



Melhores práticas de consulta para o Amazon Redshift

AWS Orientação prescritiva



AWS Orientação prescritiva: Melhores práticas de consulta para o Amazon Redshift

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens comerciais da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestige a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

Introdução	1
Visão geral	1
Público-alvo	1
Objetivos	1
Componentes da arquitetura	2
Fatores de desempenho da consulta	7
Propriedades da tabela	7
Chaves de classificação	7
Compactação de dados	8
Distribuição de dados	8
Manutenção da mesa	8
Configuração do cluster	10
Tipo de nó	10
Tamanho do nó, número de nós e fatias	10
Gerenciamento do workload	10
Aceleração de consulta breve	11
Consulta SQL	11
Estrutura da consulta	11
Compilação de código	11
Práticas recomendadas para tabelas	13
Entenda como as chaves de classificação funcionam	13
Dicas de ajuste de consultas	13
Avalie a eficácia da chave de classificação	14
Conheça sua mesa	15
Escolha o estilo certo de distribuição de tabelas	15
Práticas recomendadas para consultas	17
Evite usar a instrução SELECT * FROM	17
Identifique problemas de consulta	17
Obtenha informações resumidas sobre sua consulta	17
Evite junções cruzadas	17
Evite funções em predicados de consulta	18
Evite conversões de elenco desnecessárias	19
Use expressões CASE para agregações complexas	19
Use subconsultas	20

Use predicados	20
Adicione predicados para filtrar tabelas com junções	20
Use os operadores mais baratos para predicados	21
Use chaves de classificação nas cláusulas GROUP BY	21
Aproveite as visualizações materializadas	22
Tenha cuidado com as colunas nas cláusulas GROUP BY e ORDER BY	22
Melhores práticas para o Redshift Spectrum	23
Redução de predicados no Redshift Spectrum	24
Dicas de ajuste de consulta para o Redshift Spectrum	25
Recursos	26
Histórico do documento	27
Glossário	28
#	28
A	29
B	32
C	34
D	37
E	41
F	43
G	45
H	46
eu	48
L	50
M	51
O	56
P	58
Q	61
R	62
S	65
T	69
U	70
V	71
W	71
Z	72
.....	Ixxiv

Melhores práticas de consulta para o Amazon Redshift

Ethan Stark, Amazon Web Services (AWS)

Junho de 2024 ([histórico do documento](#))

Visão geral

Este guia fornece recomendações e melhores práticas para otimizar o desempenho de consultas e tabelas no [Amazon Redshift](#). Você pode usar o Amazon Redshift para consultar petabytes de dados estruturados e semiestruturados em seu data warehouse e em seu data lake usando SQL padrão. Este guia também fornece uma visão geral dos principais componentes da arquitetura de um data warehouse do Amazon Redshift. Esse conhecimento, junto com a compreensão dos fatores de desempenho da consulta, como propriedades da tabela, configuração do cluster e estrutura da consulta, pode ajudá-lo a criar tabelas e consultas eficientes e eficazes para seu armazém de dados do Amazon Redshift.

Público-alvo

Este guia é destinado a engenheiros de dados, arquitetos de dados e analistas de dados que projetam ou usam tabelas e consultas no Amazon Redshift.

Objetivos

Este guia pode ajudar você e sua organização a alcançar os seguintes objetivos:

- Tabelas de design para operações ideais de armazenamento e recuperação de dados
- Consultas de design para otimizar o desempenho e a economia de custos
- Otimize o desempenho do [Amazon Redshift Spectrum](#) para consultar dados diretamente de arquivos no [Amazon Simple Storage Service \(Amazon S3\)](#)

Componentes de arquitetura de um data warehouse do Amazon Redshift

Recomendamos que você tenha uma compreensão básica dos principais componentes da arquitetura em um data warehouse do Amazon Redshift. Esse conhecimento pode ajudá-lo a entender melhor como criar suas consultas e tabelas para obter um desempenho ideal.

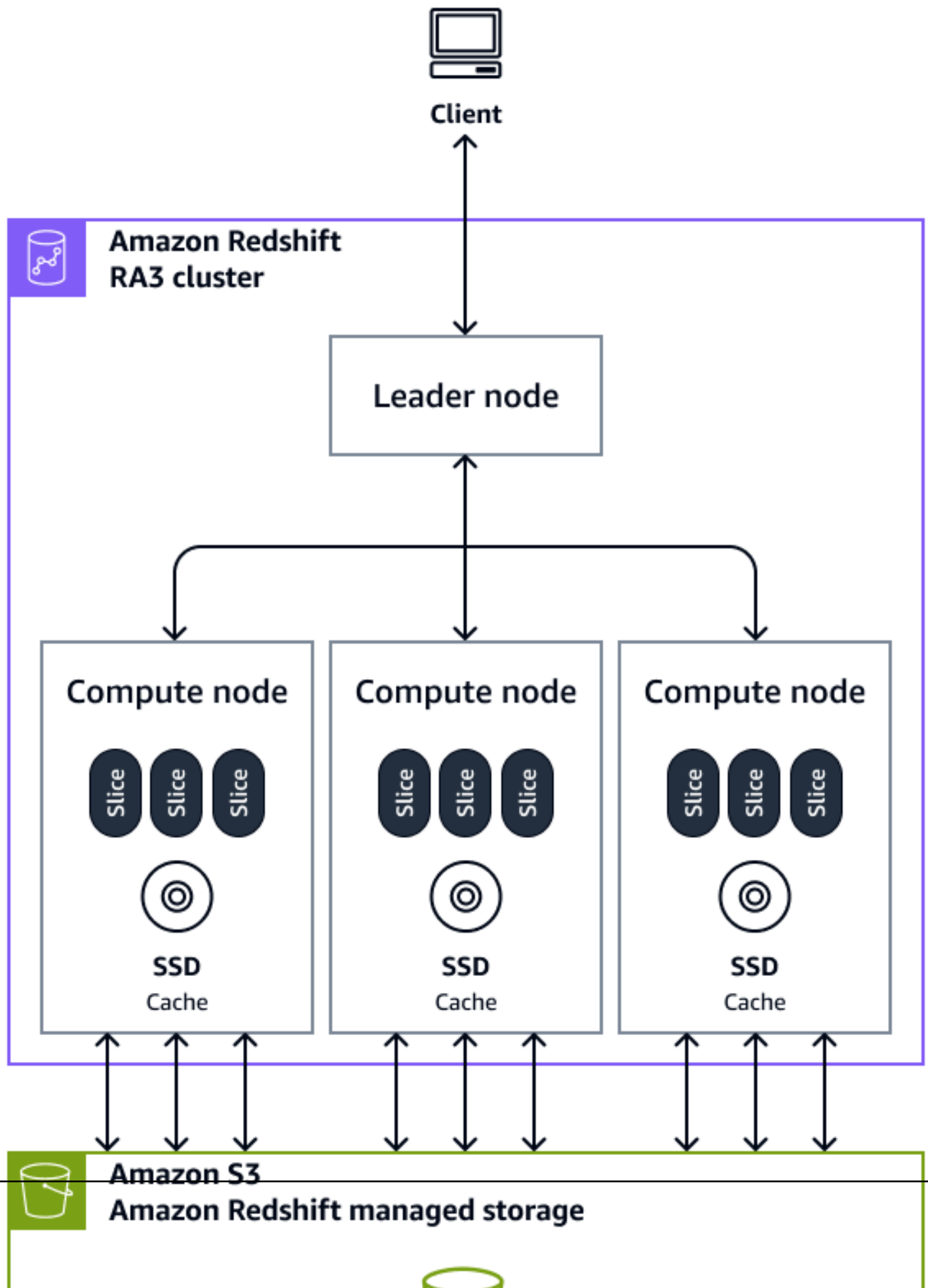
Um data warehouse no Amazon Redshift consiste nos seguintes componentes principais da arquitetura:

- **Clusters** — Um cluster, composto por um ou mais nós de computação, é o principal componente da infraestrutura de um data warehouse do Amazon Redshift. Os nós de computação são transparentes para aplicativos externos, mas seu aplicativo cliente interage diretamente somente com o nó líder. Um cluster típico tem dois ou mais nós de computação. Os nós de computação são coordenados por meio do nó líder.
- **Nó líder** — Um nó líder gerencia as comunicações dos programas clientes e de todos os nós de computação. Um nó líder também prepara os planos para executar uma consulta sempre que uma consulta for enviada a um cluster. Quando os planos estão prontos, o nó líder compila o código, distribui o código compilado para os nós de computação e, em seguida, atribui fatias de dados a cada nó de computação para processar os resultados da consulta.
- **Nó de computação** — Um nó de computação executa uma consulta. O nó líder compila o código para elementos individuais do plano para executar a consulta e atribui o código a nós de computação individuais. Os nós de computação executam o código compilado e reenviam os resultados intermediários ao nó líder para agregação final. Cada nó de computação tem sua própria CPU dedicada, memória e armazenamento em disco conectado. À medida que o workload cresce, você pode aumentar a capacidade computacional e a capacidade de armazenamento de um cluster aumentando o número de nós, atualizando o tipo de nó ou ambos.
- **Fatia do nó** — um nó de computação é particionado em unidades chamadas fatias. Cada fatia em um nó de computação recebe uma parte da memória e do espaço em disco do nó, onde processa uma parte da carga de trabalho atribuída ao nó. Assim, as fatias funcionam em paralelo para completar a operação. Os dados são distribuídos entre fatias com base no [estilo de distribuição](#) e na chave de distribuição de uma tabela específica. Uma distribuição uniforme dos dados possibilita que o Amazon Redshift atribua uniformemente cargas de trabalho às fatias e maximize os benefícios do processamento paralelo. O número de fatias por nó de computação é decidido

com base no tipo de nó. Para obter mais informações, consulte [Clusters e nós no Amazon Redshift na documentação](#) do Amazon Redshift.

- **Processamento paralelo massivo (MPP)** — O Amazon Redshift usa a arquitetura MPP para processar dados rapidamente, até mesmo consultas complexas e grandes quantidades de dados. Vários nós de computação executam o mesmo código de consulta em partes dos dados para maximizar o processamento paralelo.
- **Aplicativo cliente** — O Amazon Redshift se integra a várias ferramentas de carregamento, extração, transformação e carregamento (ETL) de dados, relatórios de business intelligence (BI), mineração de dados e análise. Todos os aplicativos cliente se comunicam com o cluster somente por meio do nó principal.

O diagrama a seguir mostra como os componentes de arquitetura de um data warehouse do Amazon Redshift trabalham juntos para acelerar as consultas.



Há sete estágios do ciclo de vida da consulta:

1. Recepção e análise de consultas:

- O nó líder recebe a consulta e analisa o SQL.
- O analisador produz uma árvore de consulta inicial, que representa a estrutura lógica da consulta original.
- O Amazon Redshift alimenta essa árvore de consultas no otimizador de consultas.

2. Otimização de consultas:

- O otimizador avalia a consulta e, se necessário, a reescreve para maximizar a eficiência.
- Esse processo de otimização pode envolver a criação de várias consultas relacionadas para substituir uma única.

3. Geração do plano de consulta:

- O otimizador gera um plano de consulta (ou vários planos, se necessário) para execução.
- O plano de consulta especifica as opções de execução, como tipos de junção, ordem de junção, métodos de agregação e requisitos de distribuição de dados.

4. Tradução do mecanismo de execução:

- O mecanismo de execução traduz o plano de consulta em etapas, segmentos e fluxos discretos:
 - Etapa — Representa uma operação individual necessária durante a execução da consulta. As etapas podem ser combinadas para permitir que os nós de computação realizem consultas, junções ou outras operações de banco de dados.
 - Segmento — Combina várias etapas que um único processo pode executar. É a menor unidade de compilação executável por uma fatia do nó de computação. (Uma fatia é a unidade de processamento paralelo no Amazon Redshift.)
 - Stream — Uma coleção de segmentos distribuídos em fatias de nós de computação disponíveis.
- O mecanismo de execução gera código compilado com base nessas etapas, segmentos e fluxos. O código compilado é executado mais rápido do que o código interpretado e consome menos capacidade computacional.
- O nó líder transmite o código compilado para os nós de computação.

5. Execução paralela:

- Essa etapa ocorre uma vez para cada fluxo.
- As fatias dos nós de computação executam segmentos de consulta em paralelo.

- Durante esse processo, o Amazon Redshift otimiza a comunicação de rede, o uso da memória e o gerenciamento de disco para transmitir resultados intermediários de uma etapa do plano de consulta para a próxima.
- Essa otimização contribui para uma execução mais rápida da consulta.

6. Processamento de fluxo:

- Essa etapa ocorre uma vez para cada fluxo.
- O mecanismo cria segmentos executáveis para cada fluxo, para um processamento paralelo eficiente.

7. Classificação final e agregação:

- O nó líder aborda qualquer classificação ou agregação final exigida pela consulta.
- Depois de concluído, o nó líder retorna os resultados ao cliente.

Para obter informações sobre componentes de arquitetura, consulte [Arquitetura do sistema de data warehouse](#) na documentação do Amazon Redshift.

Fatores de desempenho da consulta para o Amazon Redshift

Inúmeros fatores podem afetar a performance da consulta. Os seguintes aspectos de suas operações de dados, cluster e banco de dados desempenham um papel na rapidez com que suas consultas são processadas:

- [Propriedades da tabela](#)
 - [Chaves de classificação](#)(Consultor do Amazon Redshift)
 - [Compactação de dados](#)(automatizado)
 - [Distribuição de dados](#)(automatizado)
 - [Manutenção da mesa](#)(automatizado)
- [Configuração do cluster](#)
 - [Tipo de nó](#)
 - [Tamanho do nó, número de nós e fatias](#)
 - [Gerenciamento do workload](#)(automatizado)
 - [Aceleração de consulta breve](#)(automatizado)
- [Consulta SQL](#)
 - [Estrutura da consulta](#)
 - [Compilação de código](#)

Propriedades da tabela

As tabelas do Amazon Redshift são as unidades fundamentais para armazenar dados no Amazon Redshift, e cada tabela tem um conjunto de propriedades que determinam seu comportamento e acessibilidade. Essas propriedades incluem classificação, estilo de distribuição, codificação de compressão e muitas outras. Compreender essas propriedades é crucial para otimizar o desempenho, a segurança e a economia das tabelas do Amazon Redshift.

Chaves de classificação

O Amazon Redshift armazena dados em disco em ordem classificada de acordo com as chaves de classificação de uma tabela. O otimizador de consultas e o processador de consultas usam as

informações sobre onde os dados estão localizados em um nó de computação para reduzir o número de blocos que devem ser escaneados. Isso melhora significativamente a velocidade da consulta ao reduzir a quantidade de dados a serem processados. Recomendamos que você use chaves de classificação para facilitar os filtros na WHERE cláusula. Para obter mais informações, consulte [Como trabalhar com chaves de classificação](#) na documentação do Amazon Redshift.

Compactação de dados

A compactação de dados reduz os requisitos de armazenamento, o que reduz a E/S do disco e melhora o desempenho das consultas. Quando você executa uma consulta, os dados compactados são lidos na memória e depois descompactados quando a consulta é executada. Ao carregar menos dados na memória, o Amazon Redshift pode alocar mais memória para analisar os dados. Como o armazenamento colunar armazena dados semelhantes sequencialmente, o Amazon Redshift pode aplicar codificações de compactação adaptáveis especificamente vinculadas aos tipos de dados colunares. A melhor maneira de habilitar a compactação de dados nas colunas da tabela é usando a AUTO opção no Amazon Redshift para aplicar as codificações de compactação ideais ao carregar a tabela com dados. Para saber mais sobre o uso da compactação automática de dados, consulte [Carregamento de tabelas com compactação automática](#) na documentação do Amazon Redshift.

Distribuição de dados

O Amazon Redshift armazena dados nos nós de computação de acordo com o estilo de distribuição de uma tabela. Quando você executa uma consulta, o otimizador de consulta redistribui os dados aos nós de computação conforme necessário para executar junções e agregações. A escolha do estilo correto de distribuição para uma tabela ajuda a minimizar o impacto das etapas de redistribuição ao localizar os dados onde eles devem estar antes que as junções sejam executadas. Recomendamos que você use chaves de distribuição para facilitar as junções mais comuns. Para obter mais informações, consulte Como [trabalhar com estilos de distribuição de dados](#) na documentação do Amazon Redshift.

Manutenção da mesa

Embora o Amazon Redshift ofereça o melhor desempenho do setor pronto para uso imediato para a maioria das cargas de trabalho, manter os clusters do Amazon Redshift funcionando bem exige manutenção. A atualização e a exclusão de dados criam linhas inativas que devem ser limpas, e até mesmo tabelas somente de acréscimo devem ser usadas se a ordem de acréscimo não for consistente com a chave de classificação.

Vácuo

O processo de limpeza no Amazon Redshift é essencial para a integridade e a manutenção do seu cluster do Amazon Redshift. Isso também afeta o desempenho das consultas. Como as exclusões e atualizações sinalizam os dados antigos, mas na verdade não os removem, você deve usar a limpeza para recuperar o espaço em disco ocupado pelas linhas da tabela que foram marcadas para exclusão pelas operações e anteriores. `UPDATE DELETE` O Amazon Redshift pode classificar e executar automaticamente uma `VACUUM DELETE` operação em tabelas em segundo plano.

Para limpar tabelas após um carregamento ou uma série de atualizações incrementais, você também pode executar o comando `VACUUM` no banco de dados inteiro ou em tabelas individuais. Se as tabelas tiverem chaves de classificação e as cargas da tabela não estiverem otimizadas para serem classificadas à medida que são inseridas, você deverá usar vacuums para reutilizar os dados (o que pode ser crucial para o desempenho). Para obter mais informações, consulte [Limpeza de tabelas na documentação](#) do Amazon Redshift.

Analisar

A `ANALYZE` operação atualiza os metadados estatísticos nas tabelas em um banco de dados do Amazon Redshift. Manter as estatísticas atualizadas melhora a performance das consultas, permitindo que o planejador de consultas escolha os planos ideais. O Amazon Redshift monitora continuamente seu banco de dados e executa automaticamente as operações de análise em segundo plano. Para minimizar o impacto no desempenho do sistema, a `ANALYZE` operação é executada automaticamente durante os períodos em que as cargas de trabalho são leves. Se você optar por executar explicitamente `ANALYZE`, faça o seguinte:

- Execute o `ANALYZE` comando antes de executar as consultas.
- Execute o `ANALYZE` comando no banco de dados rotineiramente no final de cada ciclo regular de carregamento ou atualização.
- Execute o `ANALYZE` comando nas novas tabelas criadas por você e nas tabelas ou colunas existentes que sofram alterações significativas.
- Considere executar `ANALYZE` operações em horários diferentes para diferentes tipos de tabelas e colunas, dependendo de seu uso em consultas e de sua propensão a alterações.
- Para economizar tempo e recursos de cluster, use a `PREDICATE COLUMNS` cláusula ao executar o `ANALYZE` comando.

Configuração do cluster

Um cluster é uma coleção de nós que realizam o armazenamento e o processamento reais dos dados. Configurar seu cluster do Amazon Redshift da maneira correta é fundamental se você quiser alcançar o seguinte:

- Alta escalabilidade e simultaneidade
- Uso eficiente do Amazon Redshift
- Melhor desempenho
- Custo mais baixo

Tipo de nó

Um cluster do Amazon Redshift pode usar um dos vários tipos de nós (RA3 DC2, e DS2). Cada tipo de nó oferece diferentes tamanhos e limites para ajudá-lo a escalar seu cluster adequadamente. O tamanho do nó determina a capacidade de armazenamento, memória, CPU e preço de cada nó no cluster. A otimização de custo e desempenho começa com a escolha do tipo e tamanho corretos do nó. Para obter mais informações sobre os tipos de nós, consulte [Visão geral dos clusters do Amazon Redshift na documentação](#) do Amazon Redshift.

Tamanho do nó, número de nós e fatias

Um nó de computação é particionado em fatias. Mais nós significam mais processadores e fatias, o que permite que suas consultas sejam processadas mais rapidamente ao executar partes da consulta simultaneamente nas fatias. No entanto, mais nós também significam maiores despesas. Isso significa que você deve encontrar o equilíbrio entre custo e desempenho adequado ao seu sistema. Para obter mais informações sobre a arquitetura de cluster do Amazon Redshift, consulte Arquitetura [do sistema de armazém de dados](#) na documentação do Amazon Redshift.

Gerenciamento do workload

O gerenciamento de carga de trabalho do Amazon Redshift (WLM) permite que os usuários gerenciem de forma flexível as filas de carga de trabalho com prioridades para que consultas curtas e de execução rápida não fiquem presas em filas atrás de consultas de longa execução. O WLM automático usa algoritmos de aprendizado de máquina (ML) para criar perfis de consultas e colocá-las na fila apropriada com os recursos apropriados, enquanto gerencia a simultaneidade

de consultas e a alocação de memória. Para obter mais informações sobre o WLM, consulte [Implementação do gerenciamento de carga de trabalho](#) na documentação do Amazon Redshift.

Aceleração de consulta breve

A aceleração de consultas curtas (SQA) prioriza consultas de curta duração em vez de consultas de longa duração. O SQA executa consultas em um espaço dedicado para que as consultas SQA não sejam forçadas a esperar em filas atrás de consultas mais longas. O SQA apenas prioriza consultas que são de execução curta e estão em uma fila definida pelo usuário. Se você usa o SQA, as consultas de curta duração começam a ser executadas mais rapidamente e você pode ver os resultados mais cedo. Se você habilitar o SQA, poderá reduzir ou eliminar as filas do WLM dedicadas a consultas de curta duração. Além disso, consultas de longa duração não precisam disputar slots em uma fila do WLM. Isso significa que você pode configurar suas filas do WLM para usar menos slots de consulta. Se você usar menor simultaneidade, a taxa de transferência de consultas aumentará e o desempenho geral do sistema será aprimorado para a maioria das cargas de trabalho. Para obter mais informações sobre SQA, consulte Como [trabalhar com aceleração de consulta curta](#) na documentação do Amazon Redshift.

Consulta SQL

Uma consulta de banco de dados é uma solicitação de dados de um banco de dados. A solicitação deve vir em um cluster do Amazon Redshift usando SQL. O Amazon Redshift oferece suporte a ferramentas de cliente SQL que se conectam por meio de Java Database Connectivity (JDBC) e Open Database Connectivity (ODBC). Você pode usar a maioria das ferramentas do cliente SQL que oferecem suporte aos drivers do JDBC ou do ODBC.

Estrutura da consulta

A forma como sua consulta é escrita afeta muito seu desempenho. Recomendamos que você escreva consultas para processar e retornar o mínimo de dados necessário para atender às suas necessidades. Para obter mais informações sobre como estruturar suas consultas, consulte a seção [Melhores práticas para criar consultas do Amazon Redshift](#) deste guia.

Compilação de código

O Amazon Redshift gera e compila código para cada plano de execução de consulta. O código compilado é executado mais rápido porque ele remove a sobrecarga de usar um interpretador. Geralmente, você tem algum custo indireto na primeira vez que o código é gerado e compilado.

Como resultado, a performance de uma consulta na primeira vez que você a executa pode ser enganoso. O custo indireto pode ser especialmente perceptível quando você executa consultas pontuais. Recomendamos que você execute a consulta uma segunda vez para determinar seu desempenho típico.

O Amazon Redshift usa um serviço de compilação sem servidor para escalar compilações de consulta além dos recursos de computação de um cluster do Amazon Redshift. Os segmentos de código compilados são armazenados em cache localmente no cluster e em um cache praticamente ilimitado. Esse cache persiste após a reinicialização do cluster. As invocações subsequentes da mesma consulta são executadas mais rapidamente porque podem pular a fase de compilação. O cache não é compatível entre as versões do Amazon Redshift, portanto, o código é recompilado quando as consultas são executadas após uma atualização de versão. Usando um serviço de compilação escalável, o Amazon Redshift consegue compilar código em paralelo para oferecer uma performance consistentemente rápida. A magnitude da aceleração do workload depende da complexidade e da simultaneidade das consultas.

Melhores práticas para criar tabelas do Amazon Redshift

Esta seção fornece uma visão geral das melhores práticas para criar tabelas de banco de dados. Recomendamos que você siga essas práticas recomendadas para obter o desempenho e a eficiência ideais das consultas.

Entenda como as chaves de classificação funcionam

O Amazon Redshift armazena seus dados no disco em ordem classificada de acordo com a chave de classificação. O otimizador de consulta Amazon Redshift usa a ordem de classificação quando determina os planos de consulta ideais. Para usar as chaves de classificação de forma eficaz, recomendamos que você faça o seguinte:

- Mantenha a mesa organizada o máximo possível.
- Use VACUUM sort para restaurar o desempenho ideal.
- Evite compactar a coluna da chave de classificação.
- Se a chave de classificação estiver compactada e a `sortkey1_skew` proporção for significativamente alta, recrie a tabela sem ativar a compactação na chave de classificação.
- Evite aplicar uma função às colunas da chave de classificação. Por exemplo, na consulta a seguir, a coluna da chave de `trans_dt` : `TIMESTAMPTZ` classificação não será usada se você convertê-la em `DATE`:

```
select order_id, order_amt
from sales
where trans_dt::date = '2021-01-08'::date
```

- Execute INSERT as operações na ordem da chave de classificação.
- Use as chaves de classificação na GROUP BY cláusula quando possível.

Dicas de ajuste de consultas

Recomendamos que você faça o seguinte para ajustar suas consultas:

- Sempre solicite chaves de classificação compostas da cardinalidade mais baixa para a cardinalidade mais alta para obter a eficácia ideal.

- Se a chave inicial em uma chave de classificação composta for relativamente exclusiva (ou seja, tiver alta cardinalidade), evite adicionar colunas adicionais à sua chave de classificação. A adição de colunas adicionais tem pouco impacto no desempenho da consulta, mas aumenta os custos de manutenção.

Avalie a eficácia da chave de classificação

Para otimizar suas consultas, você deve ser capaz de avaliar a eficácia de suas consultas.

Recomendamos que você use a visualização [SVL_QUERY_SUMMARY](#) para encontrar informações gerais sobre a execução de uma consulta. Nessa exibição, você pode usar o atributo `IS_RRSCAN` para determinar se uma etapa do EXPLAIN plano usa uma varredura de faixa restrita. Você também pode usar o atributo `rows_pre_filter` para determinar a seletividade de uma chave de classificação.

Você também pode usar uma visualização administrativa GitHub chamada [v_my_last_query_summary](#). A exibição exibe informações sobre a última consulta executada.

A declaração a seguir mostra como encontrar informações gerais sobre a execução de uma consulta.

```
select lpad(' ',stm+seg+step) || label as label,
       rows,
       bytes,
       is_diskbased,
       is_rrscan,
       rows_pre_filter
from svl_query_summary
where query = pg_last_query_id()
order by stm, seg, step;
```

A consulta anterior retorna o exemplo de saída a seguir.

label	☐ rows	bytes	is_diskbased	is_rrscan	rows_pre_filter
scan tbl=163860 name=orders	☐ 1500000	24000000	f	f	1500000
project	☐ 1500000	0	f	f	0
project	☐ 1500000	0	f	f	0
hash tbl=968	☐ 1500000	24000000	f	f	0
scan tbl=163852 name=lineitem	☐ 6001215	144029160	f	t	6001215
project	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0
hjoin tbl=968	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0

Conheça sua mesa

É importante entender as propriedades críticas da sua tabela. Para saber mais sobre sua tabela, faça o seguinte:

- Use [PG_TABLE_DEF](#) para visualizar informações sobre as colunas da tabela.
- Use [SVV_TABLE_INFO para visualizar informações](#) mais abrangentes sobre uma tabela, incluindo distorção da distribuição de dados, distorção da distribuição de chaves, tamanho da tabela e estatísticas.

Escolha o estilo certo de distribuição de tabelas

Quando você executa uma consulta, o otimizador de consulta redistribui as linhas aos nós de computação conforme necessário para executar junções e agregações. O objetivo ao selecionar um estilo de distribuição de tabela é minimizar o impacto da etapa de redistribuição localizando os dados onde eles precisam estar antes de executar a consulta.

Recomendamos a seguinte abordagem para escolher o estilo correto de distribuição de tabelas:

- Evite a transmissão e a redistribuição em um plano de execução de consultas colocando as linhas no mesmo nó. Por exemplo, ao selecionar um `DISTKEY`, você pode distribuir a tabela de fatos e a tabela unidimensional em suas colunas comuns. Escolha a maior dimensão com base no tamanho do conjunto de dados filtrado. Como somente as linhas que são usadas na junção precisam ser distribuídas, considere o tamanho do conjunto de dados após a filtragem e não apenas o tamanho da tabela.

- Verifique se não há distorção na coluna em que a chave de distribuição foi criada. Caso contrário, um nó de computação poderia realizar mais trabalho pesado do que outros. Se você notar distorção, considere alterar a coluna da chave de distribuição. Uma coluna pode ser considerada candidata a uma chave de distribuição se seus valores forem distribuídos uniformemente ou se forem altos valores cardinais.
- Se a tabela usada na condição de junção for pequena (menos de 1 GB), considere o estilo de distribuição ALL.
- Você pode compactar a chave de distribuição, mas deve evitar compactar a coluna da chave de classificação (especialmente a primeira coluna da chave de classificação).

 Note

Se você usa a otimização automática de tabelas, não precisa escolher o estilo de distribuição da sua tabela. Para obter mais informações, consulte [Trabalho com otimização automática de tabelas](#) na documentação do Amazon Redshift. Para que o Amazon Redshift escolha o estilo de distribuição apropriado, especifique AUTO para o estilo de distribuição.

Melhores práticas para criar consultas do Amazon Redshift

Esta seção fornece uma visão geral das melhores práticas para criar consultas. Recomendamos que você siga as melhores práticas desta seção para obter o desempenho e a eficiência ideais da consulta.

Evite usar a instrução SELECT * FROM

Recomendamos que você evite usar a SELECT * FROM declaração. Em vez disso, sempre liste as colunas para análise. Isso reduz o tempo de execução da consulta e verifica os custos das consultas do Amazon Redshift Spectrum.

Exemplo do que evitar

```
select *  
from sales;
```

Exemplo de melhores práticas

```
select sales_date, sales_amt  
from sales;
```

Identifique problemas de consulta

Recomendamos que você verifique a visualização [STL_ALERT_EVENT_LOG](#) para identificar e corrigir possíveis problemas com sua consulta.

Obtenha informações resumidas sobre sua consulta

Recomendamos que você use as visualizações SVL_QUERY_SUMMARY e [SVL_QUERY_REPORT](#) [para obter informações resumidas sobre suas consultas](#). Você pode usar essas informações para otimizar suas consultas.

Evite junções cruzadas

Recomendamos que você evite usar junções cruzadas, a menos que seja absolutamente necessário. Sem uma condição de junção, as junções cruzadas resultam no produto cartesiano de duas tabelas.

As junções cruzadas geralmente são executadas como junções de loop aninhado (o mais lento dos tipos de junção possíveis).

Exemplo do que evitar

```
select c.c_name,  
       n.n_name  
from tpch.customer c,  
     tpch.nation n;
```

Exemplo de melhores práticas

```
select c.c_name,  
       n.n_name  
from tpch.customer c,  
join tpch.nation n  
  on n.n_nationkey = c.c_nationkey;
```

Evite funções em predicados de consulta

Recomendamos que você evite usar funções em predicados de consulta. O uso de funções em predicados de consulta pode afetar negativamente o desempenho, pois as funções geralmente adicionam sobrecarga de processamento extra a cada linha e retardam a execução geral da consulta.

Exemplo do que evitar

```
select sum(o_totalprice)  
from tpch.orders  
where datepart(year, o_orderdate) = 1992;
```

Exemplo de melhores práticas

```
select sum(o_totalprice)  
from tpch.orders  
where o_orderdate between '1992-01-01' and '1992-12-31';
```

Evite conversões de elenco desnecessárias

Recomendamos que você evite usar a conversão de conversão desnecessária nas consultas, pois os tipos de dados de conversão consomem tempo e recursos e retardam a execução da consulta.

Exemplo do que evitar

```
select sum(o_totalprice)
from tpch.orders
where o_ordertime::date = '1992-01-01';
```

Exemplo de melhores práticas

```
select sum(o_totalprice)
from tpch.orders
where o_ordertime between '1992-01-01 00:00:00' and '1992-12-31 23:59:59';
```

Use expressões CASE para agregações complexas

Recomendamos que você use uma [expressão CASE](#) para realizar agregações complexas em vez de selecionar na mesma tabela várias vezes.

Exemplo do que evitar

```
select sum(sales_amt) as us_sales
from sales
where country = 'US';

select sum(sales_amt) as ca_sales
from sales
where country = 'CA';
```

Exemplo de melhores práticas

```
select sum(case when country = 'US' then sales_amt end) as us_sales,
       sum(case when country = 'CA' then sales_amt end) as ca_sales
from sales;
```

Use subconsultas

Recomendamos que você use subconsultas nos casos em que uma tabela na consulta é usada somente para condições de predicado e a subconsulta retorna um pequeno número de linhas (menos de cerca de 200).

Exemplo do que evitar

Se uma subconsulta retornar menos de 200 linhas:

```
select sum(order_amt) as total_sales
from sales
where region_key IN
      (select region_key
       from regions
       where state = 'CA');
```

Exemplo de melhores práticas

Se uma subconsulta retornar maior ou igual a 200 linhas:

```
select sum(o.order_amt) as total_sales
from sales o
join regions r
  on r.region_key = o.region_key
  and r.state = 'CA';
```

Use predicados

Recomendamos que você use predicados para restringir o conjunto de dados o máximo possível. Os predicados são usados no SQL para filtrar e restringir os dados retornados em uma consulta. Ao especificar condições em um predicado, você pode especificar quais linhas devem ser incluídas nos resultados da consulta com base nas condições especificadas. Isso permite que você recupere somente os dados nos quais está interessado e melhora a eficiência e a precisão de suas consultas. Para obter mais informações, consulte [Condições](#) na documentação do Amazon Redshift.

Adicione predicados para filtrar tabelas com junções

Recomendamos que você adicione predicados às tabelas de filtro que participam das uniões, mesmo que os predicados apliquem os mesmos filtros. O uso de predicados para filtrar tabelas com junções

no SQL pode melhorar o desempenho da consulta, reduzindo a quantidade de dados que devem ser processados e reduzindo o tamanho do conjunto intermediário de resultados. Ao especificar as condições para a operação de junção na WHERE cláusula, o mecanismo de execução da consulta pode eliminar linhas que não correspondem às condições antes de serem unidas. Isso resulta em um conjunto de resultados menor e na execução mais rápida da consulta.

Exemplo do que evitar

```
select p.product_name, sum(o.order_amt)
from sales o
join product p
  on r.product_key = o.product_key
where o.order_date > '2022-01-01';
```

Exemplo de melhores práticas

```
select p.product_name, sum(o.order_amt)
from sales o
join product p
  on p.product_key = o.product_key
  and p.added_date > '2022-01-01'
where o.order_date > '2022-01-01';
```

Use os operadores mais baratos para predicados

No predicado, use os operadores mais baratos que você puder. Operadores de [condição de comparação](#) são preferíveis aos operadores [LIKE](#). LIKE operadores ainda são preferíveis aos operadores [SIMILAR TO](#) ou [POSIX](#).

Use chaves de classificação nas cláusulas GROUP BY

Use chaves de classificação na GROUP BY cláusula para que o planejador de consultas possa usar uma agregação mais eficiente. Uma consulta pode se qualificar para agregação de uma fase quando sua GROUP BY lista contém somente colunas de chave de classificação, uma das quais também é a chave de distribuição. As colunas da chave de classificação na GROUP BY lista devem incluir a primeira chave de classificação, seguida por outras chaves de classificação que você deseja usar na ordem da chave de classificação.

Aproveite as visualizações materializadas

Se possível, reescreva a consulta substituindo o código complexo por uma visualização materializada, o que melhorará significativamente o desempenho da consulta. Para obter mais informações, consulte [Criar visões materializadas no Amazon Redshift](#) na documentação do Amazon Redshift.

Tenha cuidado com as colunas nas cláusulas GROUP BY e ORDER BY

Se você usar ambas as ORDER BY cláusulas GROUP BY e, certifique-se de colocar as colunas na mesma ordem em ambas as GROUP BY ORDER BY cláusulas. GROUP BY exige implicitamente que os dados sejam classificados. Se sua ORDER BY cláusula for diferente, os dados deverão ser classificados duas vezes.

Exemplo do que evitar

```
select a, b, c, sum(d)
from a_table
group by b, c, a
order by a, b, c
```

Exemplo de melhores práticas

```
select a, b, c, sum(d)
from a_table
group by a, b, c
order by a, b, c
```

Melhores práticas para usar o Amazon Redshift Spectrum

Esta seção fornece uma visão geral das melhores práticas para usar o [Amazon Redshift Spectrum](#). Recomendamos que você siga estas melhores práticas para obter o desempenho ideal ao usar o Redshift Spectrum:

- Considere que os tipos de arquivo têm uma influência significativa no desempenho da consulta do Redshift Spectrum. Para melhorar o desempenho, use arquivos codificados em colunas, como ORC ou Parquet, e use o formato CSV somente para tabelas de dimensões muito pequenas.
- Use o particionamento baseado em prefixo para aproveitar a remoção de partições. Isso significa usar filtros que são inseridos nas partições do seu data lake.
- O Redshift Spectrum é escalável automaticamente para processar grandes solicitações, então faça o máximo possível no Redshift Spectrum (por exemplo, envio de predicados).
- Preste atenção aos arquivos de partição em colunas frequentemente filtradas. Se os dados forem particionados por uma ou mais colunas filtradas, o Redshift Spectrum pode aproveitar a remoção de partições e ignorar a verificação de partições e arquivos desnecessários. Uma prática comum é dividir os dados com base no tempo.
- Você pode verificar a eficácia de suas partições e a eficiência de sua consulta do Redshift Spectrum usando a consulta a seguir.

```
Select query,
       segment,
       max(assigned_partitions) as total_partitions,
       max(qualified_partitions) as qualified_partitions
From svl_s3partition
Where query=pg_last_query_id()
Group by 1,2;
```

A consulta anterior mostra o seguinte:

- `total_partitions` — O número de partições reconhecidas pelo AWS Glue Data Catalog
- `qualified_partitions` — O número de prefixos no Amazon Simple Storage Service (Amazon S3) que são acessados para a consulta do Redshift Spectrum
- Você também pode verificar a tabela do SVL_S3QUERY_SUMMARY sistema para saber mais sobre a eficácia de suas partições e a eficiência de sua consulta do Redshift Spectrum. Para fazer isso, use a instrução a seguir.

```
Select *  
From svl_s3query_summary  
Where query=pg_last_query_id();
```

A consulta anterior retorna ainda mais informações, incluindo `is_partitioned`, e `s3_returned_rows/bytes` e `s3_scanned_rows/bytes`, além de arquivos que mostram a eficiência da remoção de partições.

Redução de predicados no Redshift Spectrum

Usar o predicate pushdown evita o consumo de recursos no cluster do Amazon Redshift. Você pode transferir várias operações SQL para a camada do Redshift Spectrum. Recomendamos aproveitar isso sempre que possível.

Lembre-se do seguinte:

- Você pode avaliar alguns tipos de operações SQL completamente dentro da camada do Redshift Spectrum, incluindo as seguintes:
 - GROUP BY cláusulas
 - Condições de comparação e correspondência de padrões (por exemplo,) LIKE
 - Funções agregadas (por exemplo, COUNT, SUM, AVG, MIN, e MAX)
 - regex_replace, to_upperdate_trunc, e outras funções
- Você não pode enviar algumas operações para a camada do Redshift Spectrum, incluindo e. DISTINCT ORDER BY Execute ORDER BY somente no nível superior da consulta, se possível, pois a classificação é feita no nó principal.
- Examine seu EXPLAIN plano de consulta para verificar se a redução de predicados é eficaz. Para encontrar partes do Redshift Spectrum em um EXPLAIN comando, siga estas etapas:
 - S3 Seq Scan
 - S3 HashAggregate
 - S3 Query Scan
 - Seq Scan PartitionInfo
 - Partition Loop
- Use o menor número de colunas em sua consulta. O Redshift Spectrum pode eliminar colunas para digitalização se os dados estiverem no formato Parquet ou ORC.

- Faça uso extensivo de partições para processamento paralelo e eliminação de partições e mantenha os tamanhos dos arquivos em pelo menos 64 MB, se possível.
- Defina `TABLE PROPERTIES 'numRows' = 'nnn'` se você usar `CREATE EXTERNAL TABLE` ou `ALTER TABLE`. O Amazon Redshift não analisa tabelas externas para gerar estatísticas de tabela que o otimizador de consultas usa para gerar um plano de consulta. Se as estatísticas não estiverem definidas, o Amazon Redshift presume que as tabelas externas são as tabelas maiores e as locais são as tabelas menores.

Dicas de ajuste de consulta para o Redshift Spectrum

Recomendamos que você tenha em mente o seguinte ao ajustar suas consultas:

- O número de nós do Redshift Spectrum que seu cluster do Amazon Redshift pode engajar para uma consulta está vinculado ao número de fatias em seu cluster.
- O redimensionamento do cluster pode beneficiar os perfis computacionais locais, os perfis de armazenamento e os recursos de consulta da consulta de data lake do Amazon S3.
- O planejador de consultas do Amazon Redshift envia predicados e agregações para a camada de consulta do Redshift Spectrum sempre que possível.
- Quando grandes quantidades de dados são retornadas do Amazon S3, o processamento é limitado pelos recursos do seu cluster.
- Como o Redshift Spectrum é escalável automaticamente para processar grandes solicitações, seu desempenho geral melhora sempre que você pode enviar o processamento para a camada do Redshift Spectrum.

Recursos

- [Melhores práticas do Amazon Redshift \(documentação\)](#) do Amazon Redshift)
- [Melhores práticas do Amazon Redshift para criar consultas](#) (documentação do Amazon Redshift)
- [Ajuste do desempenho da consulta](#) (documentação do Amazon Redshift)
- [Plano de consulta](#) (documentação do Amazon Redshift)
- [Melhorando o desempenho das consultas do Amazon Redshift Spectrum](#) (documentação do Amazon Redshift)
- [Entendendo o ciclo de vida da consulta no Amazon Redshift](#) AWS (orientação prescritiva)

Histórico do documento

A tabela a seguir descreve alterações significativas feitas neste guia. Se desejar receber notificações sobre futuras atualizações, inscreva-se em um [feed RSS](#).

Alteração	Descrição	Data
AQUA removido	Removemos as informações sobre o Advanced Query Accelerator (AQUA).	14 de junho de 2024
Publicação inicial	—	3 de fevereiro de 2023

AWS Glossário de orientação prescritiva

A seguir estão os termos comumente usados em estratégias, guias e padrões fornecidos pela Orientação AWS Prescritiva. Para sugerir entradas, use o link Fornecer feedback no final do glossário.

Números

7 Rs

Sete estratégias comuns de migração para mover aplicações para a nuvem. Essas estratégias baseiam-se nos 5 Rs identificados pela Gartner em 2011 e consistem em:

- Refatorar/rearquitetar: mova uma aplicação e modifique sua arquitetura aproveitando ao máximo os recursos nativos de nuvem para melhorar a agilidade, a performance e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migre seu banco de dados Oracle local para a edição compatível com o Amazon Aurora PostgreSQL.
- Redefinir a plataforma (mover e redefinir [mover e redefinir (lift-and-reshape)]): mova uma aplicação para a nuvem e introduza algum nível de otimização a fim de aproveitar os recursos da nuvem. Exemplo: Migre seu banco de dados Oracle local para o Amazon Relational Database Service (Amazon RDS) for Oracle no. Nuvem AWS
- Recomprar (drop and shop): mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: migre seu sistema de gerenciamento de relacionamento com o cliente (CRM) para a Salesforce.com.
- Redefinir a hospedagem (mover sem alterações [lift-and-shift])mover uma aplicação para a nuvem sem fazer nenhuma alteração a fim de aproveitar os recursos da nuvem. Exemplo: Migre seu banco de dados Oracle local para o Oracle em uma EC2 instância no. Nuvem AWS
- Realocar (mover o hipervisor sem alterações [hypervisor-level lift-and-shift]): mover a infraestrutura para a nuvem sem comprar novo hardware, reescrever aplicações ou modificar suas operações existentes. Você migra servidores de uma plataforma local para um serviço em nuvem para a mesma plataforma. Exemplo: Migrar um Microsoft Hyper-V aplicativo para o. AWS
- Reter (revisitar): mantenha as aplicações em seu ambiente de origem. Isso pode incluir aplicações que exigem grande refatoração, e você deseja adiar esse trabalho para um

momento posterior, e aplicações antigas que você deseja manter porque não há justificativa comercial para migrá-las.

- Retirar: desative ou remova aplicações que não são mais necessárias em seu ambiente de origem.

A

ABAC

Consulte controle de [acesso baseado em atributos](#).

serviços abstratos

Veja os [serviços gerenciados](#).

ACID

Veja [atomicidade, consistência, isolamento, durabilidade](#).

migração ativa-ativa

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia (por meio de uma ferramenta de replicação bidirecional ou operações de gravação dupla), e ambos os bancos de dados lidam com transações de aplicações conectadas durante a migração. Esse método oferece suporte à migração em lotes pequenos e controlados, em vez de exigir uma substituição única. É mais flexível, mas exige mais trabalho do que a migração [ativa-passiva](#).

migração ativa-passiva

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia, mas somente o banco de dados de origem manipula as transações das aplicações conectadas enquanto os dados são replicados no banco de dados de destino. O banco de dados de destino não aceita nenhuma transação durante a migração.

função agregada

Uma função SQL que opera em um grupo de linhas e calcula um único valor de retorno para o grupo. Exemplos de funções agregadas incluem SUM e MAX.

AI

Veja a [inteligência artificial](#).

AIOps

Veja as [operações de inteligência artificial](#).

anonimização

O processo de excluir permanentemente informações pessoais em um conjunto de dados. A anonimização pode ajudar a proteger a privacidade pessoal. Dados anônimos não são mais considerados dados pessoais.

antipadrões

Uma solução frequentemente usada para um problema recorrente em que a solução é contraproducente, ineficaz ou menos eficaz do que uma alternativa.

controle de aplicativos

Uma abordagem de segurança que permite o uso somente de aplicativos aprovados para ajudar a proteger um sistema contra malware.

portfólio de aplicações

Uma coleção de informações detalhadas sobre cada aplicação usada por uma organização, incluindo o custo para criar e manter a aplicação e seu valor comercial. Essas informações são fundamentais para [o processo de descoberta e análise de portfólio](#) e ajudam a identificar e priorizar as aplicações a serem migradas, modernizadas e otimizadas.

inteligência artificial (IA)

O campo da ciência da computação que se dedica ao uso de tecnologias de computação para desempenhar funções cognitivas normalmente associadas aos humanos, como aprender, resolver problemas e reconhecer padrões. Para obter mais informações, consulte [O que é inteligência artificial?](#)

operações de inteligência artificial (AIOps)

O processo de usar técnicas de machine learning para resolver problemas operacionais, reduzir incidentes operacionais e intervenção humana e aumentar a qualidade do serviço. Para obter mais informações sobre como AIOps é usado na estratégia de AWS migração, consulte o [guia de integração de operações](#).

criptografia assimétrica

Um algoritmo de criptografia que usa um par de chaves, uma chave pública para criptografia e uma chave privada para descryptografia. É possível compartilhar a chave pública porque ela não é usada na descryptografia, mas o acesso à chave privada deve ser altamente restrito.

atomicidade, consistência, isolamento, durabilidade (ACID)

Um conjunto de propriedades de software que garantem a validade dos dados e a confiabilidade operacional de um banco de dados, mesmo no caso de erros, falhas de energia ou outros problemas.

controle de acesso por atributo (ABAC)

A prática de criar permissões minuciosas com base nos atributos do usuário, como departamento, cargo e nome da equipe. Para obter mais informações, consulte [ABAC AWS](#) na documentação AWS Identity and Access Management (IAM).

fonte de dados autorizada

Um local onde você armazena a versão principal dos dados, que é considerada a fonte de informações mais confiável. Você pode copiar dados da fonte de dados autorizada para outros locais com o objetivo de processar ou modificar os dados, como anonimizá-los, redigi-los ou pseudonimizá-los.

Zona de disponibilidade

Um local distinto dentro de um Região da AWS que está isolado de falhas em outras zonas de disponibilidade e fornece conectividade de rede barata e de baixa latência a outras zonas de disponibilidade na mesma região.

AWS Estrutura de adoção da nuvem (AWS CAF)

Uma estrutura de diretrizes e melhores práticas AWS para ajudar as organizações a desenvolver um plano eficiente e eficaz para migrar com sucesso para a nuvem. AWS O CAF organiza a orientação em seis áreas de foco chamadas perspectivas: negócios, pessoas, governança, plataforma, segurança e operações. As perspectivas de negócios, pessoas e governança têm como foco habilidades e processos de negócios; as perspectivas de plataforma, segurança e operações concentram-se em habilidades e processos técnicos. Por exemplo, a perspectiva das pessoas tem como alvo as partes interessadas que lidam com recursos humanos (RH), funções de pessoal e gerenciamento de pessoal. Nessa perspectiva, o AWS CAF fornece orientação para desenvolvimento, treinamento e comunicação de pessoas para ajudar a preparar a organização para a adoção bem-sucedida da nuvem. Para obter mais informações, consulte o [site da AWS CAF](#) e o [whitepaper da AWS CAF](#).

AWS Estrutura de qualificação da carga de trabalho (AWS WQF)

Uma ferramenta que avalia as cargas de trabalho de migração do banco de dados, recomenda estratégias de migração e fornece estimativas de trabalho. AWS O WQF está incluído com AWS

Schema Conversion Tool (AWS SCT). Ela analisa esquemas de banco de dados e objetos de código, código de aplicações, dependências e características de performance, além de fornecer relatórios de avaliação.

B

bot ruim

Um [bot](#) destinado a perturbar ou causar danos a indivíduos ou organizações.

BCP

Veja o [planejamento de continuidade de negócios](#).

gráfico de comportamento

Uma visualização unificada e interativa do comportamento e das interações de recursos ao longo do tempo. É possível usar um gráfico de comportamento com o Amazon Detective para examinar tentativas de login malsucedidas, chamadas de API suspeitas e ações similares. Para obter mais informações, consulte [Dados em um gráfico de comportamento](#) na documentação do Detective.

sistema big-endian

Um sistema que armazena o byte mais significativo antes. Veja também [endianness](#).

classificação binária

Um processo que prevê um resultado binário (uma de duas classes possíveis). Por exemplo, seu modelo de ML pode precisar prever problemas como “Este e-mail é ou não é spam?” ou “Este produto é um livro ou um carro?”

filtro de bloom

Uma estrutura de dados probabilística e eficiente em termos de memória que é usada para testar se um elemento é membro de um conjunto.

blue/green deployment (implantação azul/verde)

Uma estratégia de implantação em que você cria dois ambientes separados, mas idênticos. Você executa a versão atual do aplicativo em um ambiente (azul) e a nova versão do aplicativo no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Um aplicativo de software que executa tarefas automatizadas pela Internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como rastreadores da Web que indexam informações na Internet. Alguns outros bots, conhecidos como bots ruins, têm como objetivo perturbar ou causar danos a indivíduos ou organizações.

botnet

Redes de [bots](#) infectadas por [malware](#) e sob o controle de uma única parte, conhecidas como pastor de bots ou operador de bots. As redes de bots são o mecanismo mais conhecido para escalar bots e seu impacto.

ramo

Uma área contida de um repositório de código. A primeira ramificação criada em um repositório é a ramificação principal. Você pode criar uma nova ramificação a partir de uma ramificação existente e, em seguida, desenvolver recursos ou corrigir bugs na nova ramificação. Uma ramificação que você cria para gerar um recurso é comumente chamada de ramificação de recurso. Quando o recurso estiver pronto para lançamento, você mesclará a ramificação do recurso de volta com a ramificação principal. Para obter mais informações, consulte [Sobre filiais](#) (GitHub documentação).

acesso em vidro quebrado

Em circunstâncias excepcionais e por meio de um processo aprovado, um meio rápido para um usuário obter acesso a um Conta da AWS que ele normalmente não tem permissão para acessar. Para obter mais informações, consulte o indicador [Implementar procedimentos de quebra de vidro na orientação do Well-Architected](#) AWS .

estratégia brownfield

A infraestrutura existente em seu ambiente. Ao adotar uma estratégia brownfield para uma arquitetura de sistema, você desenvolve a arquitetura de acordo com as restrições dos sistemas e da infraestrutura atuais. Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e [greenfield](#).

cache do buffer

A área da memória em que os dados acessados com mais frequência são armazenados.

capacidade de negócios

O que uma empresa faz para gerar valor (por exemplo, vendas, atendimento ao cliente ou marketing). As arquiteturas de microsserviços e as decisões de desenvolvimento podem

ser orientadas por recursos de negócios. Para obter mais informações, consulte a seção [Organizados de acordo com as capacidades de negócios](#) do whitepaper [Executar microsserviços containerizados na AWS](#).

planejamento de continuidade de negócios (BCP)

Um plano que aborda o impacto potencial de um evento disruptivo, como uma migração em grande escala, nas operações e permite que uma empresa retome as operações rapidamente.

C

CAF

Consulte [Estrutura de adoção da AWS nuvem](#).

implantação canária

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substituirá a versão atual em sua totalidade.

CCoE

Veja o [Centro de Excelência em Nuvem](#).

CDC

Veja [a captura de dados de alterações](#).

captura de dados de alterações (CDC)

O processo de rastrear alterações em uma fonte de dados, como uma tabela de banco de dados, e registrar metadados sobre a alteração. É possível usar o CDC para várias finalidades, como auditar ou replicar alterações em um sistema de destino para manter a sincronização.

engenharia do caos

Introduzir intencionalmente falhas ou eventos disruptivos para testar a resiliência de um sistema. Você pode usar [AWS Fault Injection Service \(AWS FIS\)](#) para realizar experimentos que estressam suas AWS cargas de trabalho e avaliar sua resposta.

CI/CD

Veja a [integração e a entrega contínuas](#).

classificação

Um processo de categorização que ajuda a gerar previsões. Os modelos de ML para problemas de classificação predizem um valor discreto. Os valores discretos são sempre diferentes uns dos outros. Por exemplo, um modelo pode precisar avaliar se há ou não um carro em uma imagem.

criptografia no lado do cliente

Criptografia de dados localmente, antes que o alvo os AWS service (Serviço da AWS) receba.

Centro de excelência em nuvem (CCoE)

Uma equipe multidisciplinar que impulsiona os esforços de adoção da nuvem em toda a organização, incluindo o desenvolvimento de práticas recomendadas de nuvem, a mobilização de recursos, o estabelecimento de cronogramas de migração e a liderança da organização em transformações em grande escala. Para obter mais informações, consulte as [publicações CCo E](#) no Blog de Estratégia Nuvem AWS Empresarial.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem geralmente está conectada à tecnologia de [computação de ponta](#).

modelo operacional em nuvem

Em uma organização de TI, o modelo operacional usado para criar, amadurecer e otimizar um ou mais ambientes de nuvem. Para obter mais informações, consulte [Criar seu modelo operacional de nuvem](#).

estágios de adoção da nuvem

As quatro fases pelas quais as organizações normalmente passam quando migram para o Nuvem AWS:

- Projeto: executar alguns projetos relacionados à nuvem para fins de prova de conceito e aprendizado
- Fundação — Fazer investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma landing zone, definir um CCo E, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Reinvenção: otimizar produtos e serviços e inovar na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog [The Journey Toward Cloud-First & the Stages of Adoption](#) no blog de estratégia Nuvem AWS empresarial. Para obter

informações sobre como eles se relacionam com a estratégia de AWS migração, consulte o [guia de preparação para migração](#).

CMDB

Consulte o [banco de dados de gerenciamento de configuração](#).

repositório de código

Um local onde o código-fonte e outros ativos, como documentação, amostras e scripts, são armazenados e atualizados por meio de processos de controle de versão. Os repositórios de nuvem comuns incluem GitHub ou Bitbucket Cloud. Cada versão do código é chamada de ramificação. Em uma estrutura de microsserviços, cada repositório é dedicado a uma única peça de funcionalidade. Um único pipeline de CI/CD pode usar vários repositórios.

cache frio

Um cache de buffer que está vazio, não está bem preenchido ou contém dados obsoletos ou irrelevantes. Isso afeta a performance porque a instância do banco de dados deve ler da memória principal ou do disco, um processo que é mais lento do que a leitura do cache do buffer.

dados frios

Dados que raramente são acessados e geralmente são históricos. Ao consultar esse tipo de dados, consultas lentas geralmente são aceitáveis. Mover esses dados para níveis ou classes de armazenamento de baixo desempenho e menos caros pode reduzir os custos.

visão computacional (CV)

Um campo da [IA](#) que usa aprendizado de máquina para analisar e extrair informações de formatos visuais, como imagens e vídeos digitais. Por exemplo, a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

Para uma carga de trabalho, uma alteração de configuração em relação ao estado esperado. Isso pode fazer com que a carga de trabalho se torne incompatível e, normalmente, é gradual e não intencional.

banco de dados de gerenciamento de configuração (CMDB)

Um repositório que armazena e gerencia informações sobre um banco de dados e seu ambiente de TI, incluindo componentes de hardware e software e suas configurações. Normalmente, os dados de um CMDB são usados no estágio de descoberta e análise do portfólio da migração.

pacote de conformidade

Um conjunto de AWS Config regras e ações de remediação que você pode montar para personalizar suas verificações de conformidade e segurança. Você pode implantar um pacote de conformidade como uma entidade única em uma Conta da AWS região ou em uma organização usando um modelo YAML. Para obter mais informações, consulte [Pacotes de conformidade na documentação](#). AWS Config

integração contínua e entrega contínua (CI/CD)

O processo de automatizar os estágios de origem, criação, teste, preparação e produção do processo de lançamento do software. CI/CD is commonly described as a pipeline. CI/CD pode ajudá-lo a automatizar processos, melhorar a produtividade, melhorar a qualidade do código e entregar com mais rapidez. Para obter mais informações, consulte [Benefícios da entrega contínua](#). CD também pode significar implantação contínua. Para obter mais informações, consulte [Entrega contínua versus implantação contínua](#).

CV

Veja [visão computacional](#).

D

dados em repouso

Dados estacionários em sua rede, por exemplo, dados que estão em um armazenamento.

classificação de dados

Um processo para identificar e categorizar os dados em sua rede com base em criticalidade e confidencialidade. É um componente crítico de qualquer estratégia de gerenciamento de riscos de segurança cibernética, pois ajuda a determinar os controles adequados de proteção e retenção para os dados. A classificação de dados é um componente do pilar de segurança no AWS Well-Architected Framework. Para obter mais informações, consulte [Classificação de dados](#).

desvio de dados

Uma variação significativa entre os dados de produção e os dados usados para treinar um modelo de ML ou uma alteração significativa nos dados de entrada ao longo do tempo. O desvio de dados pode reduzir a qualidade geral, a precisão e a imparcialidade das previsões do modelo de ML.

dados em trânsito

Dados que estão se movendo ativamente pela sua rede, como entre os recursos da rede.

malha de dados

Uma estrutura arquitetônica que fornece propriedade de dados distribuída e descentralizada com gerenciamento e governança centralizados.

minimização de dados

O princípio de coletar e processar apenas os dados estritamente necessários. Praticar a minimização de dados no Nuvem AWS pode reduzir os riscos de privacidade, os custos e a pegada de carbono de sua análise.

perímetro de dados

Um conjunto de proteções preventivas em seu AWS ambiente que ajudam a garantir que somente identidades confiáveis acessem recursos confiáveis das redes esperadas. Para obter mais informações, consulte [Construindo um perímetro de dados em AWS](#)

pré-processamento de dados

A transformação de dados brutos em um formato que seja facilmente analisado por seu modelo de ML. O pré-processamento de dados pode significar a remoção de determinadas colunas ou linhas e o tratamento de valores ausentes, inconsistentes ou duplicados.

proveniência dos dados

O processo de rastrear a origem e o histórico dos dados ao longo de seu ciclo de vida, por exemplo, como os dados foram gerados, transmitidos e armazenados.

titular dos dados

Um indivíduo cujos dados estão sendo coletados e processados.

data warehouse

Um sistema de gerenciamento de dados que oferece suporte à inteligência comercial, como análises. Os data warehouses geralmente contêm grandes quantidades de dados históricos e geralmente são usados para consultas e análises.

linguagem de definição de dados (DDL)

Instruções ou comandos para criar ou modificar a estrutura de tabelas e objetos em um banco de dados.

linguagem de manipulação de dados (DML)

Instruções ou comandos para modificar (inserir, atualizar e excluir) informações em um banco de dados.

DDL

Consulte a [linguagem de definição de banco](#) de dados.

deep ensemble

A combinação de vários modelos de aprendizado profundo para gerar previsões. Os deep ensembles podem ser usados para produzir uma previsão mais precisa ou para estimar a incerteza nas previsões.

Aprendizado profundo

Um subcampo do ML que usa várias camadas de redes neurais artificiais para identificar o mapeamento entre os dados de entrada e as variáveis-alvo de interesse.

defense-in-depth

Uma abordagem de segurança da informação na qual uma série de mecanismos e controles de segurança são cuidadosamente distribuídos por toda a rede de computadores para proteger a confidencialidade, a integridade e a disponibilidade da rede e dos dados nela contidos. Ao adotar essa estratégia AWS, você adiciona vários controles em diferentes camadas da AWS Organizations estrutura para ajudar a proteger os recursos. Por exemplo, uma defense-in-depth abordagem pode combinar autenticação multifatorial, segmentação de rede e criptografia.

administrador delegado

Em AWS Organizations, um serviço compatível pode registrar uma conta de AWS membro para administrar as contas da organização e gerenciar as permissões desse serviço. Essa conta é chamada de administrador delegado para esse serviço. Para obter mais informações e uma lista de serviços compatíveis, consulte [Serviços que funcionam com o AWS Organizations](#) na documentação do AWS Organizations .

implantação

O processo de criar uma aplicação, novos recursos ou correções de código disponíveis no ambiente de destino. A implantação envolve a implementação de mudanças em uma base de código e, em seguida, a criação e execução dessa base de código nos ambientes da aplicação

ambiente de desenvolvimento

Veja o [ambiente](#).

controle detectivo

Um controle de segurança projetado para detectar, registrar e alertar após a ocorrência de um evento. Esses controles são uma segunda linha de defesa, alertando você sobre eventos de segurança que contornaram os controles preventivos em vigor. Para obter mais informações, consulte [Controles detectivos](#) em Como implementar controles de segurança na AWS.

mapeamento do fluxo de valor de desenvolvimento (DVSM)

Um processo usado para identificar e priorizar restrições que afetam negativamente a velocidade e a qualidade em um ciclo de vida de desenvolvimento de software. O DVSM estende o processo de mapeamento do fluxo de valor originalmente projetado para práticas de manufatura enxuta. Ele se concentra nas etapas e equipes necessárias para criar e movimentar valor por meio do processo de desenvolvimento de software.

gêmeo digital

Uma representação virtual de um sistema real, como um prédio, fábrica, equipamento industrial ou linha de produção. Os gêmeos digitais oferecem suporte à manutenção preditiva, ao monitoramento remoto e à otimização da produção.

tabela de dimensões

Em um [esquema em estrela](#), uma tabela menor que contém atributos de dados sobre dados quantitativos em uma tabela de fatos. Os atributos da tabela de dimensões geralmente são campos de texto ou números discretos que se comportam como texto. Esses atributos são comumente usados para restringir consultas, filtrar e rotular conjuntos de resultados.

desastre

Um evento que impede que uma workload ou sistema cumpra seus objetivos de negócios em seu local principal de implantação. Esses eventos podem ser desastres naturais, falhas técnicas ou o resultado de ações humanas, como configuração incorreta não intencional ou ataque de malware.

Recuperação de desastres (RD)

A estratégia e o processo que você usa para minimizar o tempo de inatividade e a perda de dados causados por um [desastre](#). Para obter mais informações, consulte [Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no](#) AWS Well-Architected Framework.

DML

Veja a [linguagem de manipulação de banco](#) de dados.

design orientado por domínio

Uma abordagem ao desenvolvimento de um sistema de software complexo conectando seus componentes aos domínios em evolução, ou principais metas de negócios, atendidos por cada componente. Esse conceito foi introduzido por Eric Evans em seu livro, *Design orientado por domínio: lidando com a complexidade no coração do software* (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como usar o design orientado por domínio com o padrão strangler fig, consulte [Modernizar incrementalmente os serviços web herdados do Microsoft ASP.NET \(ASMX\) usando contêineres e o Amazon API Gateway](#).

DR

Veja a [recuperação de desastres](#).

detecção de deriva

Rastreando desvios de uma configuração básica. Por exemplo, você pode usar AWS CloudFormation para [detectar desvios nos recursos do sistema](#) ou AWS Control Tower para [detectar mudanças em seu landing zone](#) que possam afetar a conformidade com os requisitos de governança.

DVSM

Veja o [mapeamento do fluxo de valor do desenvolvimento](#).

E

EDA

Veja a [análise exploratória de dados](#).

EDI

Veja [intercâmbio eletrônico de dados](#).

computação de borda

A tecnologia que aumenta o poder computacional de dispositivos inteligentes nas bordas de uma rede de IoT. Quando comparada à [computação em nuvem](#), a computação de ponta pode reduzir a latência da comunicação e melhorar o tempo de resposta.

intercâmbio eletrônico de dados (EDI)

A troca automatizada de documentos comerciais entre organizações. Para obter mais informações, consulte [O que é intercâmbio eletrônico de dados](#).

Criptografia

Um processo de computação que transforma dados de texto simples, legíveis por humanos, em texto cifrado.

chave de criptografia

Uma sequência criptográfica de bits aleatórios que é gerada por um algoritmo de criptografia. As chaves podem variar em tamanho, e cada chave foi projetada para ser imprevisível e exclusiva.

endianismo

A ordem na qual os bytes são armazenados na memória do computador. Os sistemas big-endian armazenam o byte mais significativo antes. Os sistemas little-endian armazenam o byte menos significativo antes.

endpoint

Veja o [endpoint do serviço](#).

serviço de endpoint

Um serviço que pode ser hospedado em uma nuvem privada virtual (VPC) para ser compartilhado com outros usuários. Você pode criar um serviço de endpoint com AWS PrivateLink e conceder permissões a outros diretores Contas da AWS ou a AWS Identity and Access Management (IAM). Essas contas ou entidades principais podem se conectar ao serviço de endpoint de maneira privada criando endpoints da VPC de interface. Para obter mais informações, consulte [Criar um serviço de endpoint](#) na documentação do Amazon Virtual Private Cloud (Amazon VPC).

planejamento de recursos corporativos (ERP)

Um sistema que automatiza e gerencia os principais processos de negócios (como contabilidade, [MES](#) e gerenciamento de projetos) para uma empresa.

criptografia envelopada

O processo de criptografar uma chave de criptografia com outra chave de criptografia. Para obter mais informações, consulte [Criptografia de envelope](#) na documentação AWS Key Management Service (AWS KMS).

ambiente

Uma instância de uma aplicação em execução. Estes são tipos comuns de ambientes na computação em nuvem:

- ambiente de desenvolvimento: uma instância de uma aplicação em execução que está disponível somente para a equipe principal responsável pela manutenção da aplicação. Ambientes de desenvolvimento são usados para testar mudanças antes de promovê-las para ambientes superiores. Esse tipo de ambiente às vezes é chamado de ambiente de teste.
- ambientes inferiores: todos os ambientes de desenvolvimento para uma aplicação, como aqueles usados para compilações e testes iniciais.
- ambiente de produção: uma instância de uma aplicação em execução que os usuários finais podem acessar. Em um pipeline de CI/CD, o ambiente de produção é o último ambiente de implantação.
- ambientes superiores: todos os ambientes que podem ser acessados por usuários que não sejam a equipe principal de desenvolvimento. Isso pode incluir um ambiente de produção, ambientes de pré-produção e ambientes para testes de aceitação do usuário.

epic

Em metodologias ágeis, categorias funcionais que ajudam a organizar e priorizar seu trabalho. Os epics fornecem uma descrição de alto nível dos requisitos e das tarefas de implementação. Por exemplo, os épicos de segurança AWS da CAF incluem gerenciamento de identidade e acesso, controles de detetive, segurança de infraestrutura, proteção de dados e resposta a incidentes. Para obter mais informações sobre epics na estratégia de migração da AWS, consulte o [guia de implementação do programa](#).

ERP

Veja o [planejamento de recursos corporativos](#).

análise exploratória de dados (EDA)

O processo de analisar um conjunto de dados para entender suas principais características. Você coleta ou agrega dados e, em seguida, realiza investigações iniciais para encontrar padrões, detectar anomalias e verificar suposições. O EDA é realizado por meio do cálculo de estatísticas resumidas e da criação de visualizações de dados.

F

tabela de fatos

A tabela central em um [esquema em estrela](#). Ele armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: aquelas que contêm medidas e aquelas que contêm uma chave externa para uma tabela de dimensões.

falham rapidamente

Uma filosofia que usa testes frequentes e incrementais para reduzir o ciclo de vida do desenvolvimento. É uma parte essencial de uma abordagem ágil.

limite de isolamento de falhas

No Nuvem AWS, um limite, como uma zona de disponibilidade, Região da AWS um plano de controle ou um plano de dados, que limita o efeito de uma falha e ajuda a melhorar a resiliência das cargas de trabalho. Para obter mais informações, consulte [Limites de isolamento de AWS falhas](#).

ramificação de recursos

Veja a [filial](#).

recursos

Os dados de entrada usados para fazer uma previsão. Por exemplo, em um contexto de manufatura, os recursos podem ser imagens capturadas periodicamente na linha de fabricação.

importância do recurso

O quanto um recurso é importante para as previsões de um modelo. Isso geralmente é expresso como uma pontuação numérica que pode ser calculada por meio de várias técnicas, como Shapley Additive Explanations (SHAP) e gradientes integrados. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

transformação de recursos

O processo de otimizar dados para o processo de ML, incluindo enriquecer dados com fontes adicionais, escalar valores ou extrair vários conjuntos de informações de um único campo de dados. Isso permite que o modelo de ML se beneficie dos dados. Por exemplo, se a data “2021-05-27 00:15:37” for dividida em “2021”, “maio”, “quinta” e “15”, isso poderá ajudar o algoritmo de aprendizado a aprender padrões diferenciados associados a diferentes componentes de dados.

solicitação rápida

Fornecer a um [LLM](#) um pequeno número de exemplos que demonstram a tarefa e o resultado desejado antes de solicitar que ele execute uma tarefa semelhante. Essa técnica é uma aplicação do aprendizado contextual, em que os modelos aprendem com exemplos (fotos) incorporados aos prompts. Solicitações rápidas podem ser eficazes para tarefas que exigem formatação, raciocínio ou conhecimento de domínio específicos. Veja também a solicitação [zero-shot](#).

FGAC

Veja o [controle de acesso refinado](#).

Controle de acesso refinado (FGAC)

O uso de várias condições para permitir ou negar uma solicitação de acesso.

migração flash-cut

Um método de migração de banco de dados que usa replicação contínua de dados por meio da [captura de dados alterados](#) para migrar dados no menor tempo possível, em vez de usar uma abordagem em fases. O objetivo é reduzir ao mínimo o tempo de inatividade.

FM

Veja o [modelo da fundação](#).

modelo de fundação (FM)

Uma grande rede neural de aprendizado profundo que vem treinando em grandes conjuntos de dados generalizados e não rotulados. FMs são capazes de realizar uma ampla variedade de tarefas gerais, como entender a linguagem, gerar texto e imagens e conversar em linguagem natural. Para obter mais informações, consulte [O que são modelos básicos](#).

G

IA generativa

Um subconjunto de modelos de [IA](#) que foram treinados em grandes quantidades de dados e que podem usar uma simples solicitação de texto para criar novos conteúdos e artefatos, como imagens, vídeos, texto e áudio. Para obter mais informações, consulte [O que é IA generativa](#).

bloqueio geográfico

Veja as [restrições geográficas](#).

restrições geográficas (bloqueio geográfico)

Na Amazon CloudFront, uma opção para impedir que usuários em países específicos acessem distribuições de conteúdo. É possível usar uma lista de permissões ou uma lista de bloqueios para especificar países aprovados e banidos. Para obter mais informações, consulte [Restringir a distribuição geográfica do seu conteúdo](#) na CloudFront documentação.

Fluxo de trabalho do GitFlow

Uma abordagem na qual ambientes inferiores e superiores usam ramificações diferentes em um repositório de código-fonte. O fluxo de trabalho do Gitflow é considerado legado, e o fluxo de [trabalho baseado em troncos](#) é a abordagem moderna e preferida.

imagem dourada

Um instantâneo de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma imagem dourada pode ser usada para provisionar software em vários dispositivos e ajudar a melhorar a velocidade, a escalabilidade e a produtividade nas operações de fabricação de dispositivos.

estratégia greenfield

A ausência de infraestrutura existente em um novo ambiente. Ao adotar uma estratégia greenfield para uma arquitetura de sistema, é possível selecionar todas as novas tecnologias sem a restrição da compatibilidade com a infraestrutura existente, também conhecida como [brownfield](#). Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e greenfield.

barreira de proteção

Uma regra de alto nível que ajuda a governar recursos, políticas e conformidade em todas as unidades organizacionais (OUs). Barreiras de proteção preventivas impõem políticas para garantir o alinhamento a padrões de conformidade. Elas são implementadas usando políticas de controle de serviço e limites de permissões do IAM. Barreiras de proteção detectivas detectam violações de políticas e problemas de conformidade e geram alertas para remediação. Eles são implementados usando AWS Config, AWS Security Hub, Amazon GuardDuty, AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda.

H

HA

Veja a [alta disponibilidade](#).

migração heterogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que usa um mecanismo de banco de dados diferente (por exemplo, Oracle para Amazon Aurora). A migração heterogênea geralmente faz parte de um esforço de redefinição da arquitetura, e converter

o esquema pode ser uma tarefa complexa. [O AWS fornece o AWS SCT](#) para ajudar nas conversões de esquemas.

alta disponibilidade (HA)

A capacidade de uma workload operar continuamente, sem intervenção, em caso de desafios ou desastres. Os sistemas HA são projetados para realizar o failover automático, oferecer consistentemente desempenho de alta qualidade e lidar com diferentes cargas e falhas com impacto mínimo no desempenho.

modernização de historiador

Uma abordagem usada para modernizar e atualizar os sistemas de tecnologia operacional (OT) para melhor atender às necessidades do setor de manufatura. Um historiador é um tipo de banco de dados usado para coletar e armazenar dados de várias fontes em uma fábrica.

dados de retenção

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de aprendizado [de máquina](#). Você pode usar dados de retenção para avaliar o desempenho do modelo comparando as previsões do modelo com os dados de retenção.

migração homogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que compartilha o mesmo mecanismo de banco de dados (por exemplo, Microsoft SQL Server para Amazon RDS para SQL Server). A migração homogênea geralmente faz parte de um esforço de redefinição da hospedagem ou da plataforma. É possível usar utilitários de banco de dados nativos para migrar o esquema.

dados quentes

Dados acessados com frequência, como dados em tempo real ou dados translacionais recentes. Esses dados normalmente exigem uma camada ou classe de armazenamento de alto desempenho para fornecer respostas rápidas às consultas.

hotfix

Uma correção urgente para um problema crítico em um ambiente de produção. Devido à sua urgência, um hotfix geralmente é feito fora do fluxo de trabalho típico de uma DevOps versão.

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente,

a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

IaC

Veja a [infraestrutura como código](#).

Política baseada em identidade

Uma política anexada a um ou mais diretores do IAM que define suas permissões no Nuvem AWS ambiente.

aplicação ociosa

Uma aplicação que tem um uso médio de CPU e memória entre 5 e 20% em um período de 90 dias. Em um projeto de migração, é comum retirar essas aplicações ou retê-las on-premises.

IIoT

Veja a [Internet das Coisas industrial](#).

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para cargas de trabalho de produção em vez de atualizar, corrigir ou modificar a infraestrutura existente. [Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e previsíveis do que infraestruturas mutáveis](#). Para obter mais informações, consulte as melhores práticas de [implantação usando infraestrutura imutável](#) no Well-Architected AWS Framework.

VPC de entrada (admissão)

Em uma arquitetura de AWS várias contas, uma VPC que aceita, inspeciona e roteia conexões de rede de fora de um aplicativo. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

migração incremental

Uma estratégia de substituição na qual você migra a aplicação em pequenas partes, em vez de realizar uma única substituição completa. Por exemplo, é possível mover inicialmente

apenas alguns microsserviços ou usuários para o novo sistema. Depois de verificar se tudo está funcionando corretamente, mova os microsserviços ou usuários adicionais de forma incremental até poder descomissionar seu sistema herdado. Essa estratégia reduz os riscos associados a migrações de grande porte.

Indústria 4.0

Um termo que foi introduzido por [Klaus Schwab](#) em 2016 para se referir à modernização dos processos de fabricação por meio de avanços em conectividade, dados em tempo real, automação, análise e IA/ML.

infraestrutura

Todos os recursos e ativos contidos no ambiente de uma aplicação.

Infraestrutura como código (IaC)

O processo de provisionamento e gerenciamento da infraestrutura de uma aplicação por meio de um conjunto de arquivos de configuração. A IaC foi projetada para ajudar você a centralizar o gerenciamento da infraestrutura, padronizar recursos e escalar rapidamente para que novos ambientes sejam reproduzíveis, confiáveis e consistentes.

Internet industrial das coisas (IIoT)

O uso de sensores e dispositivos conectados à Internet nos setores industriais, como manufatura, energia, automotivo, saúde, ciências biológicas e agricultura. Para obter mais informações, consulte [Criando uma estratégia de transformação digital industrial da Internet das Coisas \(IIoT\)](#).

VPC de inspeção

Em uma arquitetura de AWS várias contas, uma VPC centralizada que gerencia as inspeções do tráfego de rede entre VPCs (na mesma ou em diferentes Regiões da AWS) a Internet e as redes locais. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

Internet das Coisas (IoT)

A rede de objetos físicos conectados com sensores ou processadores incorporados que se comunicam com outros dispositivos e sistemas pela Internet ou por uma rede de comunicação local. Para obter mais informações, consulte [O que é IoT?](#)

interpretabilidade

Uma característica de um modelo de machine learning que descreve o grau em que um ser humano pode entender como as previsões do modelo dependem de suas entradas. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

IoT

Consulte [Internet das Coisas](#).

Biblioteca de informações de TI (ITIL)

Um conjunto de práticas recomendadas para fornecer serviços de TI e alinhar esses serviços a requisitos de negócios. A ITIL fornece a base para o ITSM.

Gerenciamento de serviços de TI (ITSM)

Atividades associadas a design, implementação, gerenciamento e suporte de serviços de TI para uma organização. Para obter informações sobre a integração de operações em nuvem com ferramentas de ITSM, consulte o [guia de integração de operações](#).

ITIL

Consulte [a biblioteca de informações](#) de TI.

ITSM

Veja o [gerenciamento de serviços de TI](#).

L

controle de acesso baseado em etiqueta (LBAC)

Uma implementação do controle de acesso obrigatório (MAC) em que os usuários e os dados em si recebem explicitamente um valor de etiqueta de segurança. A interseção entre a etiqueta de segurança do usuário e a etiqueta de segurança dos dados determina quais linhas e colunas podem ser vistas pelo usuário.

zona de pouso

Uma landing zone é um AWS ambiente bem arquitetado, com várias contas, escalável e seguro. Um ponto a partir do qual suas organizações podem iniciar e implantar rapidamente workloads e aplicações com confiança em seu ambiente de segurança e infraestrutura. Para obter mais

informações sobre zonas de pouso, consulte [Configurar um ambiente da AWS com várias contas seguro e escalável](#).

modelo de linguagem grande (LLM)

Um modelo de [IA](#) de aprendizado profundo que é pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder perguntas, resumir documentos, traduzir texto para outros idiomas e completar frases. Para obter mais informações, consulte [O que são LLMs](#).

migração de grande porte

Uma migração de 300 servidores ou mais.

LBAC

Veja controle de [acesso baseado em etiquetas](#).

privilegio mínimo

A prática recomendada de segurança de conceder as permissões mínimas necessárias para executar uma tarefa. Para obter mais informações, consulte [Aplicar permissões de privilégios mínimos](#) na documentação do IAM.

mover sem alterações (lift-and-shift)

Veja [7 Rs](#).

sistema little-endian

Um sistema que armazena o byte menos significativo antes. Veja também [endianness](#).

LLM

Veja [um modelo de linguagem grande](#).

ambientes inferiores

Veja o [ambiente](#).

M

machine learning (ML)

Um tipo de inteligência artificial que usa algoritmos e técnicas para reconhecimento e aprendizado de padrões. O ML analisa e aprende com dados gravados, por exemplo, dados da

Internet das Coisas (IoT), para gerar um modelo estatístico baseado em padrões. Para obter mais informações, consulte [Machine learning](#).

ramificação principal

Veja a [filial](#).

malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vaziar informações confidenciais ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Tróia, spyware e keyloggers.

serviços gerenciados

Serviços da AWS para o qual AWS opera a camada de infraestrutura, o sistema operacional e as plataformas, e você acessa os endpoints para armazenar e recuperar dados. O Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB são exemplos de serviços gerenciados. Eles também são conhecidos como serviços abstratos.

sistema de execução de manufatura (MES)

Um sistema de software para rastrear, monitorar, documentar e controlar processos de produção que convertem matérias-primas em produtos acabados no chão de fábrica.

MAP

Consulte [Migration Acceleration Program](#).

mecanismo

Um processo completo no qual você cria uma ferramenta, impulsiona a adoção da ferramenta e, em seguida, inspeciona os resultados para fazer ajustes. Um mecanismo é um ciclo que se reforça e se aprimora à medida que opera. Para obter mais informações, consulte [Construindo mecanismos](#) no AWS Well-Architected Framework.

conta de membro

Todos, Contas da AWS exceto a conta de gerenciamento, que fazem parte de uma organização em AWS Organizations. Uma conta só pode ser membro de uma organização de cada vez.

MES

Veja o [sistema de execução de manufatura](#).

Transporte de telemetria de enfileiramento de mensagens (MQTT)

[Um protocolo de comunicação leve machine-to-machine \(M2M\), baseado no padrão de publicação/assinatura, para dispositivos de IoT com recursos limitados.](#)

microserviço

Um serviço pequeno e independente que se comunica de forma bem definida APIs e normalmente é de propriedade de equipes pequenas e independentes. Por exemplo, um sistema de seguradora pode incluir microserviços que mapeiam as capacidades comerciais, como vendas ou marketing, ou subdomínios, como compras, reclamações ou análises. Os benefícios dos microserviços incluem agilidade, escalabilidade flexível, fácil implantação, código reutilizável e resiliência. Para obter mais informações, consulte [Integração de microserviços usando serviços sem AWS servidor](#).

arquitetura de microserviços

Uma abordagem à criação de aplicações com componentes independentes que executam cada processo de aplicação como um microserviço. Esses microserviços se comunicam por meio de uma interface bem definida usando leveza. APIs Cada microserviço nessa arquitetura pode ser atualizado, implantado e escalado para atender à demanda por funções específicas de uma aplicação. Para obter mais informações, consulte [Implementação de microserviços em. AWS](#)

Programa de Aceleração da Migração (MAP)

Um AWS programa que fornece suporte de consultoria, treinamento e serviços para ajudar as organizações a criar uma base operacional sólida para migrar para a nuvem e ajudar a compensar o custo inicial das migrações. O MAP inclui uma metodologia de migração para executar migrações legadas de forma metódica e um conjunto de ferramentas para automatizar e acelerar cenários comuns de migração.

migração em escala

O processo de mover a maior parte do portfólio de aplicações para a nuvem em ondas, com mais aplicações sendo movidas em um ritmo mais rápido a cada onda. Essa fase usa as práticas recomendadas e lições aprendidas nas fases anteriores para implementar uma fábrica de migração de equipes, ferramentas e processos para agilizar a migração de workloads por meio de automação e entrega ágeis. Esta é a terceira fase da [estratégia de migração para a AWS](#).

fábrica de migração

Equipes multifuncionais que simplificam a migração de workloads por meio de abordagens automatizadas e ágeis. As equipes da fábrica de migração geralmente incluem operações,

analistas e proprietários de negócios, engenheiros de migração, desenvolvedores e DevOps profissionais que trabalham em sprints. Entre 20 e 50% de um portfólio de aplicações corporativas consiste em padrões repetidos que podem ser otimizados por meio de uma abordagem de fábrica. Para obter mais informações, consulte [discussão sobre fábricas de migração](#) e o [guia do Cloud Migration Factory](#) neste conjunto de conteúdo.

metadados de migração

As informações sobre a aplicação e o servidor necessárias para concluir a migração. Cada padrão de migração exige um conjunto de metadados de migração diferente. Exemplos de metadados de migração incluem a sub-rede, o grupo de segurança e AWS a conta de destino.

padrão de migração

Uma tarefa de migração repetível que detalha a estratégia de migração, o destino da migração e a aplicação ou o serviço de migração usado. Exemplo: rehospede a migração para a Amazon EC2 com o AWS Application Migration Service.

Avaliação de Portfólio para Migração (MPA)

Uma ferramenta on-line que fornece informações para validar o caso de negócios para migrar para o. Nuvem AWS O MPA fornece avaliação detalhada do portfólio (dimensionamento correto do servidor, preços, comparações de TCO, análise de custos de migração), bem como planejamento de migração (análise e coleta de dados de aplicações, agrupamento de aplicações, priorização de migração e planejamento de ondas). A [ferramenta MPA](#) (requer login) está disponível gratuitamente para todos os AWS consultores e consultores parceiros da APN.

Avaliação de Preparação para Migração (MRA)

O processo de obter insights sobre o status de prontidão de uma organização para a nuvem, identificar pontos fortes e fracos e criar um plano de ação para fechar as lacunas identificadas, usando o CAF. AWS Para mais informações, consulte o [guia de preparação para migração](#). A MRA é a primeira fase da [estratégia de migração para a AWS](#).

estratégia de migração

A abordagem usada para migrar uma carga de trabalho para o. Nuvem AWS Para obter mais informações, consulte a entrada de [7 Rs](#) neste glossário e consulte [Mobilize sua organização para acelerar migrações em grande escala](#).

ML

Veja o [aprendizado de máquina](#).

modernização

Transformar uma aplicação desatualizada (herdada ou monolítica) e sua infraestrutura em um sistema ágil, elástico e altamente disponível na nuvem para reduzir custos, ganhar eficiência e aproveitar as inovações. Para obter mais informações, consulte [Estratégia para modernizar aplicativos no Nuvem AWS](#).

avaliação de preparação para modernização

Uma avaliação que ajuda a determinar a preparação para modernização das aplicações de uma organização. Ela identifica benefícios, riscos e dependências e determina o quão bem a organização pode acomodar o estado futuro dessas aplicações. O resultado da avaliação é um esquema da arquitetura de destino, um roteiro que detalha as fases de desenvolvimento e os marcos do processo de modernização e um plano de ação para abordar as lacunas identificadas. Para obter mais informações, consulte [Avaliação da prontidão para modernização de aplicativos no Nuvem AWS](#).

aplicações monolíticas (monólitos)

Aplicações que são executadas como um único serviço com processos fortemente acoplados. As aplicações monolíticas apresentam várias desvantagens. Se um recurso da aplicação apresentar um aumento na demanda, toda a arquitetura deverá ser escalada. Adicionar ou melhorar os recursos de uma aplicação monolítica também se torna mais complexo quando a base de código cresce. Para resolver esses problemas, é possível criar uma arquitetura de microsserviços. Para obter mais informações, consulte [Decompor monólitos em microsserviços](#).

MAPA

Consulte [Avaliação do portfólio de migração](#).

MQTT

Consulte Transporte de [telemetria de enfileiramento de](#) mensagens.

classificação multiclasse

Um processo que ajuda a gerar previsões para várias classes (prevendo um ou mais de dois resultados). Por exemplo, um modelo de ML pode perguntar “Este produto é um livro, um carro ou um telefone?” ou “Qual categoria de produtos é mais interessante para este cliente?”

infraestrutura mutável

Um modelo que atualiza e modifica a infraestrutura existente para cargas de trabalho de produção. Para melhorar a consistência, confiabilidade e previsibilidade, o AWS Well-Architected Framework recomenda o uso de infraestrutura [imutável](#) como uma prática recomendada.

O

OAC

Veja o [controle de acesso de origem](#).

CARVALHO

Veja a [identidade de acesso de origem](#).

OCM

Veja o [gerenciamento de mudanças organizacionais](#).

migração offline

Um método de migração no qual a workload de origem é desativada durante o processo de migração. Esse método envolve tempo de inatividade prolongado e geralmente é usado para workloads pequenas e não críticas.

OI

Veja a [integração de operações](#).

OLA

Veja o [contrato em nível operacional](#).

migração online

Um método de migração no qual a workload de origem é copiada para o sistema de destino sem ser colocada offline. As aplicações conectadas à workload podem continuar funcionando durante a migração. Esse método envolve um tempo de inatividade nulo ou mínimo e normalmente é usado para workloads essenciais para a produção.

OPC-UA

Consulte [Comunicação de processo aberto — Arquitetura unificada](#).

Comunicação de processo aberto — Arquitetura unificada (OPC-UA)

Um protocolo de comunicação machine-to-machine (M2M) para automação industrial. O OPC-UA fornece um padrão de interoperabilidade com esquemas de criptografia, autenticação e autorização de dados.

acordo de nível operacional (OLA)

Um acordo que esclarece o que os grupos funcionais de TI prometem oferecer uns aos outros para apoiar um acordo de serviço (SLA).

análise de prontidão operacional (ORR)

Uma lista de verificação de perguntas e melhores práticas associadas que ajudam você a entender, avaliar, prevenir ou reduzir o escopo de incidentes e possíveis falhas. Para obter mais informações, consulte [Operational Readiness Reviews \(ORR\)](#) no Well-Architected AWS Framework.

tecnologia operacional (OT)

Sistemas de hardware e software que funcionam com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas OT e de tecnologia da informação (TI) é o foco principal das transformações [da Indústria 4.0](#).

integração de operações (OI)

O processo de modernização das operações na nuvem, que envolve planejamento de preparação, automação e integração. Para obter mais informações, consulte o [guia de integração de operações](#).

trilha organizacional

Uma trilha criada por ela AWS CloudTrail registra todos os eventos de todas as Contas da AWS em uma organização em AWS Organizations. Essa trilha é criada em cada Conta da AWS que faz parte da organização e monitora a atividade em cada conta. Para obter mais informações, consulte [Criação de uma trilha para uma organização](#) na CloudTrail documentação.

gerenciamento de alterações organizacionais (OCM)

Uma estrutura para gerenciar grandes transformações de negócios disruptivas de uma perspectiva de pessoas, cultura e liderança. O OCM ajuda as organizações a se prepararem e fazerem a transição para novos sistemas e estratégias, acelerando a adoção de alterações, abordando questões de transição e promovendo mudanças culturais e organizacionais. Na estratégia de AWS migração, essa estrutura é chamada de aceleração de pessoas, devido à velocidade de mudança exigida nos projetos de adoção da nuvem. Para obter mais informações, consulte o [guia do OCM](#).

controle de acesso de origem (OAC)

Em CloudFront, uma opção aprimorada para restringir o acesso para proteger seu conteúdo do Amazon Simple Storage Service (Amazon S3). O OAC oferece suporte a todos os buckets

S3 Regiões da AWS, criptografia do lado do servidor com AWS KMS (SSE-KMS) e solicitações dinâmicas ao bucket S3. PUT DELETE

Identidade do acesso de origem (OAI)

Em CloudFront, uma opção para restringir o acesso para proteger seu conteúdo do Amazon S3. Quando você usa o OAI, CloudFront cria um principal com o qual o Amazon S3 pode se autenticar. Os diretores autenticados podem acessar o conteúdo em um bucket do S3 somente por meio de uma distribuição específica. CloudFront Veja também [OAC](#), que fornece um controle de acesso mais granular e aprimorado.

ORR

Veja a [análise de prontidão operacional](#).

OT

Veja a [tecnologia operacional](#).

VPC de saída (egresso)

Em uma arquitetura de AWS várias contas, uma VPC que gerencia conexões de rede que são iniciadas de dentro de um aplicativo. A [Arquitetura de Referência de AWS Segurança](#) recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

P

limite de permissões

Uma política de gerenciamento do IAM anexada a entidades principais do IAM para definir as permissões máximas que o usuário ou perfil podem ter. Para obter mais informações, consulte [Limites de permissões](#) na documentação do IAM.

Informações de identificação pessoal (PII)

Informações que, quando visualizadas diretamente ou combinadas com outros dados relacionados, podem ser usadas para inferir razoavelmente a identidade de um indivíduo. Exemplos de PII incluem nomes, endereços e informações de contato.

PII

Veja as [informações de identificação pessoal](#).

manual

Um conjunto de etapas predefinidas que capturam o trabalho associado às migrações, como a entrega das principais funções operacionais na nuvem. Um manual pode assumir a forma de scripts, runbooks automatizados ou um resumo dos processos ou etapas necessários para operar seu ambiente modernizado.

PLC

Consulte [controlador lógico programável](#).

AMEIXA

Veja o gerenciamento [do ciclo de vida do produto](#).

política

Um objeto que pode definir permissões (consulte a [política baseada em identidade](#)), especificar as condições de acesso (consulte a [política baseada em recursos](#)) ou definir as permissões máximas para todas as contas em uma organização em AWS Organizations (consulte a política de controle de [serviços](#)).

persistência poliglota

Escolher de forma independente a tecnologia de armazenamento de dados de um microserviço com base em padrões de acesso a dados e outros requisitos. Se seus microserviços tiverem a mesma tecnologia de armazenamento de dados, eles poderão enfrentar desafios de implementação ou apresentar baixa performance. Os microserviços serão implementados com mais facilidade e alcançarão performance e escalabilidade melhores se usarem o armazenamento de dados mais bem adaptado às suas necessidades. Para obter mais informações, consulte [Habilitar a persistência de dados em microserviços](#).

avaliação do portfólio

Um processo de descobrir, analisar e priorizar o portfólio de aplicações para planejar a migração. Para obter mais informações, consulte [Avaliar a preparação para a migração](#).

predicado

Uma condição de consulta que retorna `true` ou `false`, normalmente localizada em uma WHERE cláusula.

pressão de predicados

Uma técnica de otimização de consulta de banco de dados que filtra os dados na consulta antes da transferência. Isso reduz a quantidade de dados que devem ser recuperados e processados do banco de dados relacional e melhora o desempenho das consultas.

controle preventivo

Um controle de segurança projetado para evitar que um evento ocorra. Esses controles são a primeira linha de defesa para ajudar a evitar acesso não autorizado ou alterações indesejadas em sua rede. Para obter mais informações, consulte [Controles preventivos](#) em Como implementar controles de segurança na AWS.

principal (entidade principal)

Uma entidade AWS que pode realizar ações e acessar recursos. Essa entidade geralmente é um usuário raiz para um Conta da AWS, uma função do IAM ou um usuário. Para obter mais informações, consulte Entidade principal em [Termos e conceitos de perfis](#) na documentação do IAM.

privacidade por design

Uma abordagem de engenharia de sistema que leva em consideração a privacidade em todo o processo de desenvolvimento.

zonas hospedadas privadas

Um contêiner que contém informações sobre como você deseja que o Amazon Route 53 responda às consultas de DNS para um domínio e seus subdomínios em um ou mais VPCs. Para obter mais informações, consulte [Como trabalhar com zonas hospedadas privadas](#) na documentação do Route 53.

controle proativo

Um [controle de segurança](#) projetado para impedir a implantação de recursos não compatíveis. Esses controles examinam os recursos antes de serem provisionados. Se o recurso não estiver em conformidade com o controle, ele não será provisionado. Para obter mais informações, consulte o [guia de referência de controles](#) na AWS Control Tower documentação e consulte [Controles proativos](#) em Implementação de controles de segurança em AWS.

gerenciamento do ciclo de vida do produto (PLM)

O gerenciamento de dados e processos de um produto em todo o seu ciclo de vida, desde o design, desenvolvimento e lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja o [ambiente](#).

controlador lógico programável (PLC)

Na fabricação, um computador altamente confiável e adaptável que monitora as máquinas e automatiza os processos de fabricação.

encadeamento imediato

Usando a saída de um prompt do [LLM](#) como entrada para o próximo prompt para gerar respostas melhores. Essa técnica é usada para dividir uma tarefa complexa em subtarefas ou para refinar ou expandir iterativamente uma resposta preliminar. Isso ajuda a melhorar a precisão e a relevância das respostas de um modelo e permite resultados mais granulares e personalizados.

pseudonimização

O processo de substituir identificadores pessoais em um conjunto de dados por valores de espaço reservado. A pseudonimização pode ajudar a proteger a privacidade pessoal. Os dados pseudonimizados ainda são considerados dados pessoais.

publish/subscribe (pub/sub)

Um padrão que permite comunicações assíncronas entre microsserviços para melhorar a escalabilidade e a capacidade de resposta. Por exemplo, em um [MES](#) baseado em microsserviços, um microsserviço pode publicar mensagens de eventos em um canal no qual outros microsserviços possam se inscrever. O sistema pode adicionar novos microsserviços sem alterar o serviço de publicação.

Q

plano de consulta

Uma série de etapas, como instruções, usadas para acessar os dados em um sistema de banco de dados relacional SQL.

regressão de planos de consultas

Quando um otimizador de serviço de banco de dados escolhe um plano menos adequado do que escolhia antes de uma determinada alteração no ambiente de banco de dados ocorrer. Isso pode ser causado por alterações em estatísticas, restrições, configurações do ambiente, associações de parâmetros de consulta e atualizações do mecanismo de banco de dados.

R

Matriz RACI

Veja [responsável, responsável, consultado, informado \(RACI\)](#).

RAG

Consulte [Geração Aumentada de Recuperação](#).

ransomware

Um software mal-intencionado desenvolvido para bloquear o acesso a um sistema ou dados de computador até que um pagamento seja feito.

Matriz RASCI

Veja [responsável, responsável, consultado, informado \(RACI\)](#).

RCAC

Veja o [controle de acesso por linha e coluna](#).

réplica de leitura

Uma cópia de um banco de dados usada somente para leitura. É possível encaminhar consultas para a réplica de leitura e reduzir a carga no banco de dados principal.

rearquiteta

Veja [7 Rs](#).

objetivo de ponto de recuperação (RPO).

O máximo período de tempo aceitável desde o último ponto de recuperação de dados.

Isso determina o que é considerado uma perda aceitável de dados entre o último ponto de recuperação e a interrupção do serviço.

objetivo de tempo de recuperação (RTO)

O máximo atraso aceitável entre a interrupção e a restauração do serviço.

refatorar

Veja [7 Rs](#).

Região

Uma coleção de AWS recursos em uma área geográfica. Cada um Região da AWS é isolado e independente dos outros para fornecer tolerância a falhas, estabilidade e resiliência. Para obter mais informações, consulte [Especificar o que Regiões da AWS sua conta pode usar](#).

regressão

Uma técnica de ML que prevê um valor numérico. Por exemplo, para resolver o problema de “Por qual preço esta casa será vendida?” um modelo de ML pode usar um modelo de regressão linear para prever o preço de venda de uma casa com base em fatos conhecidos sobre a casa (por exemplo, a metragem quadrada).

redefinir a hospedagem

Veja [7 Rs](#).

versão

Em um processo de implantação, o ato de promover mudanças em um ambiente de produção.

realocar

Veja [7 Rs](#).

redefinir a plataforma

Veja [7 Rs](#).

recomprar

Veja [7 Rs](#).

resiliência

A capacidade de um aplicativo de resistir ou se recuperar de interrupções. [Alta disponibilidade](#) e [recuperação de desastres](#) são considerações comuns ao planejar a resiliência no. Nuvem AWS Para obter mais informações, consulte [Nuvem AWS Resiliência](#).

política baseada em recurso

Uma política associada a um recurso, como um bucket do Amazon S3, um endpoint ou uma chave de criptografia. Esse tipo de política especifica quais entidades principais têm acesso permitido, ações válidas e quaisquer outras condições que devem ser atendidas.

matriz responsável, accountable, consultada, informada (RACI)

Uma matriz que define as funções e responsabilidades de todas as partes envolvidas nas atividades de migração e nas operações de nuvem. O nome da matriz é derivado dos tipos de responsabilidade definidos na matriz: responsável (R), responsabilizável (A), consultado (C) e informado (I). O tipo de suporte (S) é opcional. Se você incluir suporte, a matriz será chamada de matriz RASCI e, se excluir, será chamada de matriz RACI.

controle responsivo

Um controle de segurança desenvolvido para conduzir a remediação de eventos adversos ou desvios em relação à linha de base de segurança. Para obter mais informações, consulte [Controles responsivos](#) em Como implementar controles de segurança na AWS.

reter

Veja [7 Rs](#).

aposentar-se

Veja [7 Rs](#).

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de [IA generativa](#) na qual um [LLM](#) faz referência a uma fonte de dados autorizada que está fora de suas fontes de dados de treinamento antes de gerar uma resposta. Por exemplo, um modelo RAG pode realizar uma pesquisa semântica na base de conhecimento ou nos dados personalizados de uma organização. Para obter mais informações, consulte [O que é RAG](#).

alternância

O processo de atualizar periodicamente um [segredo](#) para dificultar o acesso das credenciais por um invasor.

controle de acesso por linha e coluna (RCAC)

O uso de expressões SQL básicas e flexíveis que tenham regras de acesso definidas. O RCAC consiste em permissões de linha e máscaras de coluna.

RPO

Veja o [objetivo do ponto de recuperação](#).

RTO

Veja o [objetivo do tempo de recuperação](#).

runbook

Um conjunto de procedimentos manuais ou automatizados necessários para realizar uma tarefa específica. Eles são normalmente criados para agilizar operações ou procedimentos repetitivos com altas taxas de erro.

S

SAML 2.0

Um padrão aberto que muitos provedores de identidade (IdPs) usam. Esse recurso permite o login único federado (SSO), para que os usuários possam fazer login AWS Management Console ou chamar as operações da AWS API sem que você precise criar um usuário no IAM para todos em sua organização. Para obter mais informações sobre a federação baseada em SAML 2.0, consulte [Sobre a federação baseada em SAML 2.0](#) na documentação do IAM.

SCADA

Veja [controle de supervisão e aquisição de dados](#).

SCP

Veja a [política de controle de serviços](#).

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. Ele consiste no valor secreto e em seus metadados. O valor secreto pode ser binário, uma única string ou várias strings. Para obter mais informações, consulte [O que há em um segredo do Secrets Manager?](#) na documentação do Secrets Manager.

segurança por design

Uma abordagem de engenharia de sistemas que leva em conta a segurança em todo o processo de desenvolvimento.

controle de segurança

Uma barreira de proteção técnica ou administrativa que impede, detecta ou reduz a capacidade de uma ameaça explorar uma vulnerabilidade de segurança. [Existem quatro tipos principais de controles de segurança: preventivos, detectivos, responsivos e proativos.](#)

fortalecimento da segurança

O processo de reduzir a superfície de ataque para torná-la mais resistente a ataques. Isso pode incluir ações como remover recursos que não são mais necessários, implementar a prática recomendada de segurança de conceder privilégios mínimos ou desativar recursos desnecessários em arquivos de configuração.

sistema de gerenciamento de eventos e informações de segurança (SIEM)

Ferramentas e serviços que combinam sistemas de gerenciamento de informações de segurança (SIM) e gerenciamento de eventos de segurança (SEM). Um sistema SIEM coleta, monitora e analisa dados de servidores, redes, dispositivos e outras fontes para detectar ameaças e violações de segurança e gerar alertas.

automação de resposta de segurança

Uma ação predefinida e programada projetada para responder ou remediar automaticamente um evento de segurança. Essas automações servem como controles de segurança [responsivos](#) ou [detectivos](#) que ajudam você a implementar as melhores práticas AWS de segurança. Exemplos de ações de resposta automatizada incluem a modificação de um grupo de segurança da VPC, a correção de uma instância EC2 da Amazon ou a rotação de credenciais.

Criptografia do lado do servidor

Criptografia dos dados em seu destino, por AWS service (Serviço da AWS) quem os recebe.

política de controle de serviços (SCP)

Uma política que fornece controle centralizado sobre as permissões de todas as contas em uma organização em AWS Organizations. SCPs defina barreiras ou estabeleça limites nas ações que um administrador pode delegar a usuários ou funções. Você pode usar SCPs como listas de permissão ou listas de negação para especificar quais serviços ou ações são permitidos ou proibidos. Para obter mais informações, consulte [Políticas de controle de serviço](#) na AWS Organizations documentação.

service endpoint (endpoint de serviço)

O URL do ponto de entrada para um AWS service (Serviço da AWS). Você pode usar o endpoint para se conectar programaticamente ao serviço de destino. Para obter mais informações, consulte [Endpoints do AWS service \(Serviço da AWS\)](#) na Referência geral da AWS.

acordo de serviço (SLA)

Um acordo que esclarece o que uma equipe de TI promete fornecer aos clientes, como tempo de atividade e performance do serviço.

indicador de nível de serviço (SLI)

Uma medida de um aspecto de desempenho de um serviço, como taxa de erro, disponibilidade ou taxa de transferência.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme medida por um indicador de [nível de serviço](#).

modelo de responsabilidade compartilhada

Um modelo que descreve a responsabilidade com a qual você compartilha AWS pela segurança e conformidade na nuvem. AWS é responsável pela segurança da nuvem, enquanto você é responsável pela segurança na nuvem. Para obter mais informações, consulte o [Modelo de responsabilidade compartilhada](#).

SIEM

Veja [informações de segurança e sistema de gerenciamento de eventos](#).

ponto único de falha (SPOF)

Uma falha em um único componente crítico de um aplicativo que pode interromper o sistema.

SLA

Veja o contrato [de nível de serviço](#).

ESGUIO

Veja o indicador [de nível de serviço](#).

SLO

Veja o objetivo do [nível de serviço](#).

split-and-seed modelo

Um padrão para escalar e acelerar projetos de modernização. À medida que novos recursos e lançamentos de produtos são definidos, a equipe principal se divide para criar novas equipes de produtos. Isso ajuda a escalar os recursos e os serviços da sua organização, melhora a produtividade do desenvolvedor e possibilita inovações rápidas. Para obter mais informações, consulte [Abordagem em fases para modernizar aplicativos no](#). Nuvem AWS

CUSPE

Veja [um único ponto de falha](#).

esquema de estrelas

Uma estrutura organizacional de banco de dados que usa uma grande tabela de fatos para armazenar dados transacionais ou medidos e usa uma ou mais tabelas dimensionais menores para armazenar atributos de dados. Essa estrutura foi projetada para uso em um [data warehouse](#) ou para fins de inteligência comercial.

padrão strangler fig

Uma abordagem à modernização de sistemas monolíticos que consiste em reescrever e substituir incrementalmente a funcionalidade do sistema até que o sistema herdado possa ser desativado. Esse padrão usa a analogia de uma videira que cresce e se torna uma árvore estabelecida e, eventualmente, supera e substitui sua hospedeira. O padrão foi [apresentado por Martin Fowler](#) como forma de gerenciar riscos ao reescrever sistemas monolíticos. Para ver um exemplo de como aplicar esse padrão, consulte [Modernizar incrementalmente os serviços Web herdados do Microsoft ASP.NET \(ASMX\) usando contêineres e o Amazon API Gateway](#).

sub-rede

Um intervalo de endereços IP na VPC. Cada sub-rede fica alocada em uma única zona de disponibilidade.

controle de supervisão e aquisição de dados (SCADA)

Na manufatura, um sistema que usa hardware e software para monitorar ativos físicos e operações de produção.

symmetric encryption (criptografia simétrica)

Um algoritmo de criptografia que usa a mesma chave para criptografar e descriptografar dados.

testes sintéticos

Testar um sistema de forma que simule as interações do usuário para detectar possíveis problemas ou monitorar o desempenho. Você pode usar o [Amazon CloudWatch Synthetics](#) para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou diretrizes a um [LLM](#) para direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e estabelecer regras para interações com os usuários.

T

tags

Pares de valores-chave que atuam como metadados para organizar seus recursos. AWS As tags podem ajudar você a gerenciar, identificar, organizar, pesquisar e filtrar recursos. Para obter mais informações, consulte [Marcar seus recursos do AWS](#).

variável-alvo

O valor que você está tentando prever no ML supervisionado. Ela também é conhecida como variável de resultado. Por exemplo, em uma configuração de fabricação, a variável-alvo pode ser um defeito do produto.

lista de tarefas

Uma ferramenta usada para monitorar o progresso por meio de um runbook. Uma lista de tarefas contém uma visão geral do runbook e uma lista de tarefas gerais a serem concluídas. Para cada tarefa geral, ela inclui o tempo estimado necessário, o proprietário e o progresso.

ambiente de teste

Veja o [ambiente](#).

treinamento

O processo de fornecer dados para que seu modelo de ML aprenda. Os dados de treinamento devem conter a resposta correta. O algoritmo de aprendizado descobre padrões nos dados de treinamento que mapeiam os atributos dos dados de entrada no destino (a resposta que você deseja prever). Ele gera um modelo de ML que captura esses padrões. Você pode usar o modelo de ML para obter previsões de novos dados cujo destino você não conhece.

gateway de trânsito

Um hub de trânsito de rede que você pode usar para interconectar sua rede com VPCs a rede local. Para obter mais informações, consulte [O que é um gateway de trânsito](#) na AWS Transit Gateway documentação.

fluxo de trabalho baseado em troncos

Uma abordagem na qual os desenvolvedores criam e testam recursos localmente em uma ramificação de recursos e, em seguida, mesclam essas alterações na ramificação principal. A ramificação principal é então criada para os ambientes de desenvolvimento, pré-produção e produção, sequencialmente.

Acesso confiável

Conceder permissões a um serviço que você especifica para realizar tarefas em sua organização AWS Organizations e em suas contas em seu nome. O serviço confiável cria um perfil vinculado ao serviço em cada conta, quando esse perfil é necessário, para realizar tarefas de gerenciamento para você. Para obter mais informações, consulte [Usando AWS Organizations com outros AWS serviços](#) na AWS Organizations documentação.

tuning (ajustar)

Alterar aspectos do processo de treinamento para melhorar a precisão do modelo de ML. Por exemplo, você pode treinar o modelo de ML gerando um conjunto de rótulos, adicionando rótulos e repetindo essas etapas várias vezes em configurações diferentes para otimizar o modelo.

equipe de duas pizzas

Uma pequena DevOps equipe que você pode alimentar com duas pizzas. Uma equipe de duas pizzas garante a melhor oportunidade possível de colaboração no desenvolvimento de software.

U

incerteza

Um conceito que se refere a informações imprecisas, incompletas ou desconhecidas que podem minar a confiabilidade dos modelos preditivos de ML. Há dois tipos de incertezas: a incerteza epistêmica é causada por dados limitados e incompletos, enquanto a incerteza aleatória é causada pelo ruído e pela aleatoriedade inerentes aos dados. Para obter mais informações, consulte o guia [Como quantificar a incerteza em sistemas de aprendizado profundo](#).

tarefas indiferenciadas

Também conhecido como trabalho pesado, trabalho necessário para criar e operar um aplicativo, mas que não fornece valor direto ao usuário final nem oferece vantagem competitiva. Exemplos de tarefas indiferenciadas incluem aquisição, manutenção e planejamento de capacidade.

ambientes superiores

Veja o [ambiente](#).

V

aspiração

Uma operação de manutenção de banco de dados que envolve limpeza após atualizações incrementais para recuperar armazenamento e melhorar a performance.

controle de versões

Processos e ferramentas que rastreiam mudanças, como alterações no código-fonte em um repositório.

emparelhamento da VPC

Uma conexão entre duas VPCs que permite rotear o tráfego usando endereços IP privados. Para ter mais informações, consulte [O que é emparelhamento de VPC?](#) na documentação da Amazon VPC.

Vulnerabilidade

Uma falha de software ou hardware que compromete a segurança do sistema.

W

cache quente

Um cache de buffer que contém dados atuais e relevantes que são acessados com frequência. A instância do banco de dados pode ler do cache do buffer, o que é mais rápido do que ler da memória principal ou do disco.

dados mornos

Dados acessados raramente. Ao consultar esse tipo de dados, consultas moderadamente lentas geralmente são aceitáveis.

função de janela

Uma função SQL que executa um cálculo em um grupo de linhas que se relacionam de alguma forma com o registro atual. As funções de janela são úteis para processar tarefas, como calcular uma média móvel ou acessar o valor das linhas com base na posição relativa da linha atual.

workload

Uma coleção de códigos e recursos que geram valor empresarial, como uma aplicação voltada para o cliente ou um processo de back-end.

workstreams

Grupos funcionais em um projeto de migração que são responsáveis por um conjunto específico de tarefas. Cada workstream é independente, mas oferece suporte aos outros workstreams do projeto. Por exemplo, o workstream de portfólio é responsável por priorizar aplicações, planejar ondas e coletar metadados de migração. O workstream de portfólio entrega esses ativos ao workstream de migração, que então migra os servidores e as aplicações.

MINHOCA

Veja [escrever uma vez, ler muitas](#).

WQF

Consulte [Estrutura de qualificação AWS da carga de](#) trabalho.

escreva uma vez, leia muitas (WORM)

Um modelo de armazenamento que grava dados uma única vez e evita que os dados sejam excluídos ou modificados. Os usuários autorizados podem ler os dados quantas vezes forem necessárias, mas não podem alterá-los. Essa infraestrutura de armazenamento de dados é considerada [imutável](#).

Z

exploração de dia zero

Um ataque, geralmente malware, que tira proveito de uma vulnerabilidade de [dia zero](#).

vulnerabilidade de dia zero

Uma falha ou vulnerabilidade não mitigada em um sistema de produção. Os agentes de ameaças podem usar esse tipo de vulnerabilidade para atacar o sistema. Os desenvolvedores frequentemente ficam cientes da vulnerabilidade como resultado do ataque.

aviso zero-shot

Fornecer a um [LLM](#) instruções para realizar uma tarefa, mas sem exemplos (fotos) que possam ajudar a orientá-la. O LLM deve usar seu conhecimento pré-treinado para lidar com a tarefa. A

eficácia da solicitação zero depende da complexidade da tarefa e da qualidade da solicitação. Veja também a solicitação [de algumas fotos](#).

aplicação zumbi

Uma aplicação que tem um uso médio de CPU e memória inferior a 5%. Em um projeto de migração, é comum retirar essas aplicações.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.