicación local. Para obtener más información, consulte [¿Qué es IoT?](#).

interpretabilidad

Característica de un modelo de machine learning que describe el grado en que un ser humano puede entender cómo las predicciones del modelo dependen de sus entradas. Para obtener más información, consulte Interpretabilidad del modelo [de aprendizaje automático](#) con. AWS

IoT

Consulte [Internet de las cosas](#).

biblioteca de información de TI (ITIL)

Conjunto de prácticas recomendadas para ofrecer servicios de TI y alinearlos con los requisitos empresariales. La ITIL proporciona la base para la ITSM.

administración de servicios de TI (ITSM)

Actividades asociadas con el diseño, la implementación, la administración y el soporte de los servicios de TI para una organización. Para obtener información sobre la integración de las operaciones en la nube con las herramientas de ITSM, consulte la [Guía de integración de operaciones](#).

ITIL

Consulte [biblioteca de información de TI](#).

ITSM

Consulte [administración de servicios de TI](#).

L

control de acceso basado en etiquetas (LBAC)

Una implementación del control de acceso obligatorio (MAC) en la que a los usuarios y a los propios datos se les asigna explícitamente un valor de etiqueta de seguridad. La intersección entre la etiqueta de seguridad del usuario y la etiqueta de seguridad de los datos determina qué filas y columnas puede ver el usuario.

zona de aterrizaje

Una landing zone es un AWS entorno multicuenta bien diseñado, escalable y seguro. Este es un punto de partida desde el cual las empresas pueden lanzar e implementar rápidamente cargas de trabajo y aplicaciones con confianza en su entorno de seguridad e infraestructura. Para obtener más información sobre las zonas de aterrizaje, consulte [Configuración de un entorno de AWS seguro y escalable con varias cuentas](#).

modelo de lenguaje de gran tamaño (LLM)

Modelo de [IA](#) de aprendizaje profundo que se entrenó previamente con una gran cantidad de datos. Un LLM puede llevar a cabo varias tareas, como responder preguntas, resumir documentos, traducir textos a otros idiomas y completar oraciones. Para más información, consulte [¿Qué es un LLM \(modelo de lenguaje de gran tamaño\)?](#)

migración grande

Migración de 300 servidores o más.

LBAC

Consulte [control de acceso basado en etiquetas](#).

privilegio mínimo

La práctica recomendada de seguridad que consiste en conceder los permisos mínimos necesarios para realizar una tarea. Para obtener más información, consulte [Aplicar permisos de privilegio mínimo](#) en la documentación de IAM.

migrar mediante lift-and-shift

Consulte [Las 7 R](#).

sistema little-endian

Un sistema que almacena primero el byte menos significativo. Consulte también [endianidad](#).

LLM

Consulte [modelo de lenguaje de gran tamaño](#).

entornos inferiores

Consulte [entorno](#).

M

machine learning (ML)

Un tipo de inteligencia artificial que utiliza algoritmos y técnicas para el reconocimiento y el aprendizaje de patrones. El ML analiza y aprende de los datos registrados, como los datos del

Internet de las cosas (IoT), para generar un modelo estadístico basado en patrones. Para más información, consulte [Machine learning](#).

rama principal

Consulte [rama](#).

malware

Software diseñado para comprometer la seguridad o la privacidad de la computadora. El malware podría interrumpir los sistemas informáticos, filtrar información confidencial u obtener acceso no autorizado. Algunos ejemplos de malware son los virus, los gusanos, el ransomware, los troyanos, el spyware y los registradores de pulsaciones de teclas.

Servicios administrados

Servicios de AWS en el que AWS opera la capa de infraestructura, el sistema operativo y las plataformas, y se accede a los puntos finales para almacenar y recuperar datos. Amazon Simple Storage Service (Amazon S3) y Amazon DynamoDB son ejemplos de servicios administrados. También se conocen como servicios abstractos.

sistema de ejecución de fabricación (MES)

Sistema de software para seguir, supervisar, documentar y controlar los procesos de producción que convierten las materias primas en productos acabados en la zona de producción.

MAP

Consulte [Programa de aceleración de la migración](#).

MCP

Consulte [Model Context Protocol](#).

Protocolo de contexto para modelos (MCP)

Un protocolo sin estado para la comunicación entre el [agente](#) y la [herramienta](#).

Servidor MCP

Un servicio que expone una o más [herramientas](#) a través del protocolo [Model Context](#).

mecanismo

Proceso completo mediante el que se crea una herramienta, se impulsa su adopción y, a continuación, se inspeccionan los resultados para hacer ajustes. Un mecanismo es un ciclo

que se refuerza y mejora por sí mismo a medida que funciona. Para obtener más información, consulte [Creación de mecanismos](#) en el AWS Well-Architected marco.

cuenta de miembro

Todas las Cuentas de AWS demás cuentas, excepto la de administración, que forman parte de una organización AWS Organizations. Una cuenta no puede pertenecer a más de una organización a la vez.

MES

Consulte [sistema de ejecución de fabricación](#).

Message Queuing Telemetry Transport (MQTT)

[Un protocolo de comunicación ligero de máquina a máquina \(M2M\), basado en el publish/subscribe patrón, para dispositivos de IoT con recursos limitados.](#)

microservicio

Un servicio pequeño e independiente que se comunica a través de API bien definidas y que, por lo general, es propiedad de equipos pequeños e independientes. Por ejemplo, un sistema de seguros puede incluir microservicios que se adapten a las capacidades empresariales, como las de ventas o marketing, o a subdominios, como las de compras, reclamaciones o análisis. Los beneficios de los microservicios incluyen la agilidad, la escalabilidad flexible, la facilidad de implementación, el código reutilizable y la resiliencia. Para obtener más información, consulte [Integrar](#) microservicios mediante servicios sin servidor. AWS

arquitectura de microservicios

Un enfoque para crear una aplicación con componentes independientes que ejecutan cada proceso de la aplicación como un microservicio. Estos microservicios se comunican a través de una interfaz bien definida mediante API ligeras. Cada microservicio de esta arquitectura se puede actualizar, implementar y escalar para satisfacer la demanda de funciones específicas de una aplicación. Para obtener más información, consulte [Implementación de microservicios](#) en. AWS

Programa de aceleración de la migración (MAP)

Un AWS programa que proporciona soporte de consultoría, formación y servicios para ayudar a las organizaciones a crear una base operativa sólida para migrar a la nube y para ayudar a compensar el costo inicial de las migraciones. El MAP incluye una metodología de migración para ejecutar las migraciones antiguas de forma metódica y un conjunto de herramientas para automatizar y acelerar los escenarios de migración más comunes.

migración a escala

Proceso de transferencia de la mayoría de la cartera de aplicaciones a la nube en oleadas, con más aplicaciones desplazadas a un ritmo más rápido en cada oleada. En esta fase, se utilizan las prácticas recomendadas y las lecciones aprendidas en las fases anteriores para implementar una fábrica de migración de equipos, herramientas y procesos con el fin de agilizar la migración de las cargas de trabajo mediante la automatización y la entrega ágil. Esta es la tercera fase de la [estrategia de migración de AWS](#).

fábrica de migración

Cross-functional equipos que agilizan la migración de las cargas de trabajo mediante enfoques ágiles y automatizados. Los equipos de las fábricas de migración suelen estar compuestos por analistas y propietarios de operaciones, ingenieros de migración, desarrolladores y DevOps profesionales que trabajan a pasos agigantados. Entre el 20 y el 50 por ciento de la cartera de aplicaciones empresariales se compone de patrones repetidos que pueden optimizarse mediante un enfoque de fábrica. Para obtener más información, consulte la [discusión sobre las fábricas de migración](#) y la [Guía de fábricas de migración a la nube](#) en este contenido.

metadatos de migración

Información sobre la aplicación y el servidor que se necesita para completar la migración. Cada patrón de migración requiere un conjunto diferente de metadatos de migración. Algunos ejemplos de metadatos de migración son la subred de destino, el grupo de seguridad y AWS la cuenta.

patrón de migración

Tarea de migración repetible que detalla la estrategia de migración, el destino de la migración y la aplicación o el servicio de migración utilizados. Ejemplo: rehospede la migración a Amazon EC2 AWS con Application Migration Service.

Migration Portfolio Assessment (MPA)

Herramienta en línea que proporciona información a fin de validar los argumentos comerciales necesarios para migrar a la Nube de AWS. La MPA ofrece una evaluación detallada de la cartera (adecuación del tamaño de los servidores, precios, comparaciones del costo total de propiedad, análisis de los costos de migración), así como una planificación de la migración (análisis y recopilación de datos de aplicaciones, agrupación de aplicaciones, priorización de la migración y planificación de oleadas). La [herramienta MPA](#) (requiere iniciar sesión) está disponible de forma gratuita para todos los AWS consultores y consultores de los socios de APN.

Evaluación de la preparación para la migración (MRA)

Proceso que consiste en obtener información sobre el estado de preparación de una organización para la nube, identificar sus puntos fuertes y débiles y elaborar un plan de acción para cerrar las brechas identificadas mediante el AWS CAF. Para obtener más información, consulte la [Guía de preparación para la migración](#). La MRA es la primera fase de la [estrategia de migración de AWS](#).

estrategia de migración

Enfoque utilizado para migrar una carga de trabajo a la Nube de AWS. Para más información, consulte la entrada [Las 7 R](#) de este glosario y también [Mobilize your organization to accelerate large-scale migrations](#).

ML

Consulte [machine learning](#).

modernización

Transformar una aplicación obsoleta (antigua o monolítica) y su infraestructura en un sistema ágil, elástico y de alta disponibilidad en la nube para reducir los gastos, aumentar la eficiencia y aprovechar las innovaciones. Para más información, consulte [Strategy for modernizing applications in the Nube de AWS](#).

evaluación de la preparación para la modernización

Evaluación que ayuda a determinar la preparación para la modernización de las aplicaciones de una organización; identifica los beneficios, los riesgos y las dependencias; y determina qué tan bien la organización puede soportar el estado futuro de esas aplicaciones. El resultado de la evaluación es un esquema de la arquitectura objetivo, una hoja de ruta que detalla las fases de desarrollo y los hitos del proceso de modernización y un plan de acción para abordar las brechas identificadas. Para más información, consulte [Evaluating modernization readiness for applications in the Nube de AWS](#).

aplicaciones monolíticas (monolitos)

Aplicaciones que se ejecutan como un único servicio con procesos estrechamente acoplados. Las aplicaciones monolíticas presentan varios inconvenientes. Si una característica de la aplicación experimenta un aumento en la demanda, se debe escalar toda la arquitectura. Agregar o mejorar las características de una aplicación monolítica también se vuelve más complejo a medida que crece la base de código. Para solucionar problemas con la aplicación, puede utilizar

una arquitectura de microservicios. Para obtener más información, consulte [Descomposición de monolitos en microservicios](#).

MPA

Consulte [Migration Portfolio Assessment](#).

MQTT

Consulte [Message Queuing Telemetry Transport](#).

clasificación multiclase

Un proceso que ayuda a generar predicciones para varias clases (predice uno de más de dos resultados). Por ejemplo, un modelo de ML podría preguntar “¿Este producto es un libro, un automóvil o un teléfono?” o “¿Qué categoría de productos es más interesante para este cliente?”.

infraestructura mutable

Modelo que actualiza y modifica la infraestructura actual para las cargas de trabajo de producción. Para mejorar la coherencia, la confiabilidad y la previsibilidad, el AWS Well-Architected Marco recomienda el uso de una [infraestructura inmutable](#) como práctica recomendada.

O

OAC

Consulte [control de acceso de origen](#).

OAI

Consulte [identidad de acceso de origen](#).

OCM

Consulte [administración del cambio organizacional](#).

migración fuera de línea

Método de migración en el que la carga de trabajo de origen se elimina durante el proceso de migración. Este método implica un tiempo de inactividad prolongado y, por lo general, se utiliza para cargas de trabajo pequeñas y no críticas.

OI

Consulte [integración de operaciones](#).

OLA

Consulte [acuerdo de nivel operativo](#).

migración en línea

Método de migración en el que la carga de trabajo de origen se copia al sistema de destino sin que se desconecte. Las aplicaciones que están conectadas a la carga de trabajo pueden seguir funcionando durante la migración. Este método implica un tiempo de inactividad nulo o mínimo y, por lo general, se utiliza para cargas de trabajo de producción críticas.

OPC-UA

Consulte [Open Process Communications: arquitectura unificada](#).

Comunicaciones de proceso abierto: arquitectura unificada () OPC-UA

Un protocolo de comunicación de máquina a máquina (M2M) para la automatización industrial. OPC-UA proporciona un estándar de interoperabilidad con esquemas de cifrado, autenticación y autorización de datos.

acuerdo de nivel operativo (OLA)

Acuerdo que aclara lo que los grupos de TI operativos se comprometen a ofrecerse entre sí, para respaldar un acuerdo de nivel de servicio (SLA).

revisión de la preparación operativa (ORR)

Lista de comprobación de preguntas y prácticas recomendadas asociadas que son útiles para comprender, evaluar, prevenir o reducir el alcance de los incidentes y posibles errores. Para obtener más información, consulte [las revisiones de preparación operativa \(ORR\)](#) en el AWS Well-Architected marco.

tecnología operativa (TO)

Sistemas de hardware y software que funcionan con el entorno físico para controlar las operaciones, los equipos y la infraestructura industriales. En el sector de la fabricación, la integración de los sistemas de TO y tecnología de la información (TI) es un enfoque clave para las transformaciones de la [industria 4.0](#).

integración de operaciones (OI)

Proceso de modernización de las operaciones en la nube, que implica la planificación de la preparación, la automatización y la integración. Para obtener más información, consulte la [Guía de integración de las operaciones](#).

registro de seguimiento organizativo

Un registro creado por y AWS CloudTrail que registra todos los eventos Cuentas de AWS de una organización AWS Organizations. Este registro de seguimiento se crea en cada Cuenta de AWS que forma parte de la organización y realiza un seguimiento de la actividad en cada cuenta. Para obtener más información, consulte [Crear un registro para una organización](#) en la CloudTrail documentación.

administración del cambio organizacional (OCM)

Marco para administrar las transformaciones empresariales importantes y disruptivas desde la perspectiva de las personas, la cultura y el liderazgo. La OCM ayuda a las empresas a prepararse para nuevos sistemas y estrategias y a realizar la transición a ellos, al acelerar la adopción de cambios, abordar los problemas de transición e impulsar cambios culturales y organizacionales. En la estrategia de AWS migración, este marco se denomina aceleración de personal, debido a la velocidad de cambio que requieren los proyectos de adopción de la nube. Para obtener más información, consulte la [Guía de OCM](#).

control de acceso de origen (OAC)

En CloudFront, una opción mejorada para restringir el acceso y proteger el contenido del Amazon Simple Storage Service (Amazon S3). El OAC admite todos los buckets de S3 Regiones de AWS, el cifrado del lado del servidor con AWS KMS (SSE-KMS) y DELETE las solicitudes PUT y dinámicas al bucket de S3.

identidad de acceso de origen (OAI)

En CloudFront, una opción para restringir el acceso y proteger el contenido de Amazon S3. Cuando utiliza OAI, CloudFront crea un principal con el que Amazon S3 puede autenticarse. Los directores autenticados solo pueden acceder al contenido de un bucket de S3 a través de una distribución específica. CloudFront Consulte también el [OAC](#), que proporciona un control de acceso más detallado y mejorado.

ORR

Consulte [revisión de la preparación operativa](#).

OT

Consulte [tecnología operativa](#).

VPC saliente (de salida)

En una arquitectura de AWS cuentas múltiples, una VPC que gestiona las conexiones de red que se inician desde una aplicación. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

P

límite de permisos

Una política de administración de IAM que se adjunta a las entidades principales de IAM para establecer los permisos máximos que puede tener el usuario o el rol. Para obtener más información, consulte [Límites de permisos](#) en la documentación de IAM.

información de identificación personal (PII)

Información que, vista directamente o combinada con otros datos relacionados, puede utilizarse para deducir de manera razonable la identidad de una persona. Algunos ejemplos de información de identificación personal son los nombres, las direcciones y la información de contacto.

PII

Consulte [información de identificación personal](#).

manual de estrategias

Conjunto de pasos predefinidos que capturan el trabajo asociado a las migraciones, como la entrega de las funciones de operaciones principales en la nube. Un manual puede adoptar la forma de scripts, manuales de procedimientos automatizados o resúmenes de los procesos o pasos necesarios para operar un entorno modernizado.

PLC

Consulte [controlador lógico programable](#).

PLM

Consulte [administración del ciclo de vida del producto](#).

policy

Objeto que puede definir permisos (consulte [política basada en identidad](#)), especificar las condiciones de acceso (consulte [política basada en recursos](#)) o definir los permisos máximos para todas las cuentas de una organización de AWS Organizations (consulte [política de control de servicio](#)).

persistencia políglota

Elegir de forma independiente la tecnología de almacenamiento de datos de un microservicio en función de los patrones de acceso a los datos y otros requisitos. Si sus microservicios tienen la misma tecnología de almacenamiento de datos, pueden enfrentarse a desafíos de implementación o experimentar un rendimiento deficiente. Los microservicios se implementan más fácilmente y logran un mejor rendimiento y escalabilidad si utilizan el almacén de datos que mejor se adapte a sus necesidades.

evaluación de cartera

Proceso de detección, análisis y priorización de la cartera de aplicaciones para planificar la migración. Para obtener más información, consulte la [Evaluación de la preparación para la migración](#).

predicate

Condición de consulta que devuelve true o false. En general, se encuentra en una cláusula WHERE.

inserción de predicados

Técnica de optimización de consultas en bases de datos que filtra los datos de la consulta antes de transferirlos. Esta técnica reduce la cantidad de datos de la base de datos relacional que se tienen que recuperar y procesar. Además, mejora el rendimiento de las consultas.

control preventivo

Un control de seguridad diseñado para evitar que ocurra un evento. Estos controles son la primera línea de defensa para evitar el acceso no autorizado o los cambios no deseados en la red. Para obtener más información, consulte [Controles preventivos](#) en Implementación de controles de seguridad en AWS.

entidad principal

Una entidad AWS que puede realizar acciones y acceder a los recursos. Esta entidad suele ser un usuario raíz para un Cuenta de AWS rol de IAM o un usuario. Para obtener más información, consulte Entidad principal en [Términos y conceptos de roles](#) en la documentación de IAM.

Privacidad desde el diseño

Enfoque de ingeniería de sistemas que tiene en cuenta la privacidad durante todo el proceso de desarrollo.

zonas alojadas privadas

Contenedor que aloja información acerca de cómo desea que responda Amazon Route 53 a las consultas de DNS de un dominio y sus subdominios en una o varias VPC. Para obtener más información, consulte [Uso de zonas alojadas privadas](#) en la documentación de Route 53.

control proactivo

[Control de seguridad](#) que se diseñó para evitar la implementación de recursos que no cumplan con la normativa. Estos controles analizan los recursos antes de aprovisionarlos. Si el recurso no cumple con los requisitos del control, no se aprovisiona. Para obtener más información, consulte la [guía de referencia de controles](#) en la AWS Control Tower documentación y consulte [Controles proactivos](#) en Implementación de controles de seguridad en AWS.

administración del ciclo de vida del producto (PLM)

Administración de los datos y los procesos de un producto a lo largo de todo su ciclo de vida, desde el diseño, el desarrollo y el lanzamiento, pasando por el crecimiento y la madurez, hasta la reducción de su uso y su retirada.

entorno de producción

Consulte [entorno](#).

controlador lógico programable (PLC)

En el sector de la fabricación, computadora adaptable y altamente fiable que supervisa las máquinas y automatiza los procesos de fabricación.

encadenamiento de peticiones

Uso de la salida de una petición de [LLM](#) como entrada para la siguiente petición a fin de generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en tareas secundarias o para refinar o ampliar de forma iterativa una respuesta preliminar. Ayuda a mejorar la precisión

y la relevancia de las respuestas de un modelo y permite obtener resultados más detallados y personalizados.

seudonimización

El proceso de reemplazar los identificadores personales de un conjunto de datos por valores de marcadores de posición. La seudonimización puede ayudar a proteger la privacidad personal. Los datos seudonimizados siguen considerándose datos personales.

publish/subscribe (pub/sub)

Patrón que permite establecer comunicaciones asíncronas entre microservicios para mejorar la escalabilidad y la capacidad de respuesta. Por ejemplo, en un [MES](#) basado en microservicios, un microservicio puede publicar mensajes de eventos en un canal al que se pueden suscribir otros microservicios. El sistema puede agregar nuevos microservicios sin cambiar el servicio de publicación.

Q

plan de consulta

Serie de pasos, como instrucciones, que se utilizan para acceder a los datos de un sistema de base de datos relacional SQL.

regresión del plan de consulta

El optimizador de servicios de la base de datos elige un plan menos óptimo que antes de un cambio determinado en el entorno de la base de datos. Los cambios en estadísticas, restricciones, configuración del entorno, enlaces de parámetros de consultas y actualizaciones del motor de base de datos PostgreSQL pueden provocar una regresión del plan.

R

Matriz RACI

Consulte [responsable, fiable, consultada e informada \(RACI\)](#).

RAG

Consulte [generación aumentada por recuperación](#).

ransomware

Software malicioso que se ha diseñado para bloquear el acceso a un sistema informático o a los datos hasta que se efectúe un pago.

Matriz RASCI

Consulte [responsable, fiable, consultada e informada \(RACI\)](#).

RCAC

Consulte [control de acceso por filas y columnas](#).

réplica de lectura

Una copia de una base de datos que se utiliza con fines de solo lectura. Puede enrutar las consultas a la réplica de lectura para reducir la carga en la base de datos principal.

rediseñar

Consulte [Las 7 R](#).

objetivo de punto de recuperación (RPO)

La cantidad de tiempo máximo aceptable desde el último punto de recuperación de datos. Esto determina qué se considera una pérdida de datos aceptable entre el último punto de recuperación y la interrupción del servicio.

objetivo de tiempo de recuperación (RTO)

La demora máxima aceptable entre la interrupción del servicio y el restablecimiento del servicio.

refactorizar

Consulte [Las 7 R](#).

Region

Conjunto de AWS recursos en un área geográfica. Cada uno Región de AWS está aislado e independiente de los demás para proporcionar tolerancia a las fallas, estabilidad y resiliencia. Para más información, consulte [Specify which Regions de AWS your account can use](#).

regresión

Una técnica de ML que predice un valor numérico. Por ejemplo, para resolver el problema de “¿A qué precio se venderá esta casa?”, un modelo de ML podría utilizar un modelo de regresión lineal para predecir el precio de venta de una vivienda en función de datos conocidos sobre ella (por ejemplo, los metros cuadrados).

volver a alojar

Consulte [Las 7 R](#).

versión

En un proceso de implementación, el acto de promover cambios en un entorno de producción.

reubicar

Consulte [Las 7 R](#).

redefinir la plataforma

Consulte [Las 7 R](#).

recomprar

Consulte [Las 7 R](#).

resiliencia

Capacidad de una aplicación para resistir interrupciones o recuperarse de ellas. Al planificar la resiliencia en la Nube de AWS, la [alta disponibilidad](#) y la [recuperación ante desastres](#) son consideraciones comunes. Para más información, consulte [Resiliencia en la Nube de AWS](#).

política basada en recursos

Una política asociada a un recurso, como un bucket de Amazon S3, un punto de conexión o una clave de cifrado. Este tipo de política especifica a qué entidades principales se les permite el acceso, las acciones compatibles y cualquier otra condición que deba cumplirse.

matriz responsable, confiable, consultada e informada (RACI)

Una matriz que define las funciones y responsabilidades de todas las partes involucradas en las actividades de migración y las operaciones de la nube. El nombre de la matriz se deriva de los tipos de responsabilidad definidos en la matriz: responsable (R), contable (A), consultado (C) e informado (I). El tipo de soporte (S) es opcional. Si incluye el soporte, la matriz se denomina matriz RASCI y, si la excluye, se denomina matriz RACI.

control receptivo

Un control de seguridad que se ha diseñado para corregir los eventos adversos o las desviaciones con respecto a su base de seguridad. Para obtener más información, consulte [Controles receptivos](#) en Implementación de controles de seguridad en AWS.

retain

Consulte [Las 7 R](#).

retirar

Consulte [Las 7 R](#).

Generación aumentada de recuperación (RAG)

Tecnología de [IA generativa](#) mediante la que un [LLM](#) hace referencia a un origen de datos autorizado que se encuentra fuera de sus orígenes de datos de entrenamiento antes de generar una respuesta. Por ejemplo, un modelo de RAG podría hacer una búsqueda semántica en la base de conocimientos o en los datos personalizados de una organización. Para más información, consulte [¿Qué es RAG \(generación aumentada por recuperación\)?](#)

rotación

Proceso mediante el que periódicamente se actualiza un [secreto](#) para que resulte más difícil que un atacante pueda acceder a las credenciales.

control de acceso por filas y columnas (RCAC)

El uso de expresiones SQL básicas y flexibles que tienen reglas de acceso definidas. El RCAC consta de permisos de fila y máscaras de columnas.

RPO

Consulte [objetivo de punto de recuperación](#).

RTO

Consulte [objetivo de tiempo de recuperación](#).

manual de procedimientos

Conjunto de procedimientos manuales o automatizados necesarios para realizar una tarea específica. Por lo general, se diseñan para agilizar las operaciones o los procedimientos repetitivos con altas tasas de error.

S

SAML 2.0

Un estándar abierto que utilizan muchos proveedores de identidad (IdPs). Esta función permite el inicio de sesión único (SSO) federado, de modo que los usuarios pueden iniciar sesión Consola de administración de AWS o llamar a las operaciones de la AWS API sin tener que crear un usuario en IAM para todos los miembros de la organización. Para obtener más información sobre la federación basada en SAML 2.0, consulte [Acerca de la federación basada en SAML 2.0](#) en la documentación de IAM.

SCADA

Consulte [control de supervisión y adquisición de datos](#).

SCP

Consulte [política de control de servicio](#).

secreta

En AWS Secrets Manager, información confidencial o restringida, como una contraseña o credenciales de usuario, que se almacena de forma cifrada. Se compone del valor del secreto y de sus metadatos. El valor del secreto puede ser binario, una sola cadena o varias cadenas. Para más información, consulte [What's in a Secrets Manager secret?](#) en la documentación de Secrets Manager.

seguridad desde el diseño

Enfoque de ingeniería de sistemas que tiene en cuenta la seguridad durante todo el proceso de desarrollo.

control de seguridad

Barrera de protección técnica o administrativa que impide, detecta o reduce la capacidad de un agente de amenazas para aprovechar una vulnerabilidad de seguridad. Existen cuatro tipos de controles de seguridad principales: [preventivos](#), [de detección](#), [de respuesta](#) y [proactivos](#).

refuerzo de la seguridad

Proceso de reducir la superficie expuesta a ataques para hacerla más resistente a los ataques. Esto puede incluir acciones, como la eliminación de los recursos que ya no se necesitan, la implementación de prácticas recomendadas de seguridad consistente en conceder privilegios mínimos o la desactivación de características innecesarias en los archivos de configuración.

sistema de información sobre seguridad y administración de eventos (SIEM)

Herramientas y servicios que combinan sistemas de administración de información sobre seguridad (SIM) y de administración de eventos de seguridad (SEM). Un sistema de SIEM recopila, monitorea y analiza los datos de servidores, redes, dispositivos y otras fuentes para detectar amenazas y brechas de seguridad y generar alertas.

automatización de la respuesta de seguridad

Acción predefinida y programada que está diseñada para responder automáticamente a un evento de seguridad o corregirlo. Estas automatizaciones sirven como controles de seguridad [preventivos o adaptables](#) que le ayudan a implementar las mejores prácticas AWS de seguridad. La modificación de un grupo de seguridad de VPC, la aplicación de revisiones a una instancia de Amazon EC2 o la rotación de credenciales son algunos ejemplos de acciones de respuesta automatizadas.

cifrado del servidor

Cifrado de los datos en su destino, por parte de Servicio de AWS quien los recibe.

política de control de servicio (SCP)

Una política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. Las SCP definen barreras de protección o establecen límites a las acciones que un administrador puede delegar en los usuarios o roles. Puede utilizar las SCP como listas de permitidos o rechazados, para especificar qué servicios o acciones se encuentra permitidos o prohibidos. Para obtener más información, consulte [las políticas de control de servicios](#) en la AWS Organizations documentación.

punto de enlace de servicio

La URL del punto de entrada de un Servicio de AWS. Para conectarse mediante programación a un servicio de destino, puede utilizar un punto de conexión. Para obtener más información, consulte [Puntos de conexión de Servicio de AWS](#) en Referencia general de AWS.

acuerdo de nivel de servicio (SLA)

Acuerdo que aclara lo que un equipo de TI se compromete a ofrecer a los clientes, como el tiempo de actividad y el rendimiento del servicio.

indicador de nivel de servicio (SLI)

Medición de un aspecto del rendimiento de un servicio, como la tasa de errores, la disponibilidad o el rendimiento.

objetivo de nivel de servicio (SLO)

Métrica objetivo que representa el estado de un servicio medido mediante un [indicador de nivel de servicio](#).

modelo de responsabilidad compartida

Un modelo que describe la responsabilidad con AWS la que compartes la seguridad y el cumplimiento de la nube. AWS es responsable de la seguridad de la nube, mientras que usted es responsable de la seguridad en la nube. Para obtener más información, consulte el [Modelo de responsabilidad compartida](#).

Shadow AI

Aplicaciones de [IA](#) no autorizadas creadas o utilizadas fuera de los canales regulados dentro de una organización.

SIEM

Consulte [sistema de administración de eventos e información de seguridad](#).

único punto de error (SPOF)

Error en un único componente crítico de una aplicación que puede interrumpir el sistema.

SLA

Consulte [acuerdo de nivel de servicio](#).

SLI

Consulte [indicador de nivel de servicio](#).

SLO

Consulte [objetivo de nivel de servicio](#).

modelo de dividir y sembrar

Un patrón para escalar y acelerar los proyectos de modernización. A medida que se definen las nuevas funciones y los lanzamientos de los productos, el equipo principal se divide para crear nuevos equipos de productos. Esto ayuda a ampliar las capacidades y los servicios de su organización, mejora la productividad de los desarrolladores y apoya la innovación rápida. Para más información, consulte [Phased approach to modernizing applications in the Nube de AWS](#).

SPOF

Consulte [único punto de error](#).

esquema en estrella

Estructura organizativa de una base de datos que utiliza una tabla de hechos de gran tamaño para almacenar datos transaccionales o medidos y una o varias tablas dimensionales más pequeñas para almacenar los atributos de los datos. Esta estructura está diseñada para utilizarse en un [almacén de datos](#) o con fines de inteligencia empresarial.

patrón de higo estrangulador

Un enfoque para modernizar los sistemas monolíticos mediante la reescritura y el reemplazo gradual de las funciones del sistema hasta que se pueda dismantelar el sistema heredado. Este patrón utiliza la analogía de una higuera que crece hasta convertirse en un árbol estable y, finalmente, se apodera y reemplaza a su host. El patrón fue [presentado por Martin Fowler](#) como una forma de gestionar el riesgo al reescribir sistemas monolíticos. Para ver un ejemplo de cómo aplicar este patrón, consulte [Modernización gradual de los servicios web antiguos de Microsoft ASP.NET \(ASMX\) mediante contenedores y Amazon API Gateway](#).

subred

Un intervalo de direcciones IP en la VPC. Una subred debe residir en una sola zona de disponibilidad.

control de supervisión y adquisición de datos (SCADA)

En el sector de la fabricación, sistema que utiliza hardware y software para supervisar los activos físicos y las operaciones de producción.

cifrado simétrico

Un algoritmo de cifrado que utiliza la misma clave para cifrar y descifrar los datos.

pruebas sintéticas

Prueba de un sistema de manera que simule las interacciones de los usuarios para detectar posibles problemas o supervisar el rendimiento. Puede usar [Amazon CloudWatch Synthetics](#) para crear estas pruebas.

petición del sistema

Técnica para proporcionar contexto, instrucciones o pautas a un [LLM](#) para dirigir su comportamiento. Las peticiones del sistema ayudan a establecer el contexto y las reglas para las interacciones con los usuarios.

T

etiquetas

Key-value pares que actúan como metadatos para organizar sus AWS recursos. Las etiquetas pueden ayudar a administrar, identificar, organizar, buscar y filtrar recursos de . Para obtener más información, consulte [Etiquetado de los recursos de AWS](#).

variable de destino

El valor que intenta predecir en el ML supervisado. Esto también se conoce como variable de resultado. Por ejemplo, en un entorno de fabricación, la variable objetivo podría ser un defecto del producto.

lista de tareas

Herramienta que se utiliza para hacer un seguimiento del progreso mediante un manual de procedimientos. La lista de tareas contiene una descripción general del manual de procedimientos y una lista de las tareas generales que deben completarse. Para cada tarea general, se incluye la cantidad estimada de tiempo necesario, el propietario y el progreso.

entorno de prueba

Consulte [entorno](#).

entrenamiento

Proporcionar datos de los que pueda aprender su modelo de ML. Los datos de entrenamiento deben contener la respuesta correcta. El algoritmo de aprendizaje encuentra patrones en los datos de entrenamiento que asignan los atributos de los datos de entrada al destino (la respuesta que desea predecir). Genera un modelo de ML que captura estos patrones. Luego, el modelo de ML se puede utilizar para obtener predicciones sobre datos nuevos para los que no se conoce el destino.

herramienta

Una función o API que un [agente](#) puede invocar para realizar operaciones en sistemas externos.

puerta de enlace de tránsito

Centro de tránsito de red que puede utilizar para interconectar las VPC y las redes en las instalaciones. Para obtener más información, consulte [Qué es una pasarela de tránsito](#) en la AWS Transit Gateway documentación.

flujo de trabajo basado en enlaces troncales

Un enfoque en el que los desarrolladores crean y prueban características de forma local en una rama de característica y, a continuación, combinan esos cambios en la rama principal. Luego, la rama principal se adapta a los entornos de desarrollo, preproducción y producción, de forma secuencial.

acceso de confianza

Otorgar permisos a un servicio que especifique para realizar tareas en su organización AWS Organizations y en sus cuentas en su nombre. El servicio de confianza crea un rol vinculado al servicio en cada cuenta, cuando ese rol es necesario, para realizar las tareas de administración por usted. Para obtener más información, consulte [AWS Organizations Utilización con otros AWS servicios](#) en la AWS Organizations documentación.

ajuste

Cambiar aspectos de su proceso de formación a fin de mejorar la precisión del modelo de ML. Por ejemplo, puede entrenar el modelo de ML al generar un conjunto de etiquetas, incorporar etiquetas y, luego, repetir estos pasos varias veces con diferentes ajustes para optimizar el modelo.

equipo de dos pizzas

Un DevOps equipo pequeño al que puedes alimentar con dos pizzas. Un equipo formado por dos integrantes garantiza la mejor oportunidad posible de colaboración en el desarrollo de software.

U

incertidumbre

Un concepto que hace referencia a información imprecisa, incompleta o desconocida que puede socavar la fiabilidad de los modelos predictivos de ML. Hay dos tipos de incertidumbre: la incertidumbre epistémica se debe a datos limitados e incompletos, mientras que la incertidumbre aleatoria se debe al ruido y la aleatoriedad inherentes a los datos.

tareas indiferenciadas

También conocido como tareas arduas, es el trabajo que es necesario para crear y operar una aplicación, pero que no proporciona un valor directo al usuario final ni proporciona una ventaja

competitiva. Algunos ejemplos de tareas indiferenciadas son la adquisición, el mantenimiento y la planificación de la capacidad.

entornos superiores

Consulte [entorno](#).

V

succión

Una operación de mantenimiento de bases de datos que implica limpiar después de las actualizaciones incrementales para recuperar espacio de almacenamiento y mejorar el rendimiento.

control de versión

Procesos y herramientas que realizan un seguimiento de los cambios, como los cambios en el código fuente de un repositorio.

Emparejamiento de VPC

Conexión entre dos VPC que permite enrutar el tráfico mediante direcciones IP privadas. Para obtener más información, consulte [¿Qué es una interconexión de VPC?](#) en la documentación de Amazon VPC.

vulnerabilidad

Defecto de software o hardware que pone en peligro la seguridad del sistema.

W

caché caliente

Un búfer caché que contiene datos actuales y relevantes a los que se accede con frecuencia. La instancia de base de datos puede leer desde la caché del búfer, lo que es más rápido que leer desde la memoria principal o el disco.

datos templados

Datos a los que el acceso es infrecuente. Al consultar este tipo de datos, normalmente se aceptan consultas moderadamente lentas.

función de ventana

Función SQL que hace un cálculo en un grupo de filas que se relacionan de alguna manera con el registro actual. Las funciones de ventana son útiles para las tareas de procesamiento, como calcular una media móvil o acceder al valor de las filas en función de la posición relativa de la fila actual.

carga de trabajo

Conjunto de recursos y código que ofrece valor comercial, como una aplicación orientada al cliente o un proceso de backend.

flujo de trabajo

Grupos funcionales de un proyecto de migración que son responsables de un conjunto específico de tareas. Cada flujo de trabajo es independiente, pero respalda a los demás flujos de trabajo del proyecto. Por ejemplo, el flujo de trabajo de la cartera es responsable de priorizar las aplicaciones, planificar las oleadas y recopilar los metadatos de migración. El flujo de trabajo de la cartera entrega estos recursos al flujo de trabajo de migración, que luego migra los servidores y las aplicaciones.

WORM

Consulte [escritura única y lectura múltiple](#).

WQF

Consulte [AWS Workload Qualification Framework](#).

escritura única y lectura múltiple (WORM)

Modelo de almacenamiento que escribe los datos una sola vez y evita que se eliminen o modifiquen. Los usuarios autorizados pueden leer los datos tantas veces como sea necesario, pero no los pueden cambiar. Esta infraestructura de almacenamiento de datos se considera [inmutable](#).

Z

ataque de día cero

Ataque, normalmente de malware, que se aprovecha de una [vulnerabilidad de día cero](#).

vulnerabilidad de día cero

Un defecto o una vulnerabilidad sin mitigación en un sistema de producción. Los agentes de amenazas pueden usar este tipo de vulnerabilidad para atacar el sistema. Los desarrolladores suelen darse cuenta de la vulnerabilidad a raíz del ataque.

peticiones desde cero

Proporcionar a un [LLM](#) instrucciones para llevar a cabo una tarea, pero sin ejemplos (pasos) que puedan ayudar a guiarlo. El LLM debe usar los conocimientos del entrenamiento previo para llevar a cabo la tarea. La eficacia de la petición desde cero depende de la complejidad de la tarea y de la calidad de la petición. Consulte también [peticiones con pocos pasos](#).

aplicación zombi

Aplicación que utiliza un promedio de CPU y memoria menor al 5 por ciento. En un proyecto de migración, es habitual retirar estas aplicaciones.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la versión original de inglés, prevalecerá la versión en inglés.