



Otimize os custos das cargas de trabalho da Microsoft em AWS

AWS Orientação prescritiva



AWS Orientação prescritiva: Otimize os custos das cargas de trabalho da Microsoft em AWS

Copyright © 2024 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 desprestigie 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	1
Como usar este guia	2
Resultados de negócios desejados	3
Jornada de otimização de custos	4
Principais recomendações para otimizar custos	7
Visão geral	7
Principais recomendações	7
AWS Avaliação de otimização e licenciamento	9
Visão geral	9
Opções de avaliação	10
Avaliação completa	10
Escopo de cargas de trabalho	11
Colete dados	11
Análise dados	12
Planeje as próximas etapas	14
Impacto da avaliação	15
Próximas etapas	16
Recursos adicionais	16
Windows no Amazon EC2	17
Automatize os cronogramas de parada e partida	18
Visão geral	18
Estudos de caso	18
Cenário de otimização de custos	19
Recomendações de otimização de custos	21
Recursos adicionais	33
Cargas de trabalho do Windows do tamanho certo	34
Visão geral	34
Cenário de otimização de custos	34
Recomendações de otimização de custos	35
Recomendações	45
Recursos adicionais	45
Selecione o tipo de instância certo para cargas de trabalho do Windows	46

Visão geral	46
Recomendações de otimização de custos	47
Próximas etapas	56
Recursos adicionais	58
Traga licenças para cargas de trabalho do Windows e do SQL Server	58
Visão geral	58
Hosts dedicados do Amazon EC2	59
AWS opções de licenciamento	63
Trazendo licenças do Windows Server	64
Cenários de otimização de custos	65
Recomendações de otimização de custos	72
Recursos adicionais	72
Otimize os gastos com Windows no Amazon EC2	73
Visão geral	73
Conceitos básicos do Savings Plans	73
Cenários de otimização de custos	80
Recomendações de otimização de custos	83
Recursos adicionais	86
Monitore os custos usando AWS ferramentas	86
Visão geral	86
Recomendações de otimização de custos	86
Recursos adicionais	90
SQL Server	91
Escolha uma solução de alta disponibilidade e recuperação de desastres	92
Visão geral	92
Grupos de disponibilidade do SQL Server Always On	93
Instâncias de cluster de failover do SQL Server Always On	95
SIOS DataKeeper	97
Grupos de Disponibilidade Always On	99
Grupos de disponibilidade distribuídos	100
Envio de logs	101
AWS Database Migration Service	103
AWS Elastic Disaster Recovery	104
Comparação de custos	105
Recomendações de otimização de custos	110
Recursos adicionais	111

Entenda o licenciamento do SQL Server	112
Visão geral	112
AWS opções de licenciamento	112
Impacto nos custos de trazer licenças	113
Otimização de licenças	113
Recomendações de otimização de custos	110
Recursos adicionais	45
Selecione a instância EC2 certa para cargas de trabalho do SQL Server	120
Visão geral	120
Comparação de custos	121
Cenário de otimização de custos	123
Recomendações de otimização de custos	124
Recursos adicionais	128
Consolide instâncias	128
Visão geral	128
Cenário de otimização de custos	129
Recomendações de otimização de custos	130
Recursos adicionais	131
Compare as edições do SQL Server	132
Visão geral	132
Impacto nos custos	133
Recomendações de otimização de custos	135
Recursos adicionais	141
Avalie a edição SQL Server Developer	141
Visão geral	141
Impacto nos custos	142
Recursos adicionais	45
Avalie o SQL Server no Linux	145
Visão geral	145
Impacto nos custos	146
Recomendações de otimização de custos	147
Recursos adicionais	148
Otimize as estratégias de backup do SQL Server	149
Visão geral	149
Backup em nível de servidor usando instantâneos habilitados para VSS	149
Backup do SQL Server usando AWS Backup	152

Backup em nível de banco de dados	153
Recomendações de otimização de custos	162
Recursos adicionais	166
Modernize os bancos de dados do SQL Server	167
Visão geral	167
Ofertas de banco de dados	167
Comparação entre Amazon RDS e Aurora	168
Recomendações de otimização de custos	170
Recursos adicionais	175
Otimize o armazenamento para SQL Server	176
Visão geral	176
Tipos, desempenho e custo de armazenamento SSD para o Amazon EBS	177
Otimização geral de custos de SSDs para Amazon EBS	178
Recursos adicionais	180
Otimize o licenciamento do SQL Server usando o Compute Optimizer	181
Visão geral	181
Recomendações de otimização de custos	181
Configurar o Compute Optimizer	182
Recursos adicionais	184
Otimize o dimensionamento do SQL Server usando o Compute Optimizer	184
Visão geral	184
Configurar o Compute Optimizer	185
Recursos adicionais	186
Analise Trusted Advisor as recomendações para cargas de trabalho do SQL Server	186
Visão geral	186
Recomendações de otimização de custos	186
Configurar o Trusted Advisor	187
Recursos adicionais	188
Contêineres	189
Mova aplicativos do Windows para contêineres	190
Visão geral	190
Benefícios de custo	190
Recomendações de otimização de custos	192
Próximas etapas	196
Recursos adicionais	196
Otimize os custos AWS Fargate das tarefas no Amazon ECS	196

Visão geral	196
Benefícios de custo	197
Recomendações de otimização de custos	197
Próximas etapas	204
Recursos adicionais	204
Obtenha visibilidade dos seus custos do Amazon EKS	204
Visão geral	204
Benefícios de custo	204
Recomendações de otimização de custos	205
Próximas etapas	209
Recursos adicionais	209
Reorganize os aplicativos do Windows com o App2Container	209
Visão geral	209
Benefícios de custo	211
Recomendações de otimização de custos	211
Próximas etapas	211
Recursos adicionais	212
Armazenamento	213
Amazon EBS	213
Migre EBS volumes da Amazon de gp2 para gp3	214
Modificar EBS instantâneos da Amazon	218
Excluir volumes Amazon EBS não anexados	221
Amazon FSx	224
Escolha o armazenamento de SMB arquivos certo	225
Habilite a deduplicação de dados na Amazon FSx	230
Entenda a fragmentação de dados no FSx Windows File Server	232
Entenda HDD o uso do volume na Amazon FSx	237
Use uma única zona de disponibilidade	240
AWS Storage Gateway	242
Gateway de arquivos Amazon S3	243
Amazon FSx File Gateway	243
Impacto nos custos	243
Recomendações de otimização de custos	246
Recursos adicionais	248
Active Directory	249
Active Directory autogerenciado no Amazon EC2	249

Visão geral	249
Impacto nos custos	249
Recomendações de otimização de custos	250
Recursos adicionais	255
AWS Managed Microsoft AD	255
Visão geral	255
Impacto nos custos	255
Recomendações de otimização de custos	255
Recursos adicionais	257
AD Connector	257
Visão geral	257
Impacto nos custos	258
Recomendações de otimização de custos	258
Recursos adicionais	259
.NET	260
Refatore para o moderno. NETe mude para o Linux	261
Visão geral	261
Impacto nos custos	261
Recomendações de otimização de custos	262
Considerações e recursos adicionais	263
Contentorizar. NETaplicativos	264
Visão geral	264
Impacto nos custos	264
Recomendações de otimização de custos	266
Recursos adicionais	269
Use instâncias e contêineres do Graviton	269
Visão geral	269
Impacto nos custos	269
Recomendações de otimização de custos	271
Recursos adicionais	272
Support a escalabilidade dinâmica para estática. NETAplicativos de estrutura	273
Visão geral	273
Impacto nos custos	278
Recomendações de otimização de custos	279
Recursos adicionais	281
Use o armazenamento em cache para reduzir a demanda do banco de dados	281

Visão geral	281
Impacto nos custos	281
Recomendações de otimização de custos	282
Recursos adicionais	289
Considere a opção sem servidor. NET	289
Visão geral	289
Impacto nos custos	290
Recomendações de otimização de custos	290
Recursos adicionais	294
Considere bancos de dados criados especificamente	294
Visão geral	294
Impacto nos custos	299
Recomendações de otimização de custos	302
Recursos adicionais	303
Próximas etapas	305
Histórico do documento	306
Glossário	307
#	307
A	308
B	311
C	313
D	316
E	320
F	322
G	324
H	324
I	326
L	328
M	329
O	333
P	336
Q	339
R	339
S	342
T	346
U	347

V	348
W	348
Z	349
.....	cccli

Otimize os custos das cargas de trabalho da Microsoft em AWS

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

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

Visão geral

Este guia fornece recomendações, melhores práticas e estratégias para ajudar a otimizar os custos de suas cargas de trabalho da Microsoft em AWS. O guia também inclui AWS conhecimento básico, técnicas de otimização de custos e arquiteturas de referência para ajudá-lo a criar e automatizar cargas de trabalho econômicas e de alto desempenho que atendam aos seus objetivos de negócios. Coletivamente, essa orientação é chamada de Microsoft on AWS Cost Optimization (MACO). A orientação do MACO foi desenvolvida por especialistas do setor e é baseada em cenários do mundo real.

Este guia aborda as seguintes cargas de trabalho da Microsoft:

- Windows na Amazon Elastic Compute Cloud (Amazon EC2)
- SQL Server
- Contêineres
- Armazenamento
- Active Directory
- .NET

Público

Este guia é destinado a arquitetos, engenheiros, administradores, diretores, CTOs, tomadores de decisões técnicas e AWS parceiros. É útil, mas não necessário, ter experiência prévia e um conhecimento básico de AWS faturamento, tecnologias da Microsoft e administração de AWS sistemas.

Como usar este guia

Você pode usar este guia para planejar e implementar sua jornada de MACO para a nuvem. Recomendamos que você leia este guia do início ao fim para obter uma compreensão abrangente das opções e abordagens para otimizar os custos de suas cargas de trabalho da Microsoft em AWS. Você pode revisar as seguintes seções de carga de trabalho com base nas necessidades da sua organização:

- [Windows no Amazon EC2](#)
- [SQL Server](#)
- [Contêineres](#)
- [Armazenamento](#)
- [Active Directory](#)
- [.NET](#)

Important

Os exemplos de código fornecidos neste guia são apenas para fins de demonstração. É uma prática recomendada testar todo o código em um ambiente de desenvolvimento antes de usá-lo em um ambiente de produção. Antes de implementar qualquer código, recomendamos que você teste seu código em pequenos lotes e, em seguida, revise as alterações de custo resultantes do código usando [AWS Cost Explorer](#). Isso pode ajudá-lo a solucionar casos extremos e outros problemas que podem se tornar problemáticos posteriormente.

Important

Os exemplos de preços neste guia são baseados nos preços no momento da publicação. Os preços estão sujeitos a alterações. Além disso, seus custos podem variar de acordo com suas Região da AWS AWS service (Serviço da AWS) cotas e outros fatores relacionados ao seu ambiente de nuvem.

Resultados de negócios desejados

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

- Saiba como usar uma Avaliação de AWS Otimização e Licenciamento (AWS OLA) para avaliar e otimizar seus ambientes locais e na nuvem atuais, com base na utilização de recursos, licenciamento de terceiros e dependências de aplicativos.
- Desenvolva um caso de negócios para otimização de custos usando a Calculadora de AWS Modernização para cargas de trabalho da Microsoft.
- Otimize os custos para suas cargas de trabalho específicas da Microsoft, incluindo cargas de trabalho para Windows no Amazon Elastic Compute Cloud (Amazon EC2), SQL Server, contêineres, armazenamento, Active Directory e .NET.

Jornada de otimização de custos

O escopo, o tempo e o caminho específico de sua jornada de migração para a nuvem dependem de seus objetivos comerciais, requisitos técnicos e outros fatores. Esta seção fornece um exemplo de uma jornada de migração para a nuvem que se concentra no [gerenciamento financeiro da nuvem AWS](#) e segue as recomendações e as melhores práticas da MACO. Você pode usar esse exemplo para entender como projetar uma jornada de migração para a nuvem para cargas de trabalho da Microsoft.

As seguintes tarefas de alto nível ilustram a abordagem que uma organização pode adotar para implementar as recomendações e as melhores práticas do MACO:

- Estabeleça uma estratégia de marcação e habilite etiquetas de alocação de custos definidas pelo usuário. Para obter mais informações, consulte o AWS whitepaper [Best Practices for AWS tagging Resources](#).
- Defina orçamentos com base em aplicativos, equipes ou departamentos. Para obter mais informações, consulte [Gerenciando seus custos AWS Budgets](#) no AWS Billing and Cost Management User Guide.
- Faça uma Avaliação AWS de Otimização e Licenciamento (AWS OLA) para acelerar a economia. Para obter mais informações, consulte [Avaliação de AWS otimização e licenciamento](#) na AWS documentação.
- Traga sua própria licença (BYOL) para cargas de trabalho do Windows e do SQL Server usando o Amazon Elastic Compute Cloud Dedicated Hosts. Para obter mais informações, consulte a seção [Traga licenças para cargas de trabalho do Windows e do SQL Server](#) desse guia.
- Otimize seu licenciamento do SQL Server em AWS. Para obter mais informações, consulte a seção [Entenda o licenciamento do SQL Server](#) desse guia.
- Selecione o tipo de instância certo para cargas de trabalho do Windows. Para obter mais informações, consulte a seção [Selecione o tipo de instância certo para cargas de trabalho do Windows](#) desse guia.
- Selecione o tipo de instância certo para cargas de trabalho SQL. Para obter mais informações, consulte a seção [Selecione a instância EC2 certa para cargas de trabalho do SQL Server](#) desse guia.
- Migre o Amazon Elastic Block Store (Amazon EBS) de gp2 para gp3. Para obter mais informações, consulte a seção [Migre EBS volumes da Amazon de gp2 para gp3](#) desse guia.

- Controle as cargas de trabalho com o EC2 Instance Scheduler ativado. Para obter mais informações, consulte a seção [Automatize os cronogramas de parada e partida](#) desse guia.
- Elimine os custos do SQL Server para cargas de trabalho que não sejam de produção usando o SQL Server Developer Edition. Para obter mais informações, consulte a seção [Avalie a edição SQL Server Developer](#) desse guia.
- Use uma única zona de disponibilidade para o Amazon FSx for Windows File Server para cargas de trabalho de desenvolvimento e teste. Para obter mais informações, consulte a seção [Use uma única zona de disponibilidade](#) desse guia.
- Dimensione corretamente suas cargas de trabalho do Windows usando o AWS Compute Optimizer. Para obter mais informações, consulte a seção [Cargas de trabalho do Windows do tamanho certo](#) desse guia.
- Otimize os gastos com o Windows no Amazon EC2 usando Savings Plans. Para obter mais informações, consulte a seção [Otimize os gastos com Windows no Amazon EC2](#) desse guia.
- Habilite a deduplicação de dados no FSx for Windows File Server. Para obter mais informações, consulte a seção [Habilite a deduplicação de dados na Amazon FSx](#) desse guia.
- Use fragmentação de dados para sistemas de arquivos no FSx for Windows File Server. Para obter mais informações, consulte a seção [Entenda a fragmentação de dados no FSx Windows File Server](#) desse guia.
- Otimize suas estratégias de backup do SQL Server. Para obter mais informações, consulte a seção [Otimize as estratégias de backup do SQL Server](#) desse guia.
- Faça com que os aplicativos estáticos do .NET Framework suportem escalabilidade dinâmica. Para obter mais informações, consulte este guia. [Support a escalabilidade dinâmica para estática. NETAplicativos de estrutura](#)
- Use microsserviços do .NET sem servidor. Para obter mais informações, consulte a seção [Considere a opção sem servidor. NET](#) desse guia.
- Mova seus aplicativos do Windows para contêineres. Para obter mais informações, consulte a seção [Contêinerizar. NETAplicativos](#) desse guia.
- Use [AWS Compute Optimizer](#) para dimensionar corretamente os contêineres do Windows em execução no Amazon Fargate Amazon Elastic Container Service (Amazon ECS). Para obter mais informações, consulte a seção [Ativar o Compute Optimizer](#) desse guia.
- Refatore para o .NET moderno e migre para o Linux. Para obter mais informações, consulte a seção [Refatore para o moderno. NETe mude para o Linux](#) desse guia.
- Aproveite as instâncias e contêineres do Graviton. Para obter mais informações, consulte a seção [Use instâncias e contêineres do Graviton](#) desse guia.

- Modernize os bancos de dados do SQL Server. Para obter mais informações, consulte a seção [Modernize os bancos de dados do SQL Server](#) desse guia.
- Projete a infraestrutura do Active Directory. Para obter mais informações, consulte a seção [Active Directory](#) desse guia.

Para obter mais informações sobre a jornada do cliente focada no gerenciamento financeiro em nuvem com AWS, consulte o AWS whitepaper sobre o recurso [Cloud Financial Management](#).

Principais recomendações para otimizar custos

Visão geral


A otimização de custos é um dos pilares do [AWS Well-Architected](#) Framework e desempenha um papel fundamental em seus planos de migração para a nuvem. Você encontrará recomendações para otimizações de custos em todo este guia, mas esta seção destaca as recomendações de maior impacto. Você pode implementar essas recomendações rapidamente e elas terão um impacto significativo em sua organização. Essas recomendações podem ajudar a estabelecer as bases para todo o seu esforço de otimização de custos.

Principais recomendações

A tabela a seguir lista as principais recomendações para as otimizações de custo de maior impacto. A coluna “Dificuldade de implementar” classifica cada otimização com base em uma escala do que é mais fácil de implementar (1) até o que é mais difícil de implementar (5). A coluna “Economia estimada” mostra uma estimativa baseada em porcentagem de quanto sua organização pode economizar para cada otimização recomendada.

Otimizações	Dificuldade de implementar	Economia estimada
Cargas de trabalho do Windows do tamanho certo	3	25%
Traga licenças para cargas de trabalho do Windows e do SQL Server	3	30%
Avalie a edição SQL Server Developer	2	20%
Entenda o licenciamento do SQL Server	2	Até 50%
Automatize os cronogramas de parada e partida	3	Até 40%

Otimizações	Dificuldade de implementar	Economia estimada
Selecione o tipo de instância certo para cargas de trabalho do Windows	1	10— 30%
Refatore para o moderno. NETe mude para o Linux	5	10— 20%
Otimize os gastos com Windows no Amazon EC2	3	Até 20— 40%
Migre EBS volumes da Amazon de gp2 para gp3	4	Até 20%

 Important

A economia estimada na tabela anterior se aplica a cada domínio técnico individual, não ao AWS gasto geral em uma conta. Por exemplo, você pode implementar o Agendador de Instâncias em vários tipos e tamanhos de ambiente que podem alterar a economia potencial. As estimativas se aplicam especificamente aos custos de instâncias do Amazon EC2 e não implicam em nenhuma economia geral para outras. Serviços da AWS Essas estimativas são fornecidas como um indicador, não como garantia.

Os especialistas da MACO estão disponíveis para falar mais detalhadamente sobre otimizações de custos. Para marcar uma reunião para se aprofundar em seu caso de uso, entre em contato com a equipe da sua conta ou envie um e-mail [para optimize-microsoft@amazon.com](mailto:para_optimize-microsoft@amazon.com).

AWS Avaliação de otimização e licenciamento

Visão geral

Uma [Avaliação de AWS Otimização e Licenciamento \(AWS OLA\)](#) pode ajudá-lo a avaliar e otimizar seus ambientes de nuvem locais e existentes atuais, com base na utilização de recursos, licenciamento de terceiros e dependências de aplicativos. Você pode usar o AWS OLA para ajudar sua organização a criar uma estratégia de migração e licenciamento que reduza custos à medida que você migra AWS ou avalia as cargas de trabalho existentes da Microsoft. Um AWS OLA também pode ajudar você a alcançar o seguinte:

- Entenda as implantações, o desempenho dos aplicativos e os contratos existentes.
- Dimensione corretamente seus recursos.
- Desenvolva um roteiro para o. Nuvem AWS
- Reduza ou elimine custos usando os investimentos existentes e pagando somente pelo que você usa.

Recomendamos que você faça do AWS OLA o primeiro passo em sua [jornada de otimização de custos](#). Você pode trabalhar com o AWS Partner Network para concluir um AWS OLA. Eles ajudarão você a coletar dados de avaliação e fornecerão recomendações para otimizar seus custos de licenciamento e instância.

O diagrama a seguir fornece uma visão geral do processo de avaliação.



Opções de avaliação

Você pode escolher entre duas opções de AWS OLA para suas cargas de trabalho da Microsoft em AWS:

- **Versão Lite** — Nesse caso de uso, todas as suas cargas de trabalho estão no VMware. Você pode AWS fornecer uma saída do [RVTools](#). Em seguida, AWS pode oferecer um tempo de resposta de 1 a 5 dias. Essa abordagem usa point-in-time informações extraídas diretamente do VMware vCenter para desenvolver recomendações de dimensionamento e oferecer opções de preços sob demanda.
- **Versão completa** — Nesse caso de uso, você tem um ambiente misto em execução em diferentes provedores de nuvem, servidores físicos e servidores virtuais. AWS usa agentes do sistema operacional para coletar dados de uso de 14 a 30 dias. Isso permite AWS tomar decisões informadas sobre o dimensionamento de instâncias com base nos padrões de uso do seu aplicativo. AWS usa várias ferramentas de terceiros, como o Cloudamize, para concluir a análise. AWS trabalha com AWS Partner Network a ITS para ajudar a fornecer a avaliação final do custo total de propriedade (TCO) com várias opções de preços que levam em consideração modelos de preços e diferentes arquiteturas.

Avaliação completa

A avaliação completa do AWS OLA é iniciada com um telefonema de uma hora. Durante essa chamada, AWS ajuda você a determinar a AWS infraestrutura ideal para suportar sua migração, escolher um método de coleta de dados e estabelecer um cronograma para a conclusão. A implementação de ferramentas de descoberta em sua organização depende do método de coleta de dados, do tamanho da sua organização e das ferramentas que sua organização usa para gerenciar sua frota de servidores. Normalmente, são necessárias duas semanas para coletar dados de uso.

O processo completo do AWS OLA leva de 30 a 45 dias e consiste nas seguintes fases:

- Escopo de cargas de trabalho
- Colete dados
- Analise dados
- Planeje as próximas etapas

Escopo de cargas de trabalho

Primeiro, AWS trabalha com você e sua equipe para determinar o escopo da avaliação. Isso geralmente é dividido por tipo de ambiente (por exemplo, não produção e produção). O escopo inclui a localização das cargas de trabalho. Podem ser cargas de trabalho para as quais você está migrando AWS, cargas de trabalho que já estão em execução AWS (por exemplo, AWS OLA para Amazon EC2) ou cargas de trabalho em execução em outros provedores de nuvem.

Colete dados

Em seguida, AWS implanta ferramentas para ajudar na descoberta de recursos e coletar dados de desempenho de seus servidores. Essas ferramentas vêm em quatro opções de implantação:

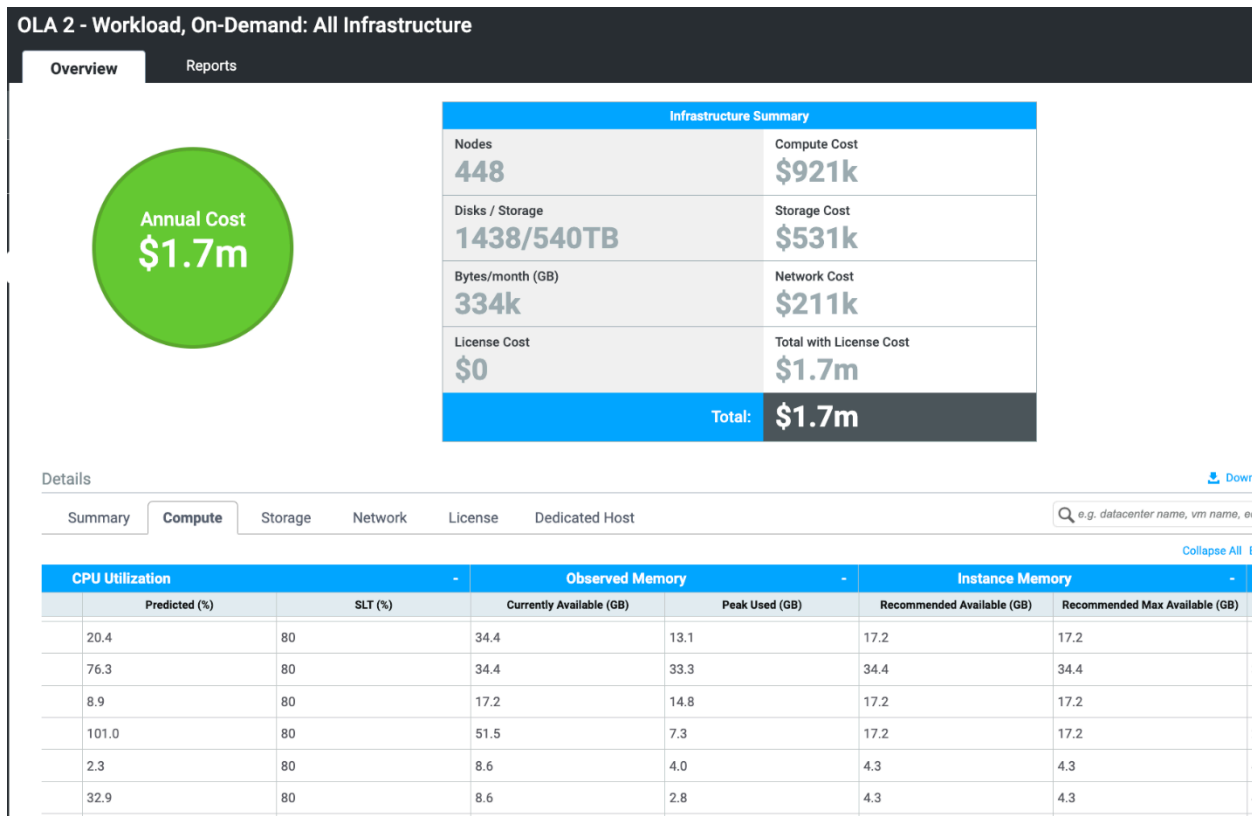
- Ferramentas que podem consultar o hipervisor (requer somente credenciais do VMware vCenter ou Hyper-V)
- Agentes que podem ser implantados em máquinas físicas ou virtuais
- Detecção sem agente usando SSH, Windows Remote Management (WinRM) ou Windows Management Instrumentation (WMI), dependendo do seu ambiente e sistema operacional
- Coleta e análise de dados em arquivos simples

Para a implantação de suas ferramentas, você pode combinar cada opção e consolidar os resultados. É fundamental garantir que qualquer opção escolhida não sobrecarregue seus recursos de TI. AWS se esforça para tornar o processo de avaliação o mais completo possível. Além de um breve telefonema para ajudar na configuração, a equipe do AWS OLA e os arquitetos de soluções especialistas da Microsoft prepararão a análise do custo total de propriedade (TCO) e as recomendações para análise.

A coleta de dados geralmente leva de duas a três semanas quando a utilização da CPU, a utilização da RAM, a taxa de transferência do armazenamento, a IOPS e a taxa de transferência da rede são analisadas. Idealmente, essa coleta ocorre durante os horários de pico do mês comercial (por exemplo, durante relatórios end-of-month financeiros). AWS deseja capturar o pico de uso porque isso fornece boas amostras estatísticas de qual deveria ser a AWS instância do tamanho certo, ao mesmo tempo em que garante que o desempenho possa exceder o que está disponível no local. AWS mescla métricas de utilização com heurísticas de desempenho de várias gerações de processadores para definir exatamente a quantidade de CPU e RAM que uma determinada carga de trabalho exige. Essas metas geralmente são menores do que as alocadas no local. Isso não apenas

reduz o custo computacional de acordo com o tamanho da instância, mas também otimiza os custos de licenciamento.

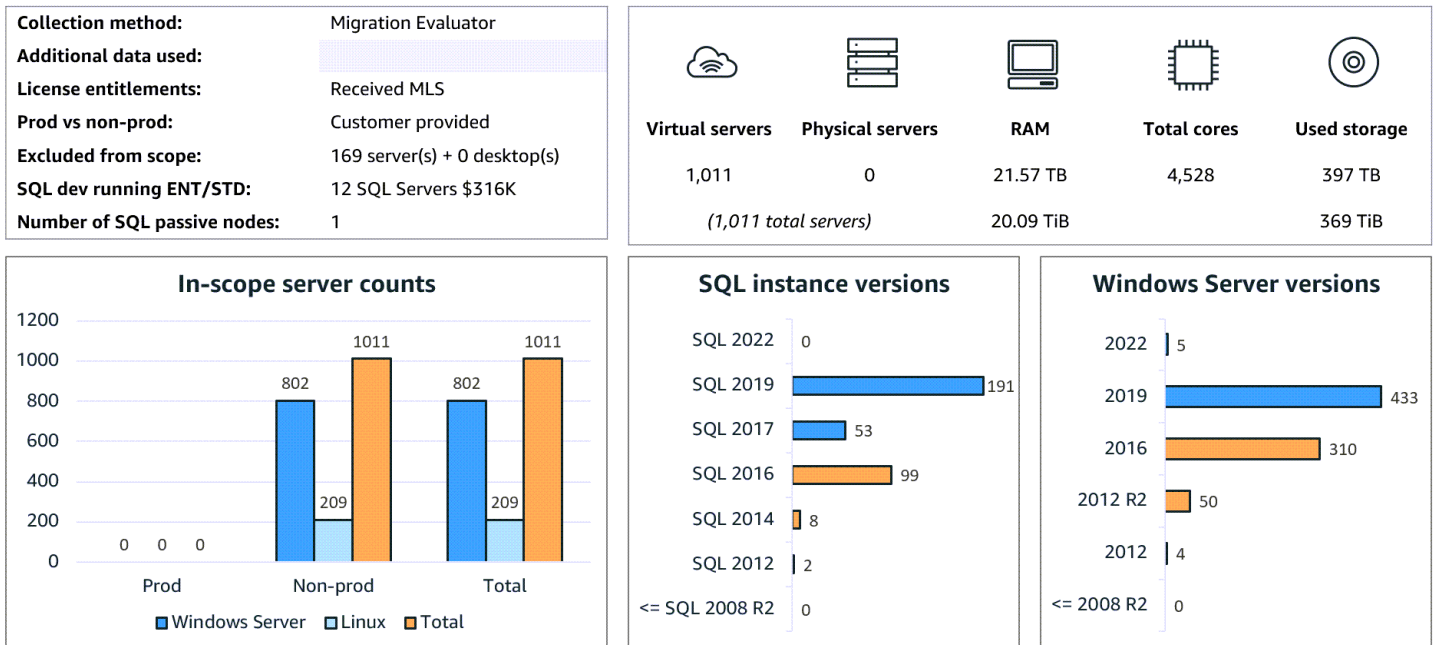
A exibição do painel a seguir mostra um exemplo de custos de infraestrutura que podem ser capturados por uma avaliação.



Analise dados

AWS faz uma apresentação resumida após a conclusão da coleta de dados. AWS analisa os dados, resume as descobertas e, em seguida, faz recomendações para o uso local e a migração para a nuvem. Você pode reduzir os custos de computação e licenciamento examinando oportunidades de consolidação, ganhos de elasticidade (onde as cargas de trabalho podem ser desativadas ou ajustadas sazonalmente), oportunidades de SKU correto (por exemplo, a edição SQL Server Enterprise está em uso, mas os requisitos de recursos e o uso de recursos sugerem que a edição SQL Server Standard é adequada). Para produtos como o SQL Server, que são licenciados pelo núcleo, geralmente faz sentido financeiro colocar as cargas de trabalho em uma instância de computação mais cara. Ou seja, se o perfil da CPU e a proporção de RAM para vCPU tiverem um efeito líquido na redução do número de núcleos licenciados para os casos de uso com licença incluída e Bring Your Own License (BYOL).

Veja a seguir um exemplo de análise com base nos dados coletados pela avaliação.



Os cenários comuns de otimização incluem a identificação de oportunidades AWS de otimização de recursos e economias em licenças de terceiros.

Exemplos de oportunidades de otimização de AWS recursos:

- Evite o excesso de provisionamento para uso máximo.
- Evite especificar demais e subutilizar recursos.
- Dimensione suas instâncias corretamente e migre para as mais novas gerações de instâncias EC2.
- Economize nos custos operacionais migrando para bancos de dados gerenciados.

Exemplos de economia de licenças de terceiros:

- Reduza os núcleos necessários para executar a mesma carga de trabalho.
- Livre-se da edição e dos pacotes complementares desnecessários do SQL Server Enterprise.
- Remova os servidores zumbis e substitua o hardware desatualizado.
- Use as opções BYOL e com licença incluída para reduzir futuros contratos comerciais.
- Modernize-se para soluções de código aberto e nativas em nuvem.

Planeje as próximas etapas

Por fim, AWS usa os dados de desempenho coletados para estimar o tamanho e o custo específicos da carga de trabalho. AWS também pode analisar de forma agregada seu ambiente com escopo e fornecer uma análise quantitativa. Isso pode ajudá-lo a determinar se a melhor opção é uma atualização local ou uma migração para o. AWS Você pode criar um caso de negócios econômico na nuvem usando o resumo da análise de TCO (conforme mostrado no exemplo a seguir) fornecido no final de um AWS OLA.

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

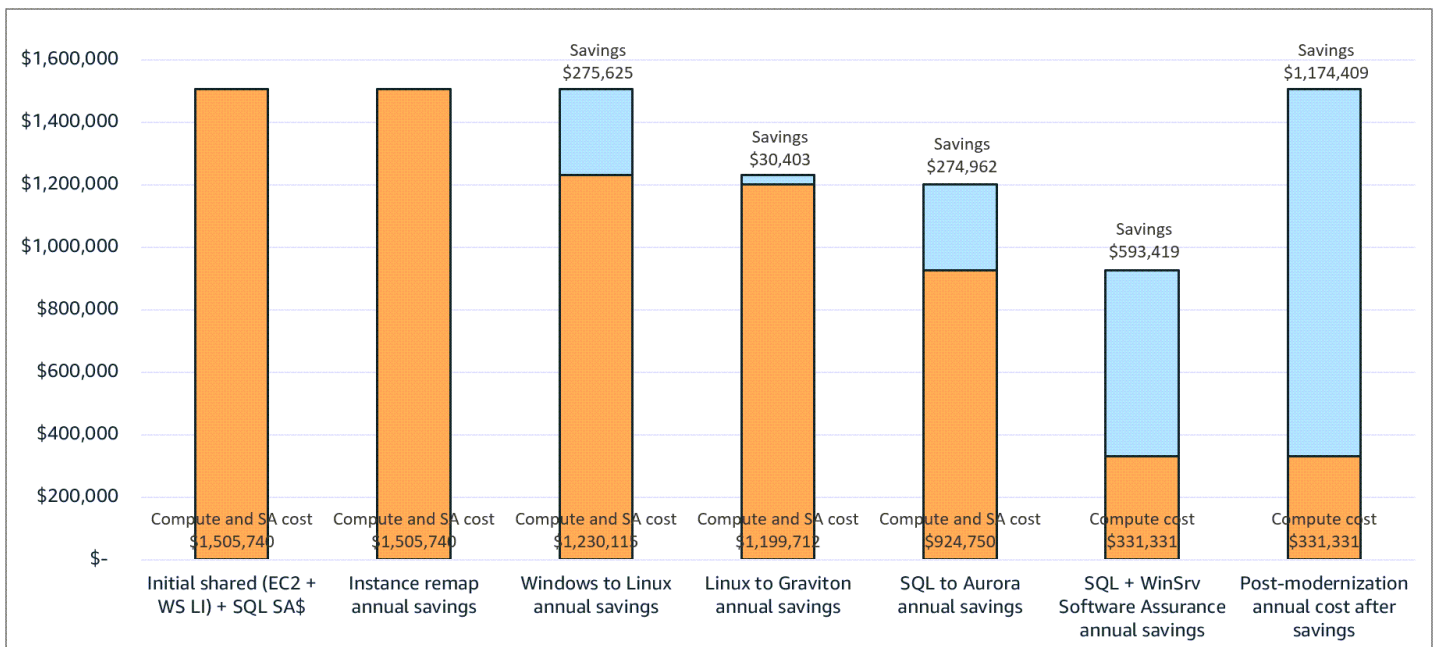
¹ Pricing model used: 3-year, no upfront RI

² Software Assurance and true-up costs provided by Microsoft

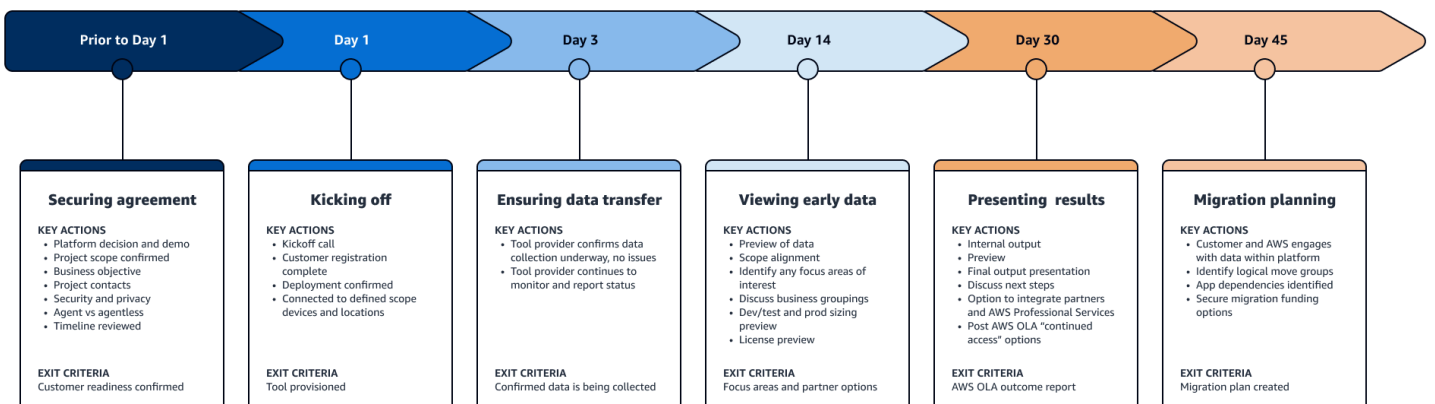
Um AWS OLA também fornece uma visão sobre o impacto que a modernização pode ter em suas cargas de trabalho existentes, fazendo sugestões como as seguintes:

- Mude para um sistema operacional Linux.
- Adicione suporte de aplicativos para processadores ARM (AWS Graviton).
- Mova cargas de trabalho do SQL Server para o Amazon Aurora.
- Elimine a garantia de software transferindo as cargas de trabalho do Windows e do SQL Server para tecnologias de código aberto.

O diagrama a seguir mostra a economia de custos que pode ser obtida por meio de técnicas de modernização, como migrar do Windows para o Linux ou do SQL Server para o Aurora.



O processo completo do AWS OLA leva aproximadamente 45 dias do início ao fim. O diagrama a seguir mostra um exemplo de cronograma.



Se você tiver um ambiente VMware puro e puder fornecer resultados do RVTools, poderá reduzir esse cronograma para uma semana útil. Além disso, AWS pode analisar um arquivo simples que inclui dados de ativos e de utilização, como média de CPU, pico de CPU, média de RAM e pico de RAM.

Impacto da avaliação

O cliente médio normalmente experimenta uma redução de custo de 20 a 30% com o esforço de dimensionamento correto. O dimensionamento correto combina a carga de trabalho de origem com as AWS instâncias de melhor tamanho com base nos dados de uso. Esses ajustes

de dimensionamento correto não apenas reduzem o custo mensal do AWS ambiente, mas frequentemente resultam em economias em outras partes da organização. Por exemplo, um ganho de 20 a 30% no licenciamento do Windows ou do SQL Server pode reduzir a próxima aprovação com a Microsoft ou liberar o licenciamento para aplicativos adicionais. Geralmente, é na consolidação e no dimensionamento correto das cargas de trabalho do SQL Server que os ganhos financeiros mais significativos são obtidos.

AWS pode ajudá-lo a categorizar os sistemas em compartimentos de modernização. Alguns sistemas são antigos e não são financeiramente viáveis de usar, enquanto outros podem ser modernizados em contêineres ou aplicativos sem servidor, onde as economias mais significativas são obtidas. A conversa com sua AWS equipe passa de tópicos generalizados sobre o que a nuvem permite para discussões mais específicas sobre como e por que cargas de trabalho específicas devem ser modernizadas. AWS também ajuda você a explorar possíveis oportunidades de inovação.

Próximas etapas

Se você está iniciando sua jornada de otimização de custos para cargas de trabalho da Microsoft que estão sendo executadas em ambientes locais ou conectados AWS, interaja com sua equipe de AWS contas e solicite um AWS OLA. AWS os membros da equipe podem responder às suas perguntas e ajudá-lo a decidir se um AWS OLA é, em última análise, a escolha certa para você e sua organização. Como alternativa, você pode [solicitar um AWS OLA online](#).

Recursos adicionais

- [AWS Avaliação de otimização e licenciamento](#) (AWS documentação)
- [AWS re:Invent 2022 - Como economizar custos e otimizar as cargas de trabalho da Microsoft em AWS \(ENT205\) \(\)](#) YouTube

Windows no Amazon EC2

[O Amazon Elastic Compute Cloud \(Amazon EC2\) é uma plataforma de computação em nuvem](#)

altamente flexível e escalável, ideal para executar suas cargas de trabalho do Windows. Você pode usar o Amazon EC2 para implantar, gerenciar e escalar suas cargas de trabalho do Windows Server na infraestrutura segura, confiável, altamente disponível e adaptável do. Nuvem AWS Considere os seguintes benefícios principais de executar cargas de trabalho do Windows no Amazon EC2:

- **Escalabilidade** — O Amazon EC2 permite que você escale facilmente suas cargas de trabalho do Windows para acomodar as mudanças nos requisitos. Você pode criar rapidamente novas instâncias do EC2 para lidar com o aumento da demanda e, com a mesma facilidade, encerrar as instâncias quando elas não forem mais necessárias. Você paga apenas pelos recursos que realmente usa.
- **Flexibilidade** — O Windows no Amazon EC2 oferece suporte a uma ampla variedade de tipos de instância que são projetados para atender a vários requisitos de carga de trabalho, desde instâncias de uso geral até instâncias otimizadas para memória ou computação. Essa flexibilidade garante que você possa escolher o melhor tipo de instância para seus aplicativos específicos baseados em Windows, maximizando o desempenho e minimizando os custos.
- **Segurança** — AWS fornece várias camadas de segurança para suas cargas de trabalho do Windows, incluindo firewalls de rede, criptografia de dados e controles de acesso seguro. Isso significa que você pode confiar que seus aplicativos e dados estão protegidos, sem deixar de ter controle total sobre suas configurações e configurações de segurança.
- **Eficiência de custos** — O modelo de pay-as-you-go preços permite que você pague somente pelos recursos que você usa, eliminando a necessidade de investimentos iniciais em hardware e software. Esse modelo também permite que você otimize seus custos, reduza os gastos de capital e aumente a eficiência operacional. É um modelo de preços ideal para empresas de todos os tamanhos.

Esta seção do guia aborda os seguintes tópicos:

- [Automatize os cronogramas de parada e partida](#)
- [Cargas de trabalho do Windows do tamanho certo](#)
- [Selecione o tipo de instância certo para cargas de trabalho do Windows](#)
- [Traga licenças para cargas de trabalho do Windows e do SQL Server](#)
- [Otimize os gastos com Windows no Amazon EC2](#)

- [Monitore os custos usando AWS ferramentas](#)

Automatize os cronogramas de parada e partida

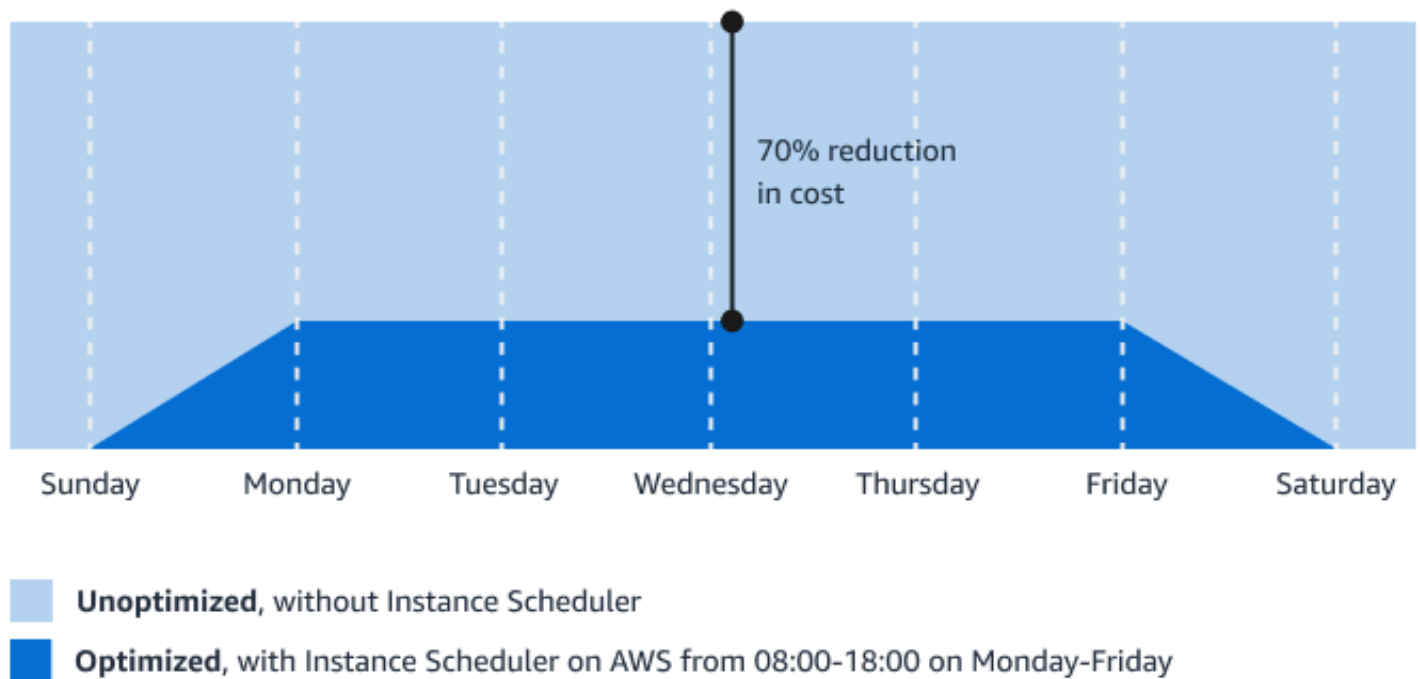
Visão geral

O [Instance Scheduler ativado AWS](#) pode ajudá-lo a reduzir os custos operacionais automatizando o início e a interrupção das instâncias do Amazon EC2 e do [Amazon Relational Database Service](#) (Amazon RDS). Se você deixar todas as suas instâncias funcionando em plena utilização contínua, poderá acabar pagando por recursos que não estão sendo usados. O Agendador de Instâncias ativado AWS permite que você desative as instâncias em momentos em que elas não são necessárias, como fora do horário comercial, fins de semana ou outros períodos em que o uso é baixo. Isso pode levar a uma economia significativa de custos ao longo do tempo.

O Agendador de Instâncias ativado AWS também oferece agendamento de instâncias entre contas, marcação automatizada e a capacidade de configurar horários ou períodos usando uma interface de linha de comando ou a janela de manutenção. [AWS Systems Manager](#) Esses recursos podem ajudar você a gerenciar suas instâncias com mais eficiência e precisão, rastrear e alocar custos em diferentes projetos ou equipes.

Estudos de caso

Considere o exemplo de uma empresa que usa o Instance Scheduler AWS em um ambiente de produção para interromper automaticamente as instâncias fora do horário comercial todos os dias. Se essa empresa deixar todas as suas instâncias funcionando em plena utilização, ela poderá obter uma economia de até 70% nas instâncias que só são necessárias durante o horário comercial normal. O gráfico a seguir mostra como a utilização semanal é reduzida de 168 horas para 50 horas.

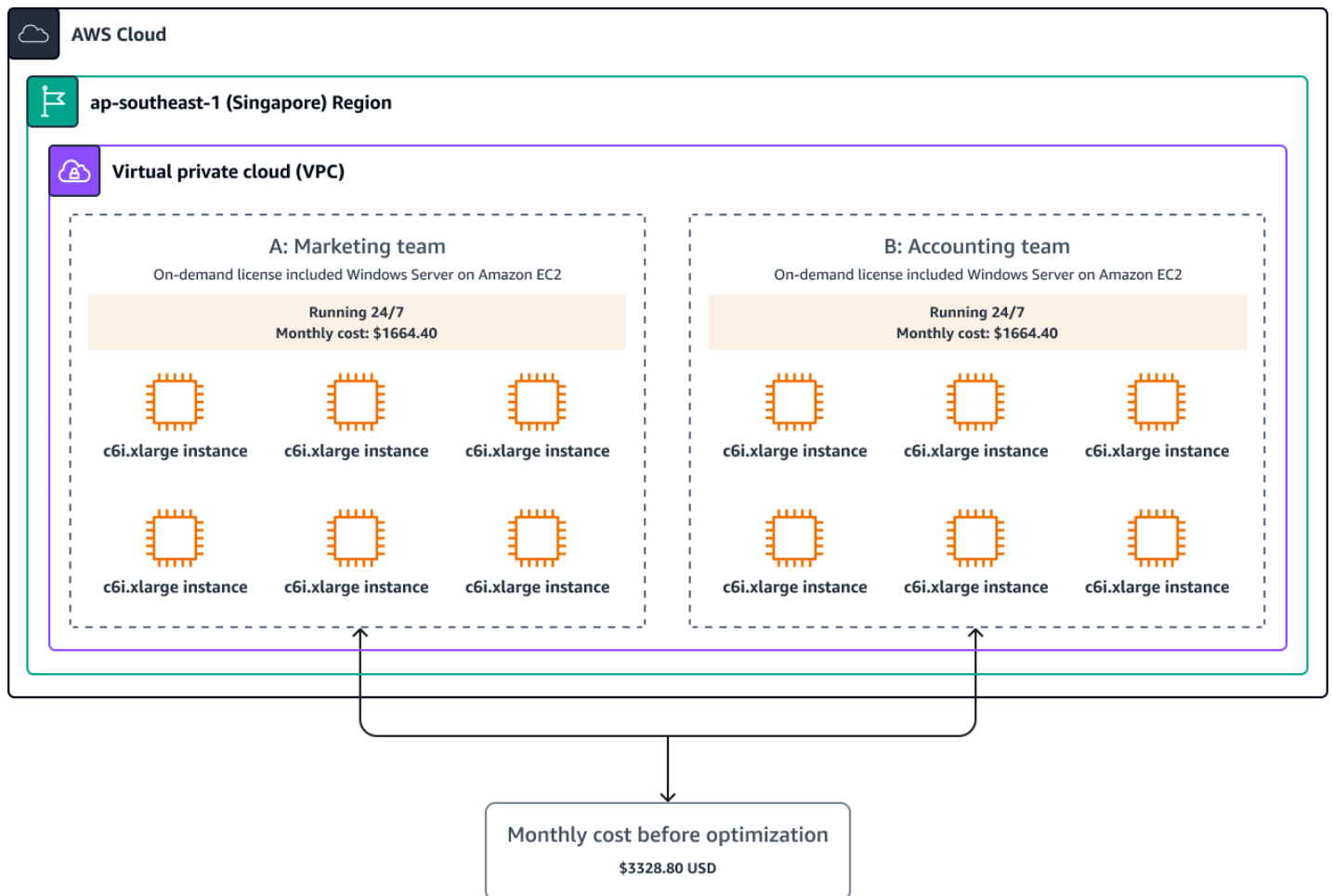


Considere outro exemplo. A concessionária de energia elétrica Jamaica Public Service Company Limited (JPS) migrou seu banco de dados para o Amazon RDS. Agora, a JPS usa o Amazon EC2 para hospedar serviços de API e executar outros aplicativos. Para o JPS, o Instance Scheduler on AWS se tornou a principal ferramenta para gerenciar ambientes que não são de produção. A JPS usou o Instance Scheduler AWS para reduzir os custos de desenvolvimento e gerenciar instâncias do EC2 com base nas necessidades da equipe e nos horários de trabalho. Isso ajudou a JPS a reduzir os custos em 40 por cento. Para obter mais informações, consulte o estudo de AWS caso [O serviço público da Jamaica migra com eficiência para a nuvem e reduz os custos em 40% usando o Agendador de AWS Instâncias](#).

Cenário de otimização de custos

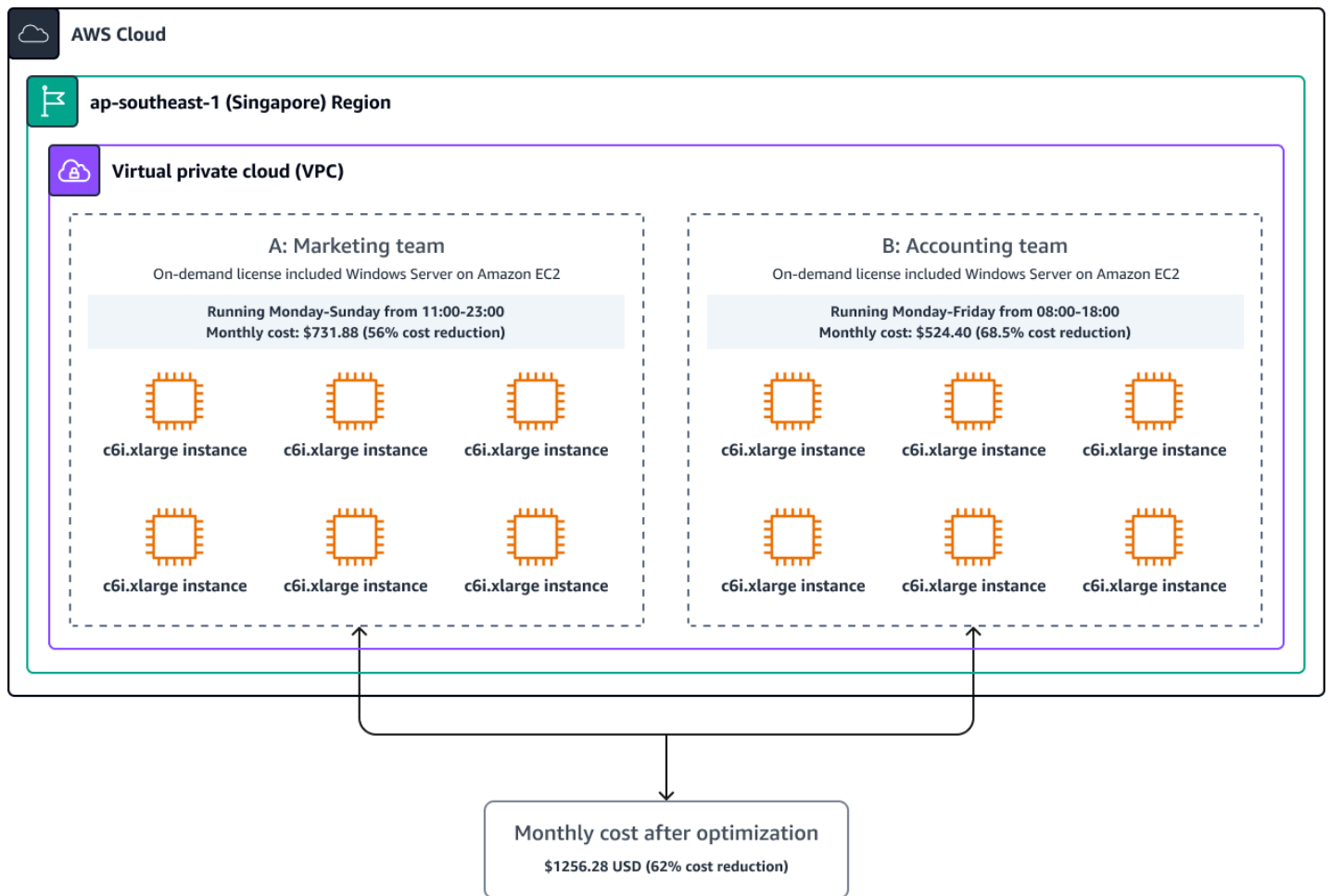
O cenário de exemplo a seguir ajuda a ilustrar as vantagens de custo de usar o Instance Scheduler no. AWS Nesse cenário, uma grande empresa de varejo em Cingapura implanta dois ambientes Windows no Amazon EC2. O primeiro ambiente, conhecido como carga de trabalho A, é utilizado pela equipe de marketing para analisar as transações em tempo real na loja enquanto as lojas estão abertas. O segundo ambiente, conhecido como carga de trabalho B, é reservado para a equipe de contabilidade, que trabalha somente durante o horário comercial normal. O cronograma operacional atual de ambos os ambientes (24 horas por dia, 7 dias por semana) não é ideal, considerando os padrões de uso atuais, e requer otimização para reduzir os custos operacionais da empresa.

O diagrama a seguir mostra o custo mensal antes da otimização.



Por exemplo, há 31 dias no mês de março, dos quais 23 são dias úteis. Se a equipe de marketing usar o Instance Scheduler AWS e operar suas instâncias somente quando necessário (ou seja, por 321 horas por mês em vez de 730 horas por mês), ela poderá economizar 932,52 USD por mês. Isso equivale a uma redução de 56% nos custos operacionais. A equipe de contabilidade também pode experimentar vantagens significativas, com o tempo de uso da instância caindo de 730 horas por mês para 230 horas. Isso resulta em uma redução de \$1.140, ou 68,5%. A empresa poderia economizar um total combinado de \$2.072,52 por mês (equivalente a uma redução de 62%) ou \$24.870,24 por ano.

O diagrama a seguir mostra o custo mensal após a otimização.



Note

O preço desse exemplo foi determinado usando o [AWS Pricing Calculator](#) em março de 2023.

Recomendações de otimização de custos

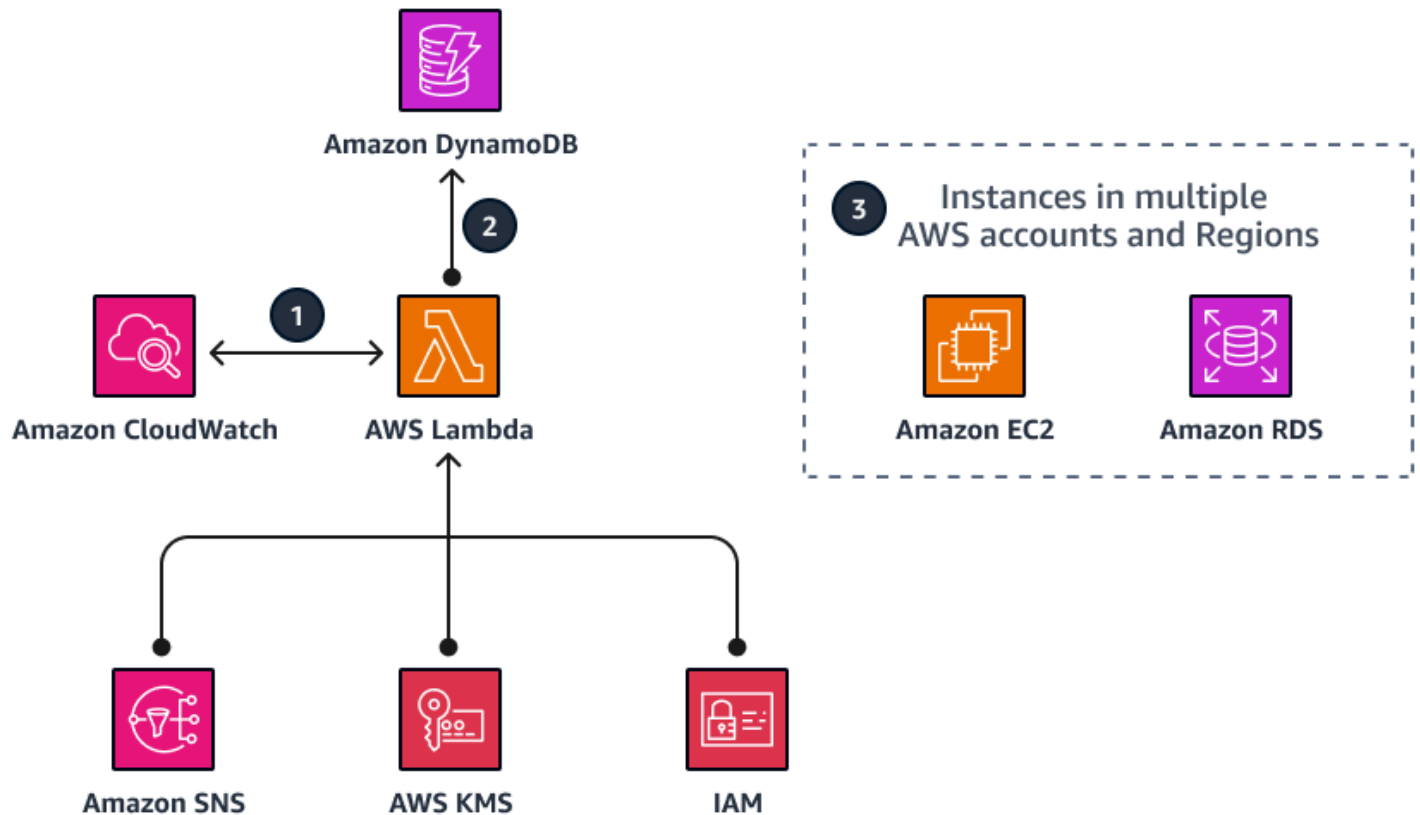
Esta seção explica como implantar e configurar o Agendador de Instâncias AWS com base no cenário de exemplo abordado na seção anterior do cenário de otimização de custos.

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando o Agendador de Instâncias em AWS:

1. Inicie a pilha do Instance Scheduler
2. Configurar períodos
3. Configurar horários

4. Marcar instâncias de

O diagrama de arquitetura a seguir mostra o que é criado na pilha do Instance Scheduler. Nuvem AWS



O diagrama mostra as seguintes etapas do fluxo de trabalho:

1. Um AWS CloudFormation modelo configura um CloudWatch evento da Amazon em um intervalo definido por você. Esse evento invoca uma AWS Lambda função. Durante a configuração, você define Regiões da AWS as contas e. Você também define uma tag personalizada que o Instance Scheduler on AWS usa para associar agendamentos às instâncias aplicáveis do Amazon EC2, instâncias do Amazon RDS e clusters.
2. Os valores de configuração do cronograma são armazenados no Amazon DynamoDB, e a função Lambda os recupera sempre que é executada. Em seguida, você pode aplicar a tag personalizada às instâncias aplicáveis.
3. Durante a configuração inicial do Agendador de Instâncias, você define uma chave de tag para identificar as instâncias aplicáveis do Amazon EC2 e do Amazon RDS. Quando você cria uma agenda, o nome que você especifica é usado como o valor da tag que identifica a agenda que você deseja aplicar ao recurso marcado.

Inicie a pilha do Instance Scheduler

Esta seção mostra como iniciar a CloudFormation pilha do Agendador de Instâncias em AWS.

Note

Você é responsável pelo custo dos Serviços da AWS usado ao executar o Instance Scheduler no AWS. Em janeiro de 2023, o custo de execução dessa solução com configurações padrão na us-east-1 região é de aproximadamente 9,90 USD por mês para cobranças do Lambda, ou menos se você tiver um crédito de uso mensal de nível gratuito do Lambda. Para obter mais informações, consulte a seção Custo do [Agendador de Instâncias no Guia de AWS Implementação](#) na Biblioteca de AWS Soluções.

Para iniciar a pilha do agendador de instâncias, conclua as etapas a seguir.

1. Faça login no [AWS Management Console](#) e escolha [Launch solution](#) (modelo disponível para download) para iniciar o `instance-scheduler-on-aws.template` CloudFormation modelo.

Note

Também é possível [fazer download do modelo](#) para usá-lo como ponto de partida para a sua própria implantação.

2. Por padrão, esse modelo é iniciado na região Leste dos EUA (Norte da Virgínia). Para iniciar o Agendador de Instâncias em uma região diferente, use o seletor de regiões na barra de navegação do console.

Note

Este exemplo usa a região Ásia-Pacífico (Cingapura).

3. Na página Criar pilha, na seção Pré-requisito - Preparar modelo, verifique se a opção Modelo está pronto está selecionada. Na seção Fonte do modelo, verifique se a opção de URL do Amazon S3 está selecionada.
4. Verifique se o URL do modelo correto está na caixa de texto URL do Amazon S3 e, em seguida, escolha Avançar.

5. Na página Especificar detalhes da pilha, insira um nome para a pilha. Para obter informações sobre limitações de nomenclatura de caracteres, consulte [Limites do IAM e do STS](#) na documentação do AWS Identity and Access Management (IAM). O nome da pilha para o exemplo deste guia é chamado `MyInstanceScheduler`.

Note

O nome da pilha não pode conter mais de 28 caracteres.

6. Em Parâmetros, revise os parâmetros do modelo e modifique-os conforme necessário.
7. Escolha Avançar. Na página Configurar opções de pilha, selecione Avançar.
8. Na página Revisar, revise e confirme as configurações. Selecione a caixa confirmando que o modelo criará recursos do IAM.
9. Selecione Criar para implantar a stack.

Configurar períodos

Depois de implantar o CloudFormation modelo, a solução cria uma tabela do DynamoDB que contém exemplos de regras e cronogramas de períodos que você pode usar como referência para criar suas próprias regras e cronogramas de períodos personalizados. Para ver um exemplo de configuração de período, consulte [Exemplos de agendamentos](#) no Agendador de Instâncias na AWS documentação.


Para concluir a etapa desse cenário, você deve gerar períodos que correspondam a cada carga de trabalho e atendam às suas necessidades específicas. Por exemplo: .

```
Period 1 (Workload A):
  Name: retail-hours
  Days: Monday to Sunday
  Hours: 1100 - 2300
Period 2 (Workload B):
  Name: office-hours
  Days: Monday to Friday
  Hours: 0800 - 1800
```

Para configurar períodos, conclua as seguintes etapas:

1. Faça login no console do [DynamoDB](#) e verifique se você está na mesma região em que lançou o modelo para CloudFormation o Instance Scheduler. AWS

2. No painel de navegação, escolha Tabelas e selecione a tabela chamada ConfigTable.
3. Escolha Explorar itens da tabela.
4. Para criar um período para o horário de expediente, selecione o período para o item de horário de expediente.
5. Na página Editar item, altere o valor da hora de início para 0800 e da hora de término para 1800. Mantenha o valor padrão em vigor durante a semana.

 Note

Os valores de início e hora de término determinam quando as instâncias devem ser iniciadas e interrompidas, enquanto o valor de dias da semana determina a quais dias da semana essa programação se aplica (de segunda a sexta-feira, neste exemplo).

6. Escolha Salvar alterações.
7. Para duplicar o período do horário de expediente e usá-lo para criar um novo período para o horário de varejo, selecione o período para o item do horário de expediente. Em seguida, no menu Ações, escolha Duplicar item.
8. Modifique os atributos para atender às suas necessidades. Os atributos a seguir são usados para atender aos requisitos do cenário de exemplo:

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

9. Selecione Create Item (Criar item).
10. No ConfigTableDynamoDB, identifique os dois períodos que você acabou de criar listados nas listas de itens.

Configurar horários

No contexto do Instance Scheduler on AWS, os cronogramas se referem à aplicação de um ou mais períodos e ao fuso horário relevante. Essas programações são então atribuídas às suas instâncias como tags. Esta seção mostra como criar duas agendas (mostradas abaixo) para acomodar os

padrões de tempo variáveis dos dois exemplos de cargas de trabalho e, em seguida, associar as agendas aos períodos que você criou na seção anterior.

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

Para criar e configurar agendamentos, conclua as seguintes etapas:

1. Faça login no console do [DynamoDB](#) e verifique se você está na mesma região em que lançou o modelo para CloudFormation o Instance Scheduler. AWS
2. No painel de navegação, escolha Tabelas e selecione a tabela chamada ConfigTable.
3. Escolha Explorar itens da tabela.
4. Para duplicar a programação do horário comercial do Reino Unido e usá-la para criar uma nova agenda para seu horário de expediente (horário comercial de Cingapura, por exemplo), selecione a programação para o uk-office-hours item. Em seguida, no menu Ações, escolha Duplicar item.
5. Modifique os atributos para atender às suas necessidades. Os atributos a seguir são usados para atender aos requisitos do cenário de exemplo:

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

6. Selecione Create Item (Criar item).
7. Repita as etapas de 4 a 6 para criar uma programação para o horário de varejo de Cingapura usando os seguintes valores de atributos:

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

8. No ConfigTableDynamoDB, identifique as duas programações e os dois períodos que você criou.

Marcar instâncias de

Depois de estabelecer seus agendamentos, você deve usar tags para alocar os agendamentos às instâncias específicas que você deseja usar. Você pode usar o editor de tags interno [AWS Resource Groups](#) para gerar e atribuir tags às suas instâncias do Amazon EC2.

1. Faça login no [AWS Management Console](#) e verifique se você está na mesma região em que lançou o CloudFormation modelo anteriormente.
2. Abra o [console Resource Groups](#). No painel de navegação, expanda Marcação e escolha Editor de tags.
3. Na seção Localizar recursos para marcar, em Regiões, escolha suas Regiões. Para tipos de recursos, escolha Amazon EC2 ou Amazon RDS. Esse cenário se concentra nas instâncias do Amazon EC2 na carga de trabalho A. A equipe de marketing está usando a carga de trabalho A na região de Cingapura. Os recursos dessa carga de trabalho já estão marcados com uma chave de departamento e um valor de marketing. Você pode usar essa tag para pesquisar as instâncias.
4. Escolha Recursos de pesquisa.
5. Selecione as instâncias que você deseja incluir na agenda na lista de resultados da pesquisa e escolha Gerenciar tags dos recursos selecionados.
6. Na seção Editar tags de todos os recursos selecionados, escolha Adicionar tag para adicionar as tags de agendamento do Instance Scheduler às suas instâncias do EC2. Você pode usar as chaves e os valores de tag que correspondem aos cronogramas (criados anteriormente no DynamoDB).
7. Em Chave de tag, adicione Programação. Em Valor da etiqueta, insira singapore-retail-hours.
8. Escolha Revisar e aplicar alterações de tag.
9. Para aplicar a tag a todas as instâncias do EC2 que você selecionou, escolha Aplicar alterações a todas as selecionadas.
10. Repita as etapas de 3 a 9 para qualquer agenda adicional que você queira aplicar.

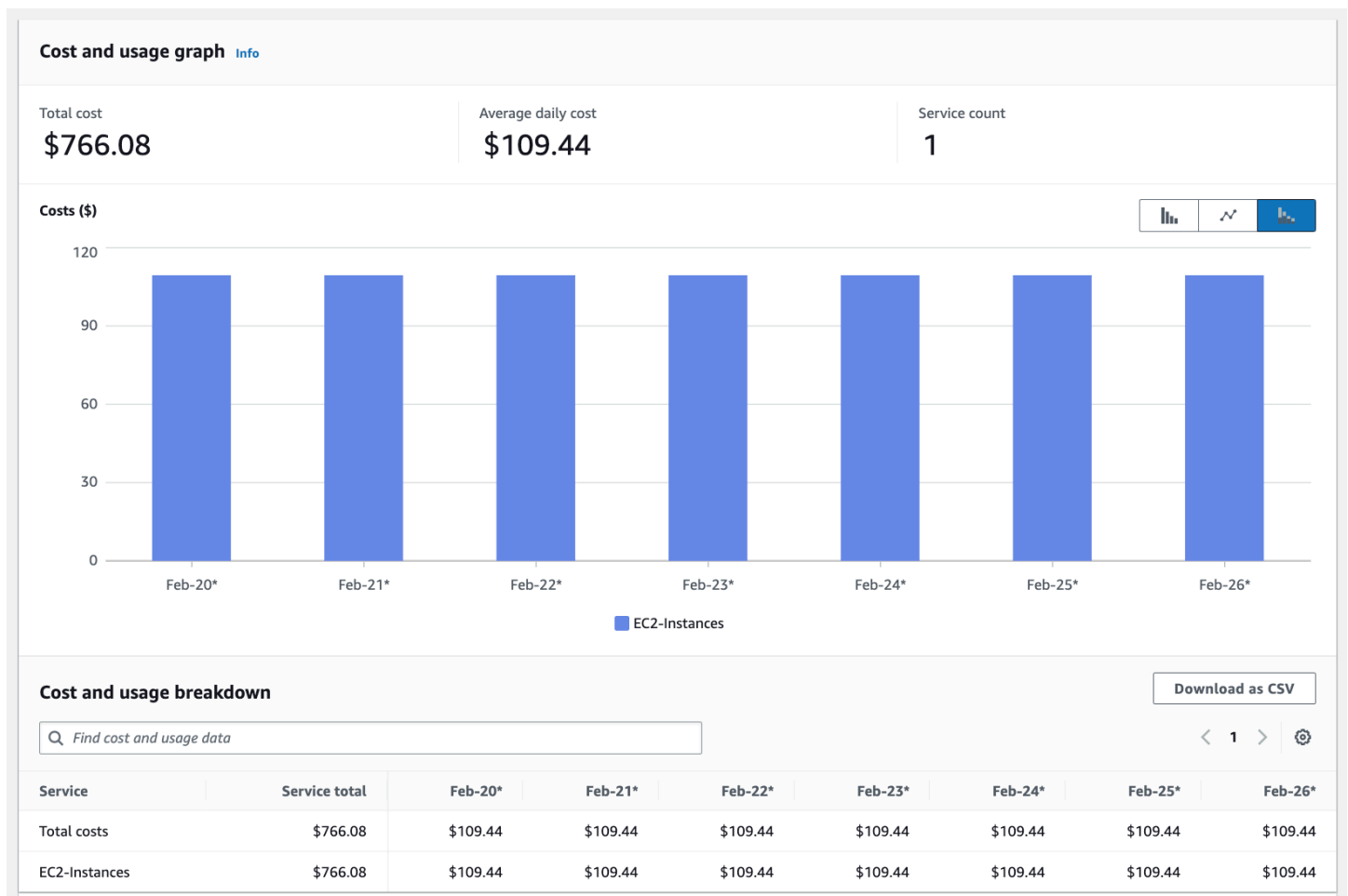
Resultados da validação

Recomendamos que você use [AWS Cost Explorer](#) para medir os benefícios de custo de usar o Instance Scheduler on AWS. Você pode usar o Cost Explorer para fazer o seguinte:

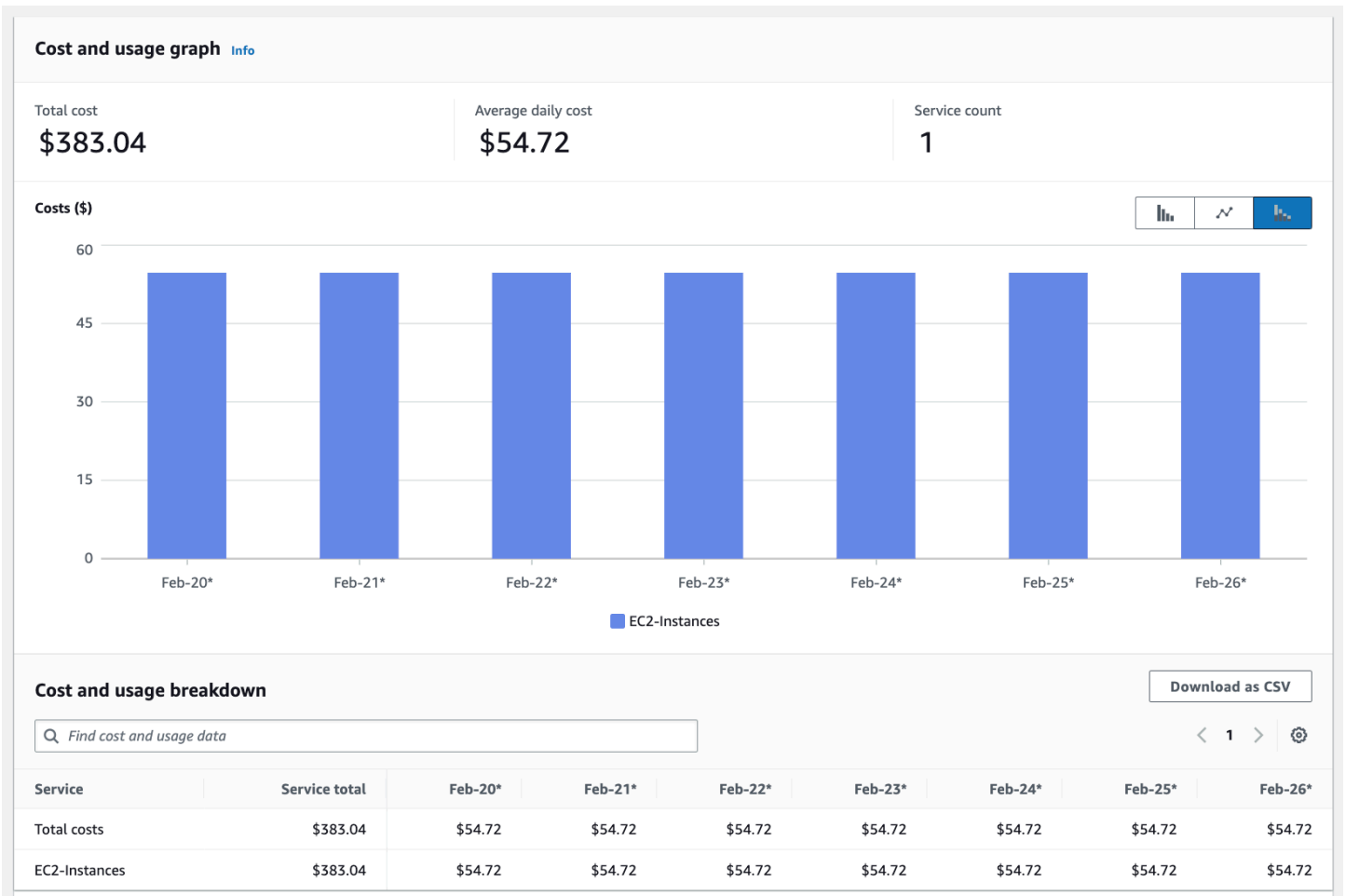
- Visualize e analise os custos associados às suas instâncias do EC2, incluindo instâncias gerenciadas pelo Instance Scheduler.
- Filtre sua visualização do Cost Explorer por tags para que você possa se concentrar em cargas de trabalho específicas e obter uma visão granular da economia de custos obtida com o uso do Instance Scheduler.
- Obtenha insights sobre o impacto financeiro do uso do Instance Scheduler.
- Identifique oportunidades para maior otimização de custos e tome decisões baseadas em dados para otimizar seus AWS gastos.

Os gráficos a seguir ilustram o custo operacional da carga de trabalho A e da carga de trabalho B durante um período de sete dias (de segunda a domingo) antes da otimização usando o Instance Scheduler.

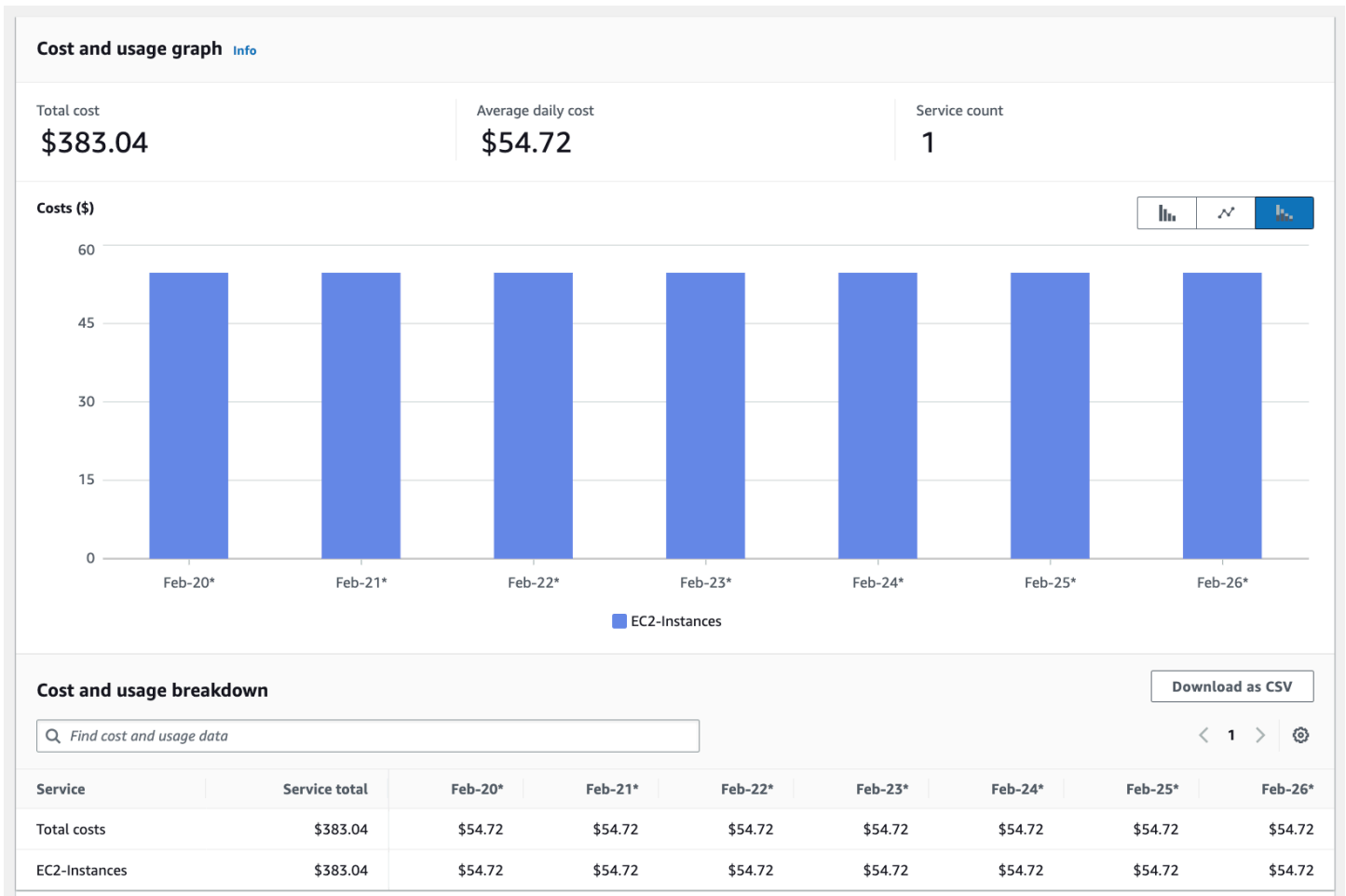
Despesas totais combinadas das cargas de trabalho A e B



Despesas da carga de trabalho A

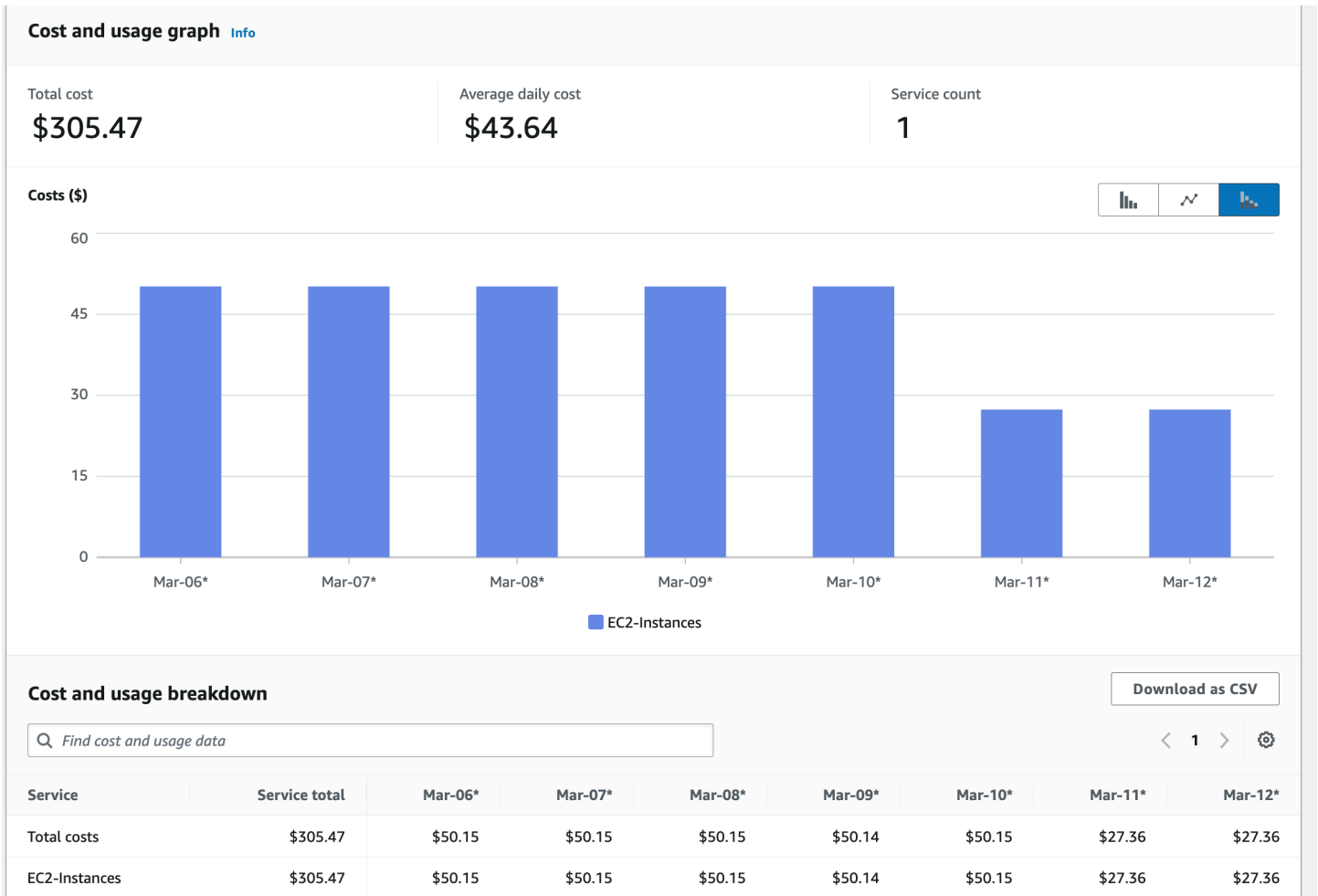


Despesas da carga de trabalho B

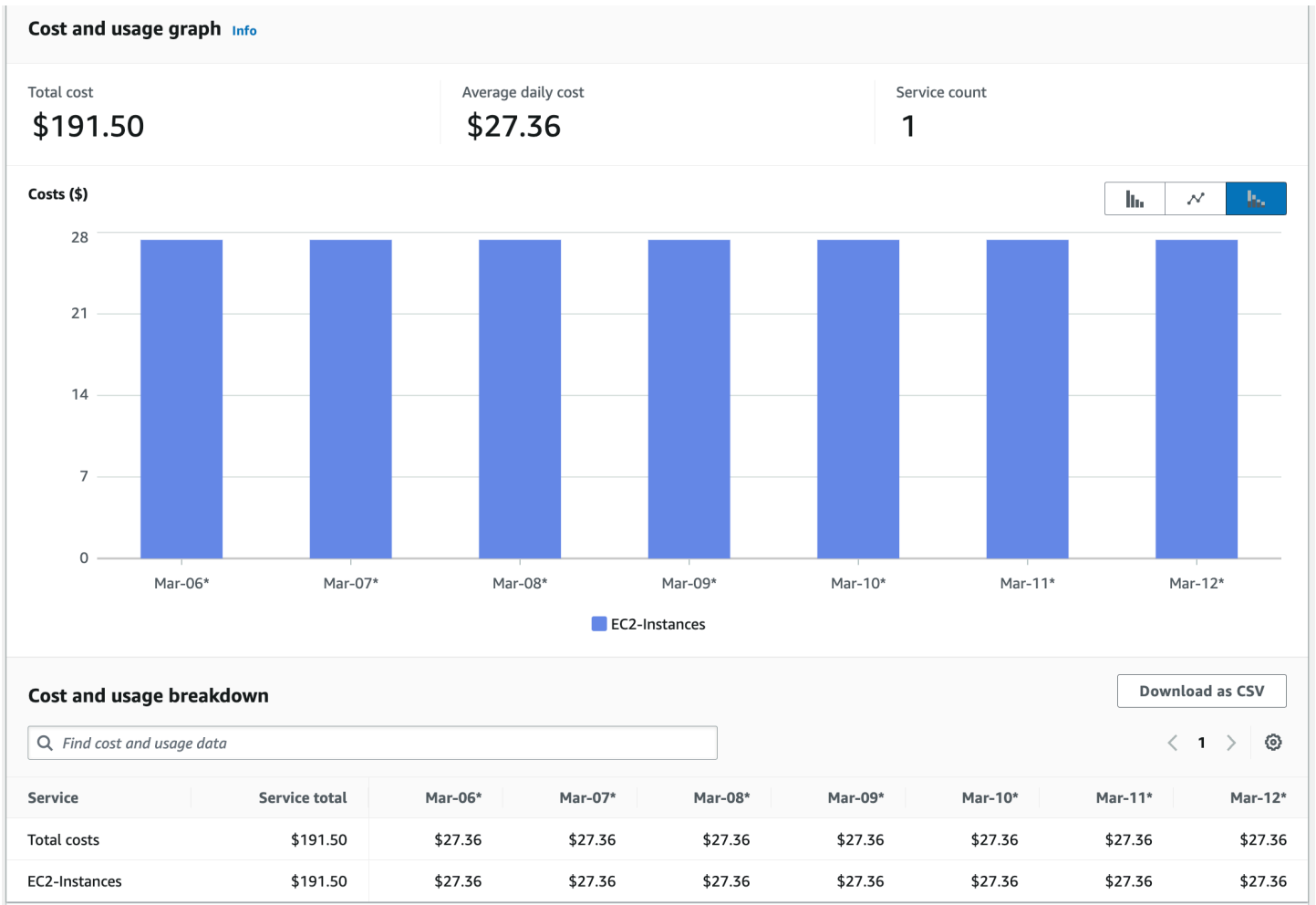


Nesse cenário, o Cost Explorer mostra as reduções de custo resultantes da implementação do Instance Scheduler no. AWS Os gráficos a seguir mostram os custos operacionais da carga de trabalho A e da carga de trabalho B por um período de sete dias (segunda a domingo) após a otimização.

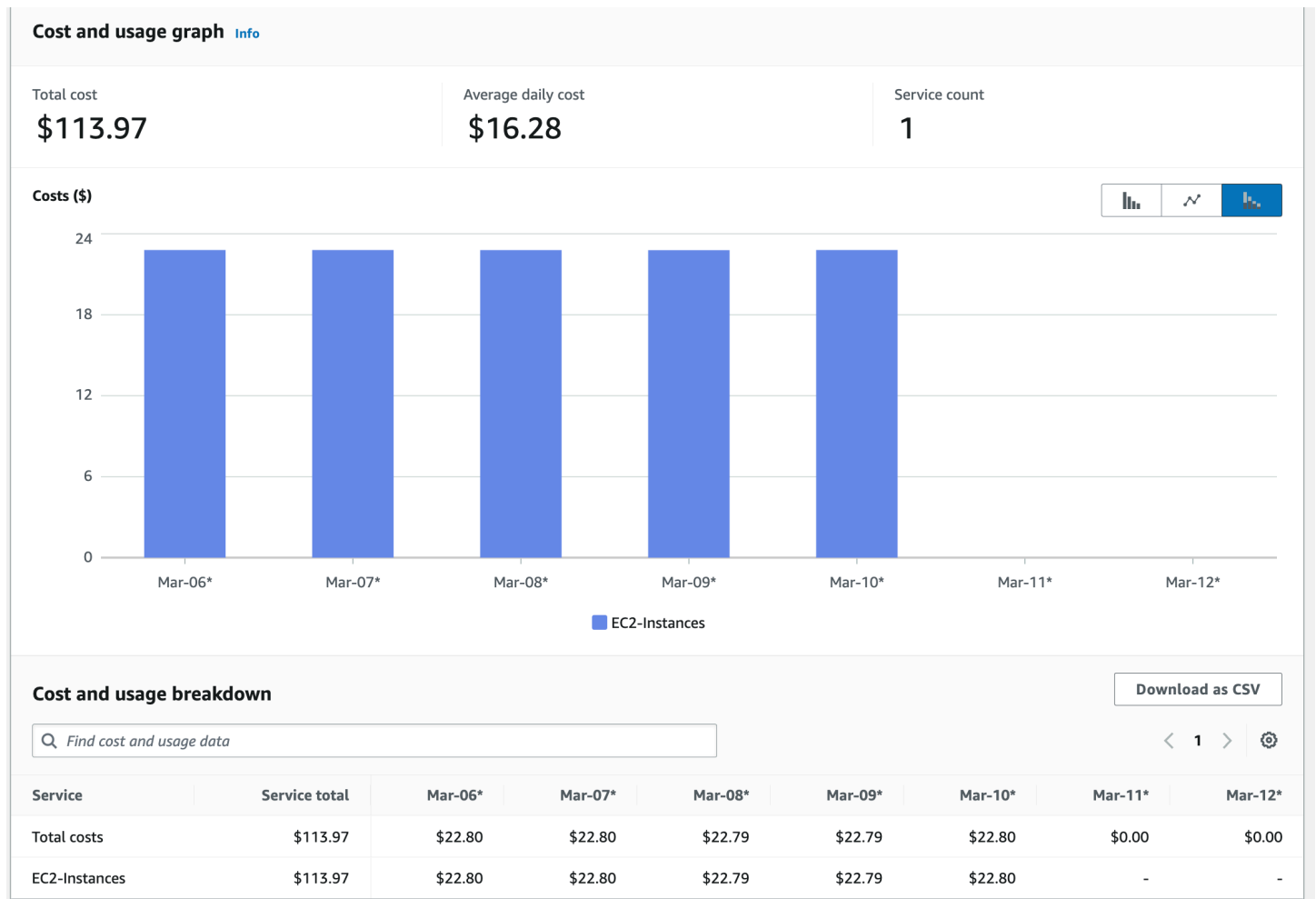
Despesas totais combinadas das cargas de trabalho A e B



Despesas da carga de trabalho A



Despesas da carga de trabalho B



Recursos adicionais

- [Automatize a inicialização e a interrupção de AWS instâncias](#) (Agendador de instâncias na AWS documentação)
- [De volta ao básico: Usando um programador de instâncias para controlar os custos de recursos do Amazon EC2 e do Amazon RDS](#) () YouTube
- [Marcando seus AWS recursos](#) (Guia do usuário de AWS recursos de marcação)
- [Analisando seus custos com AWS Cost Explorer](#) (AWS Billing and Cost Management documentação)

Cargas de trabalho do Windows do tamanho certo

Visão geral

O dimensionamento correto é uma das ferramentas mais potentes para economizar custos. AWS oferece vários métodos para coletar informações de dimensionamento correto, desde a análise de possíveis cargas de trabalho usando uma [Avaliação de AWS Otimização e Licenciamento \(AWS OLA\)](#) até a revisão de cargas de trabalho existentes usando [AWS Cost Explorer](#)

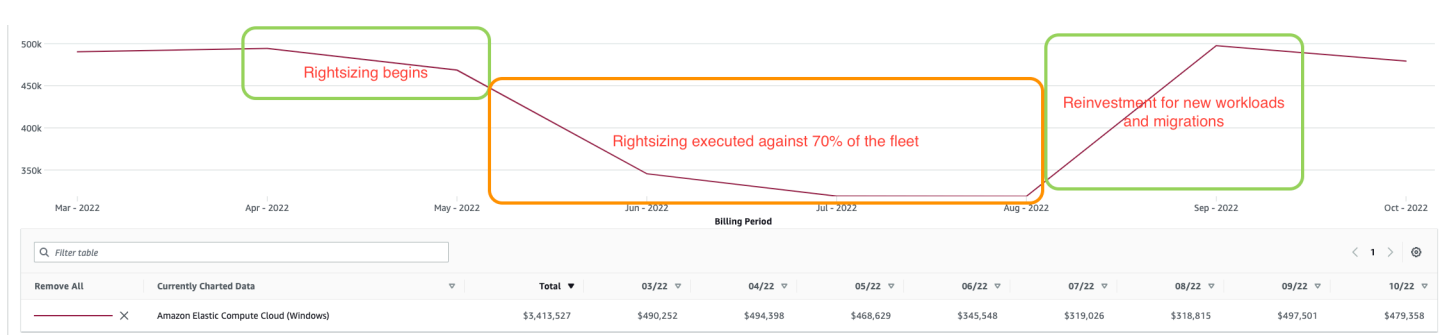
Esta seção mostra como usar [AWS Compute Optimizer](#) para identificar oportunidades de dimensionamento correto do Amazon EC2. O Compute Optimizer ajuda a evitar o provisionamento excessivo e o subprovisionamento dos seguintes tipos de recursos: AWS

- Tipos de instância do [Amazon Elastic Compute Cloud \(Amazon EC2\)](#)
- Volumes do [Amazon Elastic Block Store \(Amazon EBS\)](#)
- Serviços [do Amazon Elastic Container Service \(Amazon ECS\)](#) em AWS Fargate
- [AWS Lambda funções com base nos dados de utilização fornecidos pela Amazon CloudWatch](#)

Cenário de otimização de custos

Medir a eficácia do dimensionamento correto pode ser um desafio, pois os esforços de dimensionamento correto podem ser direcionados a um aplicativo específico, a uma equipe ou a toda a organização. Por exemplo, considere uma organização que migra milhares de instâncias para AWS, com 90% de sua frota composta por cargas de trabalho do Windows. A organização pode empregar o Compute Optimizer para analisar sua frota e descobrir um excesso significativo de provisionamento em suas contas e. Regiões da AWS Em seguida, eles podem usar a [AWS Systems Manager automação](#) para dimensionar corretamente sua frota por meio de várias janelas de manutenção. Como resultado, a organização consegue ajustar o tipo de instância do tamanho certo para 70 por cento de sua frota e obtém uma economia de custos de 35 por cento.

O painel a seguir ilustra a economia alcançada ao longo de vários meses quando esse exemplo de organização implementou estrategicamente as recomendações de dimensionamento corretas do Compute Optimizer. O objetivo deles era operar suas cargas de trabalho existentes da forma mais eficiente possível, a fim de retomar uma migração paralisada de um data center de colocation próximo ao final do contrato.



Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando o Compute Optimizer:

- Ativar o Compute Optimizer
- Habilitar a coleta de métricas de memória para nós do Windows
- Consuma as recomendações do Compute Optimizer
- Marque instâncias para obter o tamanho certo
- Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento
- Implemente as recomendações de dimensionamento corretas com a automação AWS Systems Manager
- Considere métodos alternativos de redimensionamento
- Analise os custos antes e depois no Cost Explorer

Ativar o Compute Optimizer

Você pode ativar o [Compute](#) Optimizer no nível da organização ou de uma única conta em AWS Organizations. A configuração de toda a organização fornece relatórios contínuos para instâncias novas e existentes em toda a sua frota para todas as contas dos membros. Isso permite que o dimensionamento correto seja uma atividade recorrente em vez de uma point-in-time atividade.

Nível da organização

Para a maioria das organizações, a forma mais eficiente de usar o Compute Optimizer é no nível organizacional. Isso fornece visibilidade de várias contas e várias regiões em sua organização e centraliza os dados em uma fonte para análise. Para habilitar isso no nível da organização, faça o seguinte:

1. Faça login na sua [conta de gerenciamento do Organizations](#) com uma função que tenha as [permissões necessárias](#) e opte por se inscrever em todas as contas dessa organização. A organização deve ter [todos os recursos habilitados](#).
2. Depois de ativar a conta de gerenciamento, você pode entrar na conta, ver todas as outras contas de membros e consultar suas recomendações.

Note

É uma prática recomendada configurar uma [conta de administrador delegado](#) para o Compute Optimizer. Isso permite que você exerça o princípio do menor privilégio. Dessa forma, você pode minimizar o acesso à conta de gerenciamento da organização e, ao mesmo tempo, fornecer acesso ao serviço de toda a organização.

Nível de conta única

Se você tem como alvo uma conta com altos custos, mas não tem acesso a ela AWS Organizations, você ainda pode ativar o Compute Optimizer para essa conta e região. Para saber mais sobre o processo de aceitação, consulte [Introdução AWS Compute Optimizer](#) na documentação do Compute Optimizer.

Habilitar a coleta de métricas de memória para nós do Windows

As métricas de memória fornecem ao Compute Optimizer as métricas essenciais necessárias para fazer recomendações bem informadas de dimensionamento correto em sua organização. Isso se deve à análise de CPU, memória, rede e armazenamento que está sendo conduzida antes de oferecer uma recomendação.

Para passar métricas de memória das instâncias do Windows EC2 para o Compute Optimizer, você deve habilitar CloudWatch o agente e configurar as métricas de memória a serem coletadas a cada 60 segundos. Não há custo adicional para usar métricas de memória com CloudWatch.

Ative o CloudWatch agente e configure as métricas de memória

Baixe o [ComputeOptimizearquivo.yml](#). Você pode usar esse arquivo para ativar a coleta de memória para todas as instâncias da sua conta. O arquivo de modelo gera os seguintes componentes:

- [AWS Systems Manager Armazenamento de parâmetros](#) — armazena a configuração do CloudWatch agente necessária para coletar métricas de memória.

- AWS Identity and Access Management Função (IAM) com [políticas AWSAWS Systems Manager gerenciadas](#) anexadas — Isso é para o documento Systems Manager Automation.
- [AWS Systems Manager documentos](#) — Isso instala e configura o CloudWatch agente (substituindo qualquer CloudWatch configuração existente).
- AWS Systems Manager Associação [do State Manager](#) — Isso permite que os documentos do Systems Manager sejam executados em todas as instâncias da sua conta.

⚠ Important

A execução desse modelo substitui qualquer CloudWatch configuração existente nas instâncias.

Depois, faça o seguinte:

1. Faça login no AWS Management Console e abra o [CloudFormation console](#).
2. No painel de navegação, escolha Pilhas.
3. escolha Criar pilha e, em seguida, escolha Com recursos existentes (importar recursos).
4. Escolha Próximo.
5. Em Recurso de modelo, selecione Carregar um arquivo de modelo.
6. Escolha o arquivo e, em seguida, faça o upload do `ComputeOptimize.yml` arquivo.
7. Escolha Próximo.
8. Na página Especificar detalhes da pilha, em Nome da pilha, insira um nome para sua pilha e escolha Avançar.
9. Na página Identificar recursos, insira os valores do identificador dos recursos que você está importando.
- 10 Escolha Importar recursos.
- 11 Depois que a pilha for implantada, escolha a guia Saídas para encontrar a chave, o valor e a descrição da sua associação.

Monitore o progresso da associação

1. Depois que a implantação da CloudFormation pilha for concluída, abra o [console do Systems Manager](#).

2. No painel de navegação, na seção Gerenciamento de nós, escolha State Manager.
3. Na página Associações, escolha o ID de associação da sua associação.
4. Escolha a guia Execution history (Histórico de execução).
5. Na coluna ID de execução, escolha a ID de execução da sua associação. O status deve ser Sucesso.

Veja as métricas em CloudWatch

Recomendamos que você espere pelo menos cinco minutos para que as métricas sejam preenchidas. CloudWatch

1. Abra o [console de CloudWatch](#).
2. No painel de navegação, expanda a seção Métricas e escolha Todas as métricas.
3. Confirme se as métricas aparecem no namespace CWAgent.

Note

Para aplicar as configurações a qualquer nova instância, execute novamente a associação.

Consuma as recomendações do Compute Optimizer


Considere um exemplo que se concentra em fazer as alterações de tamanho corretas em uma única conta e região. Neste exemplo, o Compute Optimizer está ativado no nível da organização em todas as contas. Lembre-se de que o dimensionamento correto é um processo disruptivo que, na maioria dos casos, é executado com precisão pelos proprietários do aplicativo durante uma janela de manutenção programada de várias semanas.

Se você navegar até o Compute Optimizer de dentro da conta de gerenciamento de uma organização (conforme mostrado nas etapas a seguir), poderá escolher a conta que deseja investigar. Neste exemplo, há seis instâncias em execução em uma única conta na us-east-1 região. Todas as seis instâncias estão superprovisionadas. O objetivo é redimensionar as instâncias com base nas recomendações do Compute Optimizer.

Identifique instâncias superprovisionadas e exporte os detalhes das recomendações

1. Faça login AWS Management Console e abra o console do [Compute Optimizer](#).

2. No painel de navegação, escolha Painel.
3. Na caixa de pesquisa na página do Painel, insira region=Leste dos EUA (Norte da Virgínia). Em seguida, insira Findings=overprovisioned. Esses filtros permitem que você veja todas as instâncias superprovisionadas na região. us-east-1
4. Para analisar as recomendações detalhadas para instâncias EC2 superprovisionadas, role para baixo até o cartão de instâncias EC2 e escolha Exibir recomendações.
5. Escolha Exportar e salve o arquivo para uso futuro.
6. Para o bucket do S3, insira o nome do bucket do Amazon S3 que você deseja que seja o destino do arquivo de exportação.

 Note

Para salvar recomendações para análise futura, você deve ter um bucket S3 disponível para gravação no Compute Optimizer em cada região. Para obter mais informações, consulte a [política de bucket do Amazon S3 AWS Compute Optimizer](#) na documentação do Compute Optimizer.

7. Na seção Filtros de exportação, marque a caixa de seleção Incluir recomendações para todas as contas de membros na organização.
8. Em Tipo de recurso, escolha instâncias EC2.
9. Na seção Colunas a serem incluídas, marque a caixa de seleção Selecionar tudo.
10. Escolha Exportar.

Escolha instâncias com base nas recomendações

As recomendações de instância são baseadas nas métricas de desempenho coletadas e analisadas pelo Compute Optimizer. É essencial estar ciente das cargas de trabalho em execução na instância para garantir que você escolha a melhor instância. [Este exemplo pressupõe que você possa escolher entre a última geração de instâncias R6i, R5 e T3 do Amazon EC2.](#) As instâncias T3 têm capacidade de intermitência e têm menos recursos de largura de banda de rede. As instâncias R5 e R6 têm o mesmo custo por hora e são quase idênticas. No entanto, a instância R6 tem uma maior capacidade de largura de banda de rede, apresenta a última geração de processadores Intel e oferece o mesmo espaço computacional que o R5. Neste exemplo, R6 é a melhor opção para redimensionar.

1. No console do [Compute Optimizer](#), escolha Recomendações para instâncias do EC2 na barra de navegação. Esta página mostra uma comparação do tipo de instância atual com as opções recomendadas para substituí-la.
2. Para obter o ID da instância que você deseja dimensionar corretamente, abra o [console do Amazon S3](#) a partir da conta de gerenciamento em. AWS Organizations
3. No painel de navegação, escolha Buckets e, em seguida, escolha o bucket que você está usando para armazenar seus resultados exportados.
4. Na guia Objetos, selecione seu arquivo de exportação na lista de objetos e escolha Baixar.
5. Para extrair as informações da instância do arquivo, você pode usar o botão Texto em colunas na guia Dados no Microsoft Excel.

Note

Os IDs de instância são representados como Amazon Resource Names (ARNs). Certifique-se de definir o delimitador como “/” e extrair o ID da instância. Como alternativa, você pode escrever um script ou usar um ambiente de desenvolvimento integrado (IDE) para cortar o ARN.

6. No Excel, filtre a coluna de descoberta para exibir somente as instâncias de OVER_PROVISIONED. Essas são as instâncias que você está almejando para o tamanho certo.
7. Salve os IDs da instância em um editor de texto para facilitar o acesso posterior.

Marque instâncias para obter o tamanho certo

Marcar suas cargas de trabalho é uma ferramenta poderosa para organizar seus recursos em. AWS As etiquetas permitem que você obtenha uma visibilidade refinada dos custos e facilitem o estorno. Para obter mais informações sobre estratégias e métodos para adicionar tags aos AWS recursos, consulte o AWS whitepaper [Best Practices for AWS Tagging](#) Resources. Neste exemplo, você pode usar o [Editor de AWS tags](#) para fazer ajustes nas instâncias superprovisionadas que você deseja redimensionar durante uma janela de manutenção. Você também pode usar essa tag para visualizar os custos antes e depois da alteração.

1. Faça login AWS Management Console e abra o [AWS Resource Groups console](#) da conta que contém as instâncias destinadas ao redimensionamento.
2. Na barra de navegação, na seção Marcação, escolha Editor de tags.
3. Para Regiões, selecione sua região de destino.

4. Para Tipos de recursos, escolha `AWS::EC2::Instance`.
5. Escolha Recursos de pesquisa.
6. Na página de resultados da pesquisa de recursos, selecione todas as instâncias que você deseja dimensionar corretamente e escolha Gerenciar tags dos recursos selecionados.
7. Escolha Adicionar Tag.
8. Em Chave de tag, insira Rightsizing. Em Valor do Tag, insira ativado. Em seguida, escolha Revisar e aplicar alterações de tag.

Note

Você pode incluir metadados adicionais, como Team ou Business Unit, para ajudar na filtragem posterior no Cost Explorer.

Depois de criar e aplicar tags definidas pelo usuário aos seus recursos, pode levar até 24 horas para que as tags apareçam na sua página de tags de alocação de custos para ativação. Depois de selecionar suas tags para ativação, pode levar mais 24 horas para que elas se tornem ativas.

Para usuários avançados, você pode usar [AWS CloudShell](#) dentro da conta e da região de destino para marcar várias instâncias. Por exemplo: .

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

Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento

Recomendamos ativar a tag de alocação de custos definida pelo usuário. Isso permite que a tag Rightsizing seja reconhecida e filtrável nas ferramentas de AWS cobrança (por exemplo, Cost

Explorer e). AWS Cost and Usage Report Se você não habilitar isso, a opção de filtragem de tags e os dados não estarão disponíveis. Para obter informações sobre o uso de tags de alocação de custos, consulte [Ativação de tags de alocação de custos definidas pelo usuário](#) na documentação. AWS Billing and Cost Management

1. Faça login no AWS Management Console e abra o [AWS Billing console](#).
2. No painel de navegação, na seção Faturamento, escolha Tags de alocação de custos.
3. Na guia Tags de alocação de custos definidas pelo usuário, insira Rightsizing.
4. Selecione a chave da tag Rightsizing e, em seguida, escolha Ativar.

Depois de 24 horas, a tag deve aparecer no Cost Explorer.

Implemente recomendações de dimensionamento corretas com o Systems Manager Automation

O redimensionamento é um cenário que exige que uma instância seja interrompida e iniciada. Nesse cenário, talvez seja necessário lidar com essa interrupção em uma janela de manutenção e precisar que equipes diferentes lidem com seu próprio redimensionamento. Antes de alterar um tipo de instância, revise as [considerações sobre tipos de instância compatíveis](#) na documentação do Amazon EC2.

As etapas de exemplo nesta seção implementam recomendações de dimensionamento correto por conta e região usando um documento do Systems Manager Automation chamado [AWS-ResizeInstance](#). Essa abordagem é típica da maioria das organizações, pois a maioria das organizações exige tipos de instância diferentes para finalidades diferentes. Você também pode usar o mesmo documento de AWS-ResizeInstance automação para direcionar implantações de uma ou várias contas.

1. Faça login AWS Management Console e abra o [console do Systems Manager](#).
2. No painel de navegação, na seção Recursos compartilhados, escolha Documentos.
3. Na barra de pesquisa, insira AWS- ResizeInstance e escolha AWS- nos resultados ResizeInstance da pesquisa.
4. Escolha Execute automation.
5. Na página Execute automation runbook, escolha Execução simples.
6. Na seção Parâmetros de entrada, insira InstanceIdInstanceType. Mantenha o restante dos valores padrão.

7. Escolha Executar e aguarde até que a automação siga as etapas para alterar o tipo de instância.

Considere métodos alternativos de redimensionamento

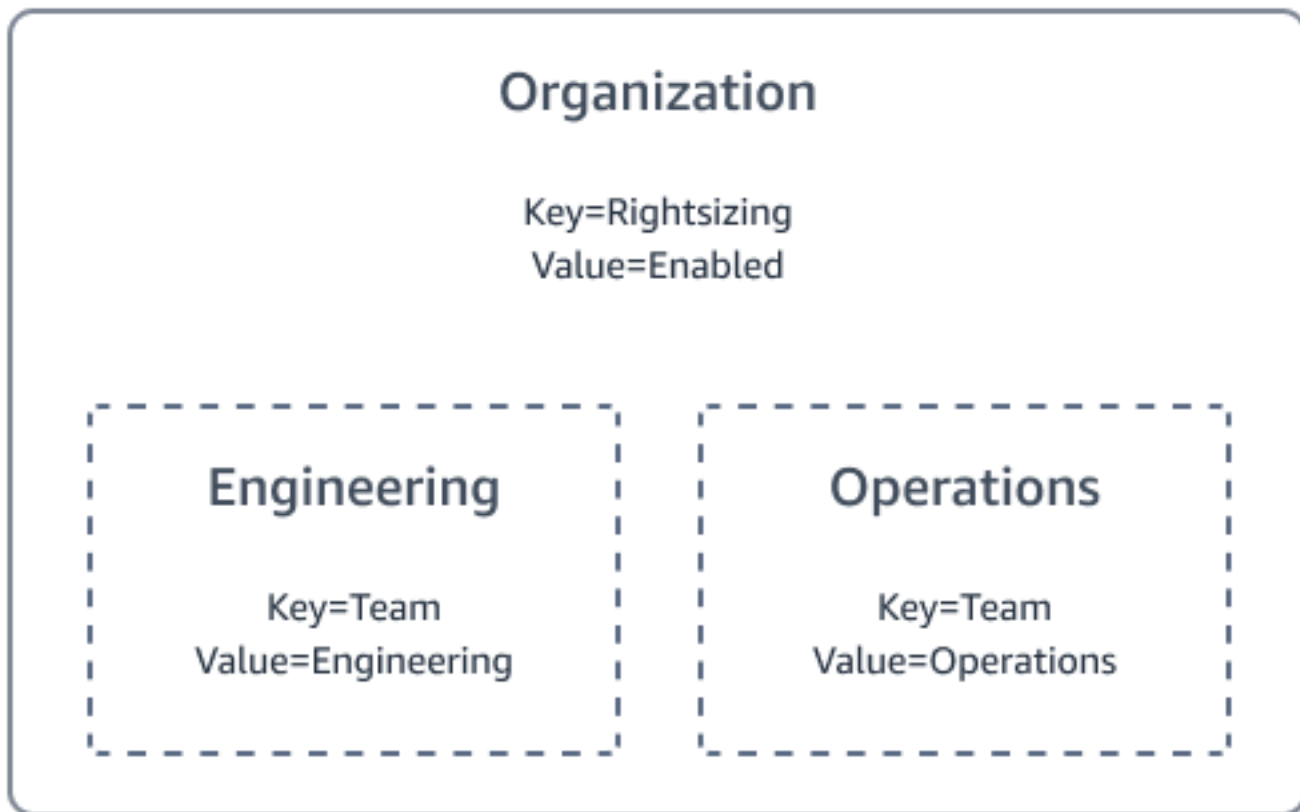
Se você estiver usando um modelo de execução para implantar suas instâncias, poderá atualizar o modelo de execução com o tipo de instância do tamanho certo e, em seguida, realizar uma atualização da instância para substituir as instâncias pela versão do tamanho certo.

Se você planeja usar o processo de dimensionamento correto em várias contas e regiões, deve criar um documento personalizado do Systems Manager Automation. Este documento permite alimentar várias instâncias como parâmetro e direcionar instâncias para o mesmo tipo de instância de destino (por exemplo, todas as instâncias em transição para t3a.medium, independentemente do tipo de instância de origem).

Analise os custos antes e depois no Cost Explorer

Depois de dimensionar corretamente seus recursos, você pode usar o Cost Explorer para mostrar os custos antes e depois usando a tag Rightsizing. Lembre-se de que você pode usar [tags de recursos](#) para controlar os custos. Ao usar várias camadas de tags, você pode obter visibilidade granular de seus custos. No exemplo abordado neste guia, a tag Rightsizing é usada para aplicar uma tag genérica a todas as instâncias de destino. Em seguida, uma tag de equipe é usada para organizar ainda mais os recursos. A próxima etapa é introduzir tags de aplicativo para mostrar ainda mais o impacto nos custos de operação de um aplicativo específico.

O diagrama a seguir mostra a estrutura de tags de uma organização.



Considere o exemplo de uma empresa que dimensiona corretamente os servidores web de produção de propriedade da equipe de operações. No Cost Explorer, a tag Rightsizing é definida como ativada e a tag Team está definida como operações. Neste exemplo, o esforço de dimensionamento correto reduz os custos operacionais de 0,89 centavos para 0,28 centavos por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de \$7.945,92. Após o dimensionamento correto, o custo anual cai para \$2.499,84. Isso se traduz em uma redução de 68,5% nos custos anuais da carga de trabalho. Imagine o impacto disso em uma grande organização. Lembre-se de que isso é feito em um ambiente de amostra e as instâncias estão, em sua maioria, ociosas. Em um ambiente de produção, você pode ver economias entre 10 e 35 por cento.

Agora, considere o impacto do dimensionamento correto do bastião de produção de propriedade da equipe de engenharia. No Cost Explorer, a tag Rightsizing é definida como ativada e a tag Team está definida como engenharia. Neste exemplo, o esforço de dimensionamento correto reduz os custos de 0,75 centavos para 0,44 centavos por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de \$6.696,00. Após o dimensionamento correto, o custo anual cai para \$3.928,32.

Se você usar várias tags, poderá filtrar os dados até obter detalhes granulares de custo. Neste exemplo, a tag Equipe reduz o ruído para que você possa ver o impacto em nível de equipe. Como

a tag Rightsizing está ativada, você também pode filtrar por qualquer instância que tenha essa tag com o valor de habilitado ou nenhum valor presente. Isso pode fornecer uma visão global de seus esforços de dimensionamento correto, especialmente quando visualizados na conta de gerenciamento (pagador) no nível do Cost Explorer. Essa visualização permite que você veja todas as contas e instâncias.

Considere um exemplo no nível de conta única em que a tag Rightsizing está definida como ativada. Os custos operacionais caem de \$1,64 por hora para \$0,72 centavos por hora. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de \$14.641,92. Após o dimensionamento correto, o custo anual cai para \$6.428,16. Isso se traduz em uma redução de 56% nos custos de computação dessa conta.

Antes de embarcar na viagem do tamanho certo, considere o seguinte:

- AWS oferece muitas opções para redução de custos. Isso inclui o [AWS OLA](#), em que AWS analisa suas instâncias locais antes de migrar para o. AWS O AWS OLA também fornece recomendações corretas de dimensionamento e orientação de licenciamento.
- Complete todo o tamanho correto antes de comprar [Savings Plans](#). Isso pode ajudá-lo a evitar compras excessivas em seu compromisso com Savings Plans.

Recomendações

Recomendamos as seguintes etapas:

1. Analise seu cenário atual e considere a conversão de volumes gp2 do Amazon EBS em volumes gp3.
2. Analise os [Savings Plans](#).

Recursos adicionais

- [AWS Compute Optimizer](#)(AWS documentação)
- [Melhores práticas para a marcação de AWS recursos](#) (AWS whitepapers)
- [Como coletar dados de AWS Compute Optimizer e AWS Trusted Advisor entre seu AWS Organizations](#) (YouTube)
- [Otimizando o desempenho e reduzindo os custos de licenciamento: aproveitando as instâncias do AWS Compute Optimizer SQL Server do Amazon EC2](#) (Microsoft Workloads no blog) AWS

Selecione o tipo de instância certo para cargas de trabalho do Windows

Visão geral

Uma distinção significativa entre cargas de trabalho operando na nuvem em comparação com ambientes locais é a prática de provisionamento excessivo. Ao comprar hardware físico para uso local, você faz uma despesa de capital que deve durar por um período predeterminado, normalmente de 3 a 5 anos. Para acomodar o crescimento previsto durante a vida útil do hardware, o hardware é adquirido com mais recursos do que sua carga de trabalho exige atualmente. Consequentemente, o hardware físico geralmente é provisionado em excesso, muito além das necessidades de sua carga de trabalho real.

A tecnologia de máquina virtual (VM) surgiu como um meio eficaz de utilizar recursos de hardware excedentes. Os administradores superprovisionaram VMs com vCPUs e RAM, permitindo que o hipervisor gerenciasse o uso de recursos físicos entre servidores ocupados e ociosos alocando recursos não utilizados para cada VM. Ao gerenciar VMs, os recursos de vCPU e RAM alocados para cada VM funcionavam mais como governadores de recursos do que como indicadores do uso real. A superalocação de recursos da VM pode facilmente exceder três vezes os recursos computacionais disponíveis.

[O Amazon Elastic Compute Cloud \(Amazon EC2\) evita o](#) provisionamento excessivo de VMs no hardware subjacente, pois isso é desnecessário. A computação em nuvem é uma despesa operacional, não uma despesa de capital, e você paga apenas pelo que usa. Se sua carga de trabalho exigir mais recursos no futuro, provisione-os quando você realmente precisar deles, em vez de fazer isso preventivamente.

Há centenas de opções para escolher os tipos certos de [instância do Amazon EC2](#). Se você planeja migrar uma carga de trabalho do Windows para a nuvem, AWS oferece um [AWS OLA](#) para ajudá-lo a entender melhor sua carga de trabalho atual e fornecer um exemplo de seu desempenho em AWS. A análise do AWS OLA visa combinar o tipo e o tamanho de instância EC2 adequados ao seu uso real no local.

Se você já tem cargas de trabalho em execução no Amazon EC2 e busca estratégias de otimização de custos, esta seção do guia ajuda a identificar diferenças entre as instâncias do Amazon EC2 e sua aplicabilidade às cargas de trabalho típicas do Windows.

Recomendações de otimização de custos

Para otimizar os custos dos seus tipos de instância do EC2, recomendamos que você faça o seguinte:

- Escolha a família de instâncias certa para sua carga de trabalho
- Entenda as variações de preço entre as arquiteturas de processadores
- Entenda as diferenças entre preço e desempenho nas gerações do EC2
- Migre para instâncias mais novas
- Use instâncias com capacidade de intermitência

Escolha a família de instâncias certa para sua carga de trabalho

É importante escolher a família de instâncias certa para sua carga de trabalho.

As instâncias do Amazon EC2 são divididas nos seguintes grupos:

- Uso geral
- Otimizadas para computação
- Otimizado para memória
- Computação acelerada
- Otimizada para armazenamento
- Otimizado para HPC

A maioria das cargas de trabalho do Windows se encaixa nas seguintes categorias:

- Uso geral
- Otimizadas para computação
- Otimizado para memória

Para simplificar ainda mais, considere uma instância básica do EC2 em cada categoria:


- Otimizado para computação — C6i
- Uso geral — M6i

- Memória otimizada — R6i

A geração anterior de instâncias do EC2 exibiu pequenas diferenças nos tipos de processadores. Por exemplo, as instâncias otimizadas para computação C5 têm processadores mais rápidos do que as instâncias M5 de uso geral ou as instâncias otimizadas para memória R5. Todas as instâncias EC2 de última geração (C6i, M6i, R6i, C6a, M6a e R6a) usam o mesmo processador em todas as famílias de instâncias. Como o processador é consistente entre as instâncias de última geração, a diferença de preço entre as famílias de instâncias agora depende mais da quantidade de RAM. Quanto mais RAM uma instância tiver, mais cara ela será.

O exemplo a seguir ilustra o preço por hora de uma instância de 4 vCPUs baseada em Intel em execução na região. us-east-1

Instância	vCPUs	RAM	Custo por hora
c6i.xlarge	4	8	0,17 US\$
m6i.xlarge	4	16	0,19 US\$
r6i.xlarge	4	32	\$0,25

 Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

Instâncias intermitentes

Embora seja uma prática recomendada na computação em nuvem desativar os recursos computacionais não utilizados para evitar cobranças, nem todas as cargas de trabalho podem ser desligadas e ativadas sempre que necessárias. Algumas cargas de trabalho permanecem inativas por longos períodos, mas devem estar acessíveis 24 horas por dia.

As instâncias intermitentes (T3) oferecem uma maneira de manter cargas de trabalho com picos ou de baixa utilização on-line o dia todo, mantendo os custos de computação baixos. As instâncias EC2 com capacidade de intermitência têm uma quantidade máxima de recursos de vCPU que a instância pode usar por breves períodos. Essas instâncias empregam um sistema baseado em créditos de

CPU com capacidade de intermitência. Esses créditos são acumulados durante os períodos de inatividade ao longo do dia. As instâncias intermitentes oferecem proporções variáveis de vCPU para RAM, tornando-as alternativas para instâncias otimizadas para computação em alguns casos e para outras instâncias de uso geral em outros.

O exemplo a seguir ilustra o preço por hora de uma instância T3 (ou seja, instância com capacidade de intermitência) em execução na região. us-east-1

Instância	vCPUs	MEMÓRIA RAM (GB)	Custo por hora
t3.nano	2	0,5	\$0,0052
t3.micro	2	1	\$0,0104
t3.small	2	2	\$0.0208
t3.medium	2	4	\$0,0416
t3.large	2	8	\$0,0832
t3.xlarge	4	16	0,164 US\$
t3.2xlarge	8	32	\$0,328

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

Entenda as variações de preço entre as arquiteturas de processadores

Os processadores [Intel](#) têm sido o padrão para instâncias EC2 desde o início. Gerações anteriores de instâncias do EC2, como C5, M5 e R5, não indicam a Intel como a arquitetura do processador (já que era o padrão). As novas gerações de instâncias do EC2, como C6i, M6i e R6i, incluem um “i” para indicar o uso de um processador Intel.

A mudança na anotação da arquitetura do processador se deve à introdução de opções adicionais de processador. O processador mais comparável ao Intel é o [AMD](#) (indicado com um “a”). Os processadores AMD EPYC usam a mesma arquitetura x86 e oferecem desempenho semelhante aos

processadores Intel, mas a um preço mais baixo. Conforme demonstrado nos exemplos de preços a seguir, as instâncias AMD EC2 oferecem um desconto de aproximadamente 10% nos custos de computação em comparação com as da Intel.

Instância Intel	Custo por hora	Instância AMD	Preço	% de diferença
c6i.xlarge	0,17 US\$	c6a.xlarge	\$0,153	10%
m6i.xlarge	\$0,192	m6a.xlarge	\$0,1728	10%
r6i.xlarge	\$0,252	r6a.xlarge	\$0,268	10%

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

A terceira principal opção de arquitetura de processador são os [processadores AWS Graviton](#) (indicados com um “g”) em instâncias EC2. Projetados por AWS, os processadores Graviton oferecem a melhor relação preço/desempenho no Amazon EC2. Os processadores Graviton atuais não são apenas 20% mais baratos do que seus equivalentes da Intel, mas também oferecem um aumento de desempenho de 20% ou mais. Espera-se que a próxima geração de processadores Graviton amplie ainda mais essa diferença de desempenho, com testes mostrando um aumento adicional de 25% no desempenho.

O Windows Server não pode ser executado nos processadores Graviton, que são baseados na arquitetura ARM. Na verdade, o Windows Server opera somente em processadores x86. Embora você não possa obter um aumento de 40% no desempenho de preço usando instâncias baseadas em Graviton para Windows Server, você ainda pode usar processadores Graviton com cargas de trabalho específicas da Microsoft. Por exemplo, [versões mais recentes do .NET podem ser executadas no Linux](#). Isso significa que essas cargas de trabalho podem usar processadores ARM e se beneficiar de instâncias Graviton EC2 mais rápidas e acessíveis.

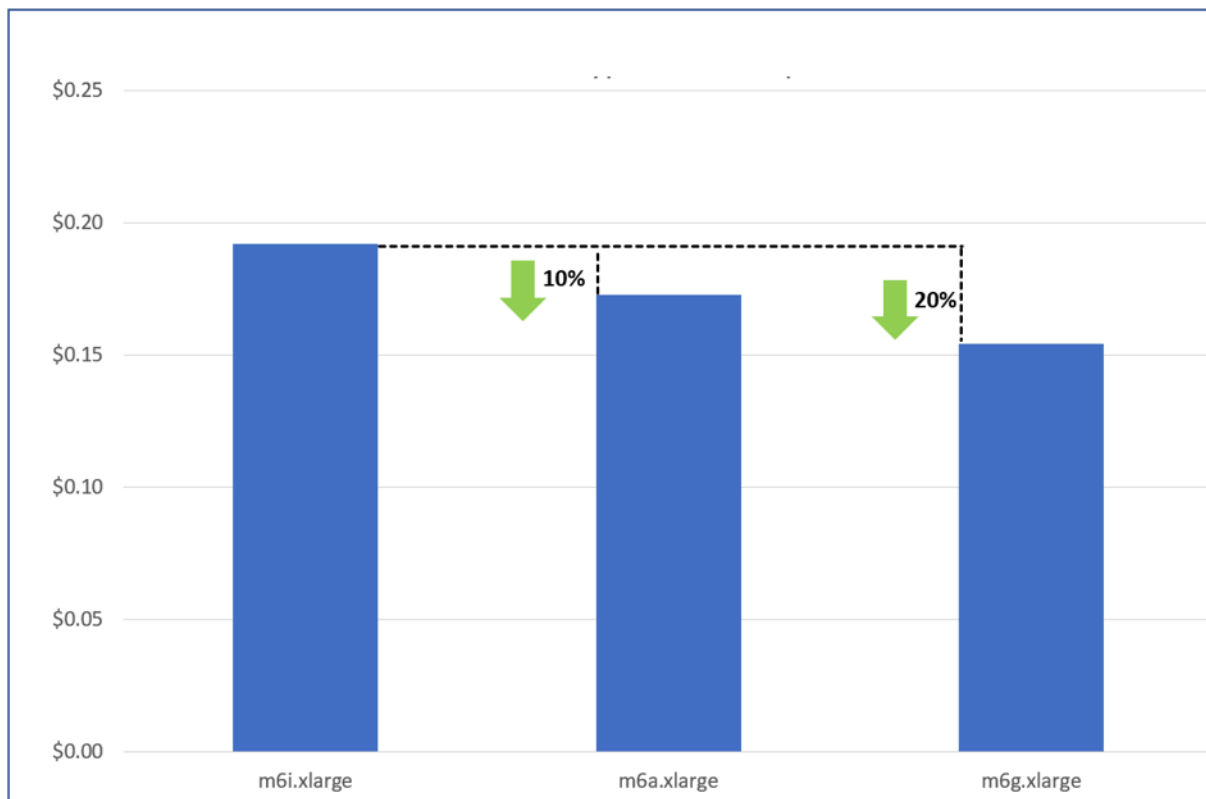
O exemplo a seguir ilustra o preço por hora de uma instância do Graviton em execução na us-east-1 região.

Instância Intel	Custo por hora	Instância de Graviton	Custo por hora	% de diferença
c6i.xlarge	0,17 US\$	c6g.xlarge	\$0,136	20%
m6i.xlarge	\$0,192	m6g.xlarge	\$0,154	20%
r6i.xlarge	\$0,252	r6g.xlarge	\$0,2016	20%

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

O gráfico a seguir compara os preços das instâncias da série M.



Entenda as diferenças de preço/desempenho entre as gerações do EC2

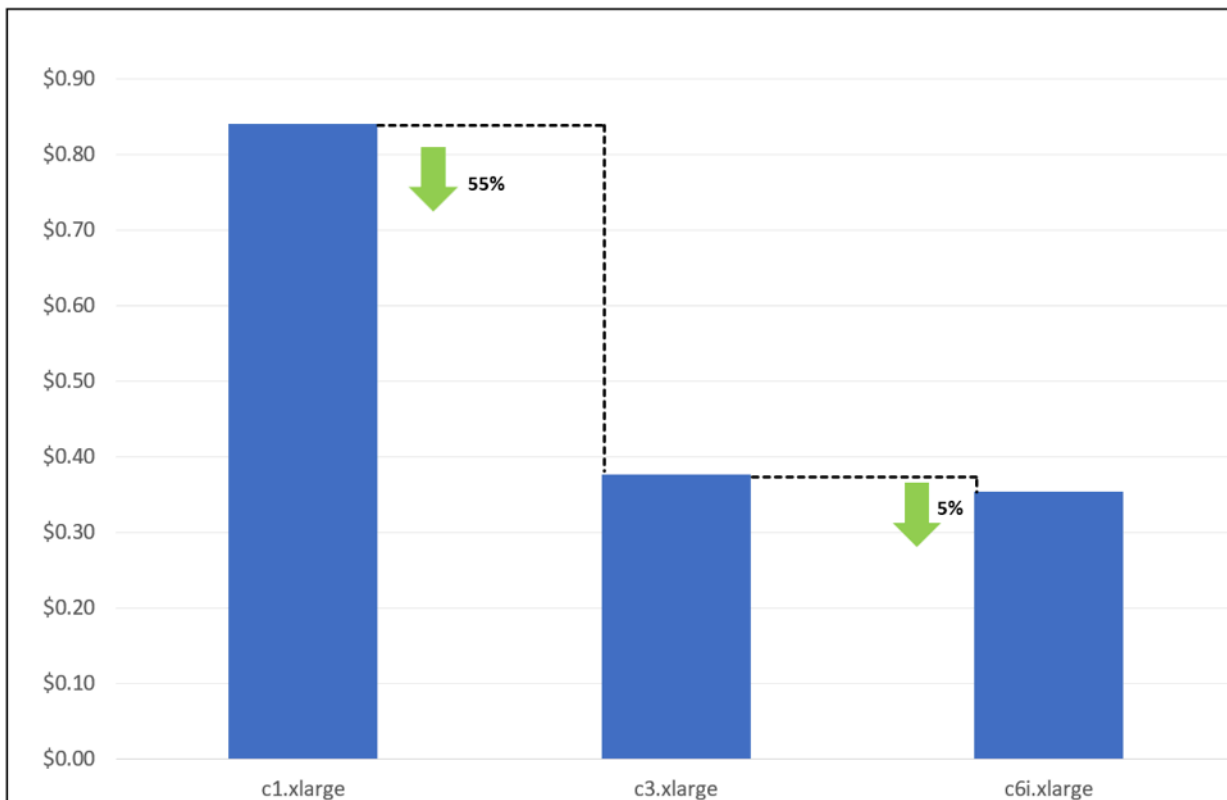
Uma das características mais consistentes do Amazon EC2 é que cada nova geração oferece melhor desempenho de preço do que sua antecessora. Como mostra a tabela a seguir, o preço das instâncias EC2 de nova geração diminui a cada versão subsequente.

Instância otimizada para computação	Custo por hora	Instância de uso geral	Custo por hora	Instância otimizada para memória	Custo por hora
C1.xlarge	\$0,52	M1.x grande	\$0,35	r1.xlarge	n/a
C3.xlarge	\$0,21	M3.x grande	\$0,266	r3.xlarge	\$0,333
C5.xlarge	0,17 US\$	M5.x grande	\$0,192	r5.xlarge	\$0,252

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

O gráfico a seguir compara os custos das diferentes gerações de instâncias da série C.



No entanto, a 6ª geração de instâncias tem o mesmo preço da 5ª geração, conforme mostra a tabela a seguir.

Instância otimizada para computação	Custo por hora	Instância de uso geral	Custo por hora	Instância otimizada para memória	Custo por hora
C5.xlarge	0,17 US\$	M5.x grande	\$0,192	r5.xlarge	\$0,252
C6I.XLarge	0,17 US\$	M6i.xlarge	\$0,192	r6i.xlarge	\$0,252

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

Apesar de ter o mesmo custo, a nova geração oferece desempenho de preço superior devido a processadores mais rápidos, maior taxa de transferência de rede e maior taxa de transferência e IOPS do Amazon Elastic Block Store (Amazon EBS).

Uma das melhorias mais significativas de preço/desempenho é o aprimoramento da instância [X2i](#). Essa geração da instância oferece um desempenho de preço até 55% maior do que a geração anterior. Como mostra a tabela a seguir, o x2iedn demonstra melhorias em todos os aspectos de desempenho (tudo pelo mesmo preço da geração anterior).

Instância	Custo por hora	vCPUs	RAM	Velocidade do processador	Armazenamento de instâncias	Redes	Taxa de transferência do Amazon EBS	IOPS do EBS
x1e.2xlarge	\$1,66	8	244	2,3 GHz	SSD DE 237 GB	10 Gbps	125 Mb/s	7400
x1iedn.2xlarge	\$1,66	8	256	3,5 GHz	SSD NVMe de 240 GB	25 Gbps	2500 MB/s	65000

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

Cenários de exemplo de

Considere o exemplo de uma empresa de análise que rastreia veículos de entrega e deseja melhorar o desempenho do SQL Server. Depois que uma PME MACO analisa os gargalos de desempenho dessa empresa, a empresa faz a transição de instâncias x1e.2xlarge para instâncias x2iedn.xlarge. O novo tamanho da instância é menor, mas os aprimoramentos nas instâncias x2 permitem maior desempenho e otimização do SQL Server por meio do uso de extensões de pool de buffer. Isso permite que a empresa faça o downgrade da edição SQL Server Enterprise para a edição SQL

Server Standard. Também permite que a empresa reduza seu licenciamento do SQL Server de 8 vCPUs para 4 vCPUs.

Antes da otimização:

Servidor	EC2 instance (Instância do EC2)	Edição do SQL Server	Custo mensal
ProdDB1	x1e.2xlarge	Enterprise	\$3.918,64
ProdDB2	x1e.2xlarge	Enterprise	\$3.918,64
Total			\$7.837,28

Após a otimização:

Servidor	EC2 instance (Instância do EC2)	Edição do SQL Server	Custo mensal
ProdDB1	x2iedn.xlarge	Padrão	\$1.215,00
ProdDB2	x2iedn.xlarge	Padrão	\$1.215,00
Total			\$2.430,00

Ao todo, a mudança de instâncias x1e.2xlarge para instâncias x2iedn.xlarge permite que a empresa, no cenário de exemplo, economize 5.407 USD por mês em seus servidores de banco de dados de produção. Isso reduz o custo total da carga de trabalho em 69%.

Note

Os preços são baseados nos preços por hora sob demanda na us-east-1 região.

Migre para instâncias mais novas

[As gerações mais antigas do Amazon EC2 são executadas no hipervisor Xen, enquanto as gerações mais novas operam no Sistema Nitro.AWS](#) O Sistema Nitro fornece quase todos os recursos de

computação e memória do hardware host para suas instâncias. Isso resulta em melhor desempenho geral. Há considerações especiais ao [migrar de instâncias baseadas em Xen para instâncias baseadas](#) em Nitro. Por exemplo, as [AMIs AWS do Windows são](#) definidas com configurações e personalizações padrão usadas pela mídia de instalação da Microsoft. As personalizações incluem drivers e configurações que oferecem suporte aos tipos de instância de última geração ([instâncias criadas no Sistema Nitro](#)).

Se você estiver iniciando instâncias a partir de AMIs personalizadas do Windows ou de AMIs do Windows fornecidas pela Amazon que foram criadas antes de agosto de 2018, recomendamos que você conclua as etapas de [Migrar para os tipos de instância de última geração](#) na documentação do Amazon EC2.

Use instâncias com capacidade de intermitência

Embora as instâncias com capacidade de intermitência sejam uma boa maneira de economizar nos custos de computação, recomendamos que você as evite nos seguintes cenários:

- [As especificações mínimas do Windows Server](#) com a experiência de desktop exigem 2 GB de RAM. Evite usar instâncias t3.micro ou t3.nano com o Windows Server porque elas não têm a quantidade mínima de RAM.
- Se sua carga de trabalho estiver alta, mas não ficar ociosa por tempo suficiente para criar créditos de intermitência, usar instâncias EC2 normais é mais eficiente do que usar instâncias com capacidade de intermitência. Recomendamos [monitorar seus créditos de CPU](#) para verificar isso.
- Recomendamos que você evite usar instâncias intermitentes com o SQL Server na maioria dos cenários. O licenciamento do SQL Server é baseado no número de vCPUs atribuídas a uma instância. Se o SQL Server ficar inativo a maior parte do dia, você pagaria por licenças SQL que não está utilizando totalmente. Nesses cenários, recomendamos que você consolide várias instâncias do SQL Server em um servidor maior.

Próximas etapas

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos com as instâncias Windows do Amazon EC2:

- Use a instância EC2 de última geração para obter a melhor relação preço/desempenho.
- Use instâncias EC2 com processadores AMD para uma redução de dez por cento nos custos de computação.

- Maximize a utilização dos recursos escolhendo um tipo de instância do EC2 que corresponda à sua carga de trabalho.

A tabela a seguir mostra exemplos de pontos de partida típicos para cargas de trabalho do Windows. Opções adicionais estão disponíveis, como volumes de armazenamento de instâncias para aprimorar as cargas de trabalho do SQL Server ou instâncias EC2 com proporções muito maiores de VCPU/RAM. Recomendamos que você teste suas cargas de trabalho minuciosamente e use ferramentas de monitoramento AWS Compute Optimizer para ajudar a fazer os ajustes necessários.

Workload	Típico	Opcional
Active Directory	T3, M6i	R6i
Servidores de arquivos	T3, M6i	C6i
Servidores da web	T3, C6i	M6i, R6i
SQL Server	R6i	x2iedn, x2iEzn

Se você precisar alterar o tipo de instância do EC2, o processo normalmente envolve apenas uma simples reinicialização do servidor. Para obter mais informações, consulte [Alterar o tipo de instância](#) na documentação do Amazon EC2.

Antes de alterar o tipo de instância, recomendamos que você considere o seguinte:

- Você deve interromper suas instâncias apoiadas pelo Amazon EBS antes de poder alterar o tipo de instância. Certifique-se de planejar o tempo de inatividade enquanto sua instância estiver parada. Interromper a instância e alterar o tipo de instância pode levar alguns minutos, e o tempo necessário para iniciar a instância pode variar dependendo dos scripts de startup da aplicação. Para obter mais informações, consulte [Pare e inicie sua instância](#) na documentação do Amazon EC2.
- Quando você interrompe e inicia uma instância, AWS move a instância para um novo hardware. Se sua instância tiver um endereço IPv4 público, AWS liberará o endereço e fornecerá à instância um novo endereço IPv4 público. Se você precisar de um endereço IPv4 público que não mude, use um endereço [IP elástico](#).
- Você não pode alterar o tipo de instância se a [hibernação](#) estiver ativada na instância.
- Você não pode alterar o tipo de instância de uma [instância spot](#).

- Se sua instância estiver em um grupo de Auto Scaling, o Amazon EC2 Auto Scaling marca a instância interrompida como não íntegra e poderá encerrá-la e iniciar uma instância substituta. Para evitar isso, é possível suspender os processos de escalabilidade para o grupo enquanto estiver alterando o tipo de instância. Para obter mais informações, consulte [Suspender e retomar um processo para um grupo de Auto Scaling](#) na documentação do Amazon EC2 Auto Scaling.
- Quando você altera o tipo de instância de uma instância com volumes de armazenamento de instâncias NVMe, a instância atualizada pode ter volumes adicionais de armazenamento de instâncias, porque todos os volumes de armazenamento de instâncias NVMe estão disponíveis mesmo que não estejam especificados na Amazon Machine Image (AMI) ou no mapeamento de dispositivos de blocos de instâncias. Caso contrário, a instância atualizada tem o mesmo número de volumes de armazenamento de instância que você especificou ao iniciar a instância original.

Recursos adicionais

- [Tipos de instância do Amazon EC2 \(documentação\)](#) AWS
- [AWS Avaliação de otimização e licenciamento](#) (AWS documentação)

Traga licenças para cargas de trabalho do Windows e do SQL Server

Visão geral

Se você tiver investimentos significativos em cargas de trabalho da Microsoft e contratos de licenciamento corporativo existentes, poderá escolher entre várias AWS opções para oferecer suporte a essas cargas de trabalho, incluindo as opções de [licença incluída \(fornecida por AWS\)](#) e [Bring Your Own License \(BYOL\)](#). Você pode usar os [hosts dedicados do Amazon EC2](#) para aproveitar totalmente os contratos de licenciamento existentes da Microsoft e trazer o Windows Server para o. AWS Isso pode economizar até 50% nos custos de instância do Amazon EC2. Como as licenças do Windows representam aproximadamente metade dos custos da instância, colocar o Windows Server AWS em hosts dedicados pode resultar em economias substanciais. Como o Windows Server não pode ser colocado em [locação padrão \(compartilhada\)](#), os hosts dedicados são a escolha ideal se você quiser usar suas licenças existentes para o Windows Server. AWS

Os hosts dedicados não servem apenas para instâncias BYOL do Windows Server. Eles também oferecem a flexibilidade de adequar seu licenciamento local às cargas de trabalho existentes do

SQL Server. Os hosts dedicados expõem os núcleos físicos do servidor subjacente e permitem que você licencie o SQL Server no nível do núcleo físico. Isso não é possível na locação padrão (compartilhada) em que o licenciamento do SQL Server é baseado no número de CPUs virtuais alocadas para a instância. Esse recurso permite que você licencie cargas de trabalho do SQL Server de AWS forma consistente com sua estratégia de licenciamento local. Consequentemente, você pode economizar até 50% nos custos de licenciamento do SQL Server em comparação com a locação padrão (compartilhada), além da economia nos custos de instância, usando licenças qualificadas do Windows. Para obter mais informações sobre esse cenário, consulte a seção [Compreender o licenciamento do SQL Server](#) deste guia.

Hosts dedicados do Amazon EC2

Um host dedicado do Amazon EC2 é essencialmente o mesmo host EC2 AWS usado para executar suas ofertas de computação do EC2. A diferença é que esses hosts são totalmente dedicados a um único cliente e fornecem acesso exclusivo à infraestrutura física subjacente. Você pode usar hosts dedicados para executar suas instâncias em hardware totalmente dedicado ao seu uso, em vez de compartilhar recursos com outros AWS clientes. Isso lhe dá maior controle sobre os recursos da nuvem e permite que você reduza custos trazendo suas próprias licenças de software, como Windows Server e SQL Server, para o AWS.

Lembre-se do seguinte:

- Um host dedicado é um servidor físico totalmente dedicado a um único cliente. Você obtém visibilidade dos soquetes e dos núcleos físicos do host dedicado para poder atender aos requisitos de conformidade de licenciamento, como contratos de licenciamento de software por soquete, por núcleo ou por VM.
- Hosts dedicados que podem suportar vários tamanhos de instância da mesma família de instâncias são conhecidos como hosts dedicados heterogêneos. Essas [famílias de instâncias](#) incluem T3, A1, C5, M5, R5, C5n, R5n e M5n. Por outro lado, outras famílias de instâncias oferecem suporte a apenas um tamanho de instância no mesmo host dedicado. Eles são chamados de hosts dedicados homogêneos.
- Os anfitriões dedicados são cobrados por anfitrião. Isso significa que você é cobrado por host dedicado, independentemente de quantas instâncias estejam sendo executadas nele. O preço do host dedicado varia de acordo com a família da instância, a região e a opção de pagamento selecionada. Você pode escolher a configuração ideal para sua carga de trabalho para alcançar os resultados desejados de desempenho e custo.

Este diagrama ilustra as diferenças entre instâncias de locação compartilhada e hosts dedicados.



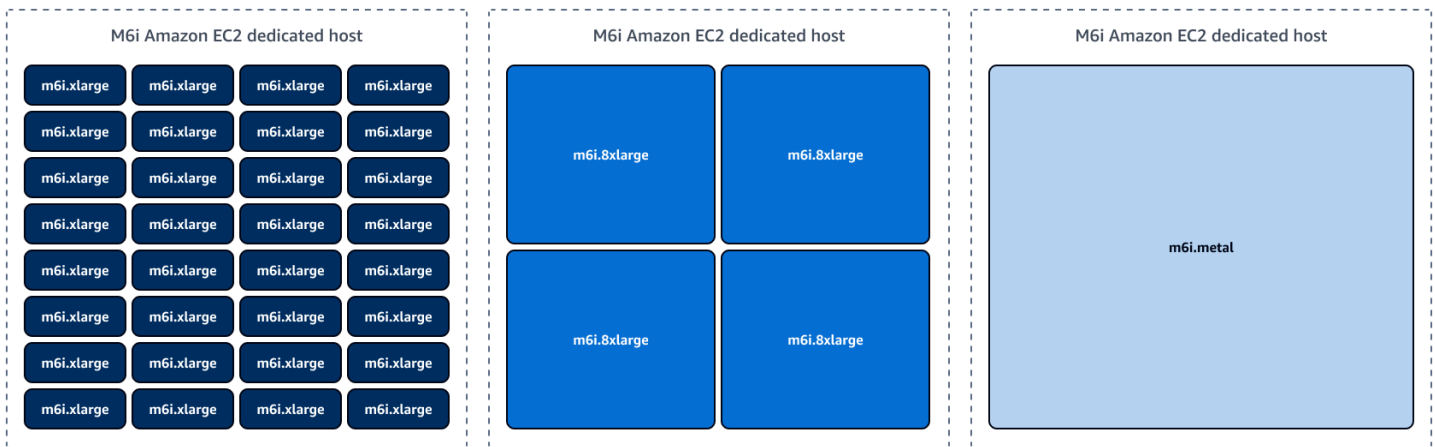
Hosts dedicados homogêneos

Considere um cenário em que um host dedicado M6i é usado. Os hosts dedicados M6i e R6i têm dois soquetes, 64 núcleos físicos e oferecem suporte a tipos de instância do mesmo tamanho. Eles são chamados de hosts dedicados homogêneos. Isso significa que o número de instâncias que você pode executar em um único host dedicado M6i depende do tamanho da instância.

Por exemplo: .

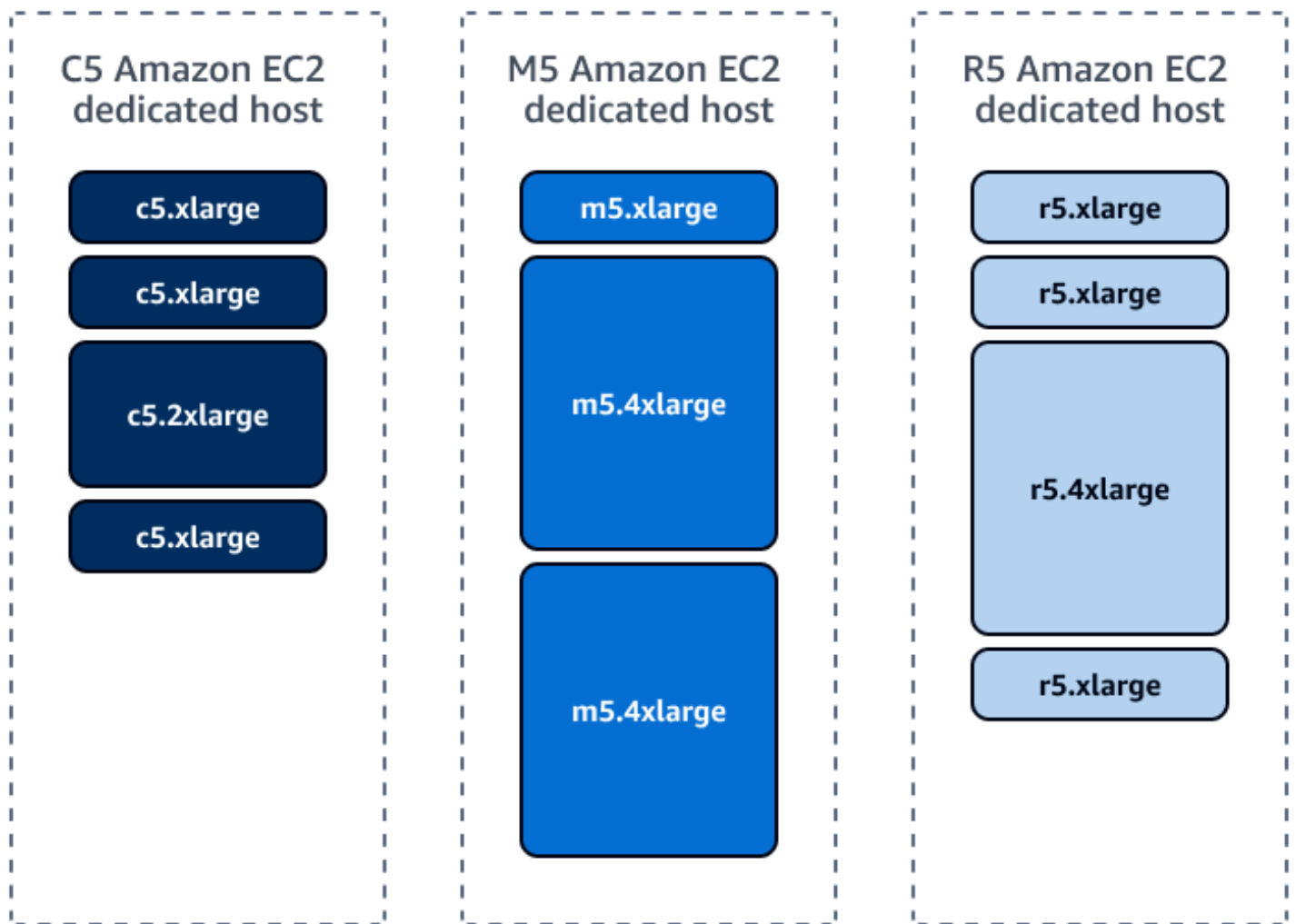
- No caso do xlarge (4 vCPUs), você pode executar no máximo 32 instâncias m6i.xlarge nesse host dedicado.
- No caso de 8xlarge (32 vCPUs), você pode executar no máximo 4 instâncias m6i.8xlarge nesse host dedicado.
- No caso do metal (128 vCPUs), você pode executar no máximo 1 instância m6i.metal nesse host dedicado.

O diagrama a seguir mostra as opções de host dedicado para instâncias M6.



Hosts dedicados heterogêneos

Os hosts dedicados que suportam vários tamanhos de instância no mesmo host são chamados de hosts dedicados heterogêneos do Amazon EC2. O diagrama a seguir mostra um exemplo de hosts dedicados C5, M5 e R5 com vários tamanhos de instância, como 2xlarge, xlarge e 4xlarge.



Gerenciamento de host dedicado

Recomendamos que você considere o seguinte em relação ao gerenciamento de hosts dedicados do Amazon EC2:

- Para aproveitar ao máximo os hosts dedicados, você pode [compartilhar um único host entre várias contas em sua organização](#). O compartilhamento de host permite a otimização de recursos e pode resultar em economia de custos ao usar todos os slots disponíveis no host. Ao compartilhar um host dedicado entre unidades de negócios, você pode centralizar sua infraestrutura de TI e melhorar a utilização dos recursos, mantendo a separação entre as cargas de trabalho. Se você faz parte de uma organização AWS Organizations e o compartilhamento está habilitado em sua organização, os consumidores em sua organização recebem automaticamente acesso ao Host dedicado compartilhado. Caso contrário, os consumidores receberão um convite para participar

do compartilhamento de recursos e acesso ao Host dedicado compartilhado depois de aceitar o convite.

- Você pode executar o Windows Server 2022 em hosts dedicados sob o modelo de licença incluída, já que o Windows Server 2019 é a versão mais recente em que você pode fazer BYOL. Se você quiser usar o Windows Server 2022 em hosts dedicados, deverá usar instâncias com licença incluída do Windows Server 2022.
- [AWS License Manager](#) é uma solução abrangente para gerenciar licenças de software de vários fornecedores em ambientes locais AWS e internos. Se você [usa o License Manager](#), pode obter maior visibilidade e controle sobre como as licenças de software são usadas, resultando em economia de custos e maior conformidade. Você pode usar o License Manager para definir regras para emular seus termos de licenciamento exclusivos. Isso permite que você aplique essas regras e evite o uso indevido da licença. Isso pode reduzir o risco de não conformidade e melhorar os processos de gerenciamento de licenças.
- Você pode usar o License Manager para automatizar o posicionamento, a liberação e a recuperação de hosts usando [grupos de recursos de hosts](#). Isso pode aumentar a produtividade e reduzir a sobrecarga de gerenciamento. O License Manager também fornece uma visão centralizada do uso de licenças em ambientes locais AWS e em todos os ambientes com base nas regras de licenciamento, facilitando o gerenciamento de compras incrementais de licenciamento, conformidade e auditorias de fornecedores em toda a organização. Além disso, o License Manager se integra com AWS Organizations e AWS Resource Access Manager (AWS RAM) para compartilhar configurações de licença entre contas e regiões. Isso permite que você crie relatórios para todo o seu ambiente com base em um cronograma e gerencie as regras de licenciamento de forma centralizada em uma. Conta da AWS Em última análise, isso pode melhorar a governança e reduzir a complexidade.
- Ao projetar a alta disponibilidade para hosts dedicados em uma única região, certifique-se de ter alocado no mínimo dois hosts dedicados em no mínimo duas zonas de disponibilidade para cargas de trabalho críticas de produção. Para obter mais informações, consulte os [hosts dedicados do Amazon EC2 para Microsoft Windows sobre implantação](#) de AWS referência.
- Para cada família de instâncias de host dedicado, há um limite no número de instâncias que você pode executar para cada tamanho de instância. Para obter mais informações, consulte a [tabela de configuração de hosts dedicados](#) na documentação do Amazon EC2.

AWS opções de licenciamento

As licenças são classificadas nas seguintes categorias principais:

- Licença incluída — Essa opção de licenciamento permite que você compre e use licenças sob demanda, pagando somente pelo que você usa. É ideal para casos de uso em que você busca flexibilidade no uso do licenciamento e deseja evitar custos iniciais. Você pode escolher entre uma variedade de produtos Windows Server, SQL Server e outros produtos da Microsoft.
- Produtos BYOL com mobilidade de licenças — Se você já tem licenças existentes e deseja usá-las na nuvem, essa opção de licenciamento permite que você traga suas próprias licenças para a nuvem por meio do programa [Microsoft License Mobility](#). Produtos com mobilidade de licenças, como o SQL Server com Software Assurance (SA), podem ser transferidos para locação compartilhada ou dedicada. Isso reduz os custos da AWS instância.
- Produtos BYOL sem mobilidade de licenças — Para produtos da Microsoft, como o Windows Server, que não possuem mobilidade de licenças, AWS fornece opções dedicadas para o uso desses produtos na nuvem. Além disso, os hosts dedicados permitem o licenciamento no nível do núcleo físico, possibilitando economizar 50% ou mais nas licenças necessárias para executar suas cargas de trabalho. Os hosts dedicados são uma excelente opção para cargas de trabalho estáveis e previsíveis que são executadas na maior parte do tempo.

Trazendo licenças do Windows Server

Trazer suas próprias licenças do Windows é uma das estratégias mais eficazes para otimização de licenças, pois permite que você aproveite os investimentos existentes e reduza suas AWS despesas. Cenários específicos de BYOL não exigem benefícios de SA ou mobilidade de licenças, mas a infraestrutura dedicada do Amazon EC2 é sempre necessária. Para se qualificar, você deve ter comprado licenças perpétuas antes de 1º de outubro de 2019 ou adicionadas como uma atualização em uma inscrição corporativa ativa em vigor antes de 1º de outubro de 2019. Nesses cenários específicos de BYOL, você pode atualizar somente licenças para versões disponíveis antes de 1º de outubro de 2019. Por exemplo, se você abandonou o SA em 2017, você tem os direitos de implantação somente no Windows Server 2016, não em 2019. No entanto, 2019 é a última versão elegível para BYOL. AWS Para obter mais informações, consulte [Licenciamento — Windows Server](#) na AWS documentação.

A obtenção de licenças pode afetar significativamente o custo de execução das cargas de trabalho da Microsoft. AWS Ao trazer suas próprias licenças, você não precisa pagar custos adicionais de licenciamento pelas instâncias em execução na nuvem, o que pode levar a uma economia considerável.

A tabela a seguir mostra o custo mensal sob demanda da execução de uma única instância c5.xlarge 24 horas por dia, 7 dias por semana, em várias configurações.

Configuração	Custo mensal (USD)
Edição Windows Server + SQL Server Enterprise	\$1,353.00 (LI)
Edição padrão do Windows Server + SQL Server	\$609.00 (LI)
Somente Windows Server	\$259.00 (LI)
Somente computação (Linux)	\$127,00

Você pode usar suas licenças existentes para reduzir os custos de licenciamento e economizar dinheiro em sua fatura geral AWS .

Para se qualificar para BYOL em hosts dedicados do Amazon EC2, você deve trazer suas próprias licenças de software, como para Windows Server e SQL Server. O BYOL permite que você use suas licenças existentes AWS e pode resultar em economia de custos. Para trazer suas próprias licenças, você deve ter os direitos de licença do fornecedor do software e também fornecer a mídia ou a imagem de instalação do software. A mídia ou a imagem de instalação podem ser usadas para executar instâncias em hosts dedicados. Para saber mais sobre a criação de uma AMI BYOL, consulte [Como criar AMIs Bring-Your-Own-License do Windows Server localmente com o VM Import/Export no blog Microsoft Workloads on](#). AWS

Note

Um tipo de licença definido como Automático equivale a uma opção de [AWS licença incluída](#). Essa opção pode resultar em gastos indesejados sob demanda. Você precisará mudar os [tipos de licenciamento](#).

Cenários de otimização de custos

O dimensionamento correto e a otimização das licenças são um componente essencial da otimização de custos em AWS. Se você implementar as estratégias certas, poderá reduzir os custos de licenciamento, manter a conformidade e obter o melhor valor possível do seu investimento em licenciamento usando os hosts dedicados do Amazon EC2 e a opção BYOL.

Esta seção aborda os seguintes exemplos de cenários:

- Economia de custos com hosts dedicados T3
- Comparando locação compartilhada com hosts dedicados com SQL Server BYOL
- Implantações altamente disponíveis do SQL Server

Economia de custos com hosts dedicados T3

Os hosts dedicados T3 diferem de outros hosts dedicados do Amazon EC2 que tradicionalmente fornecem recursos fixos de CPU. Os hosts dedicados T3, por outro lado, oferecem suporte a instâncias com capacidade de intermitência que são capazes de compartilhar recursos da CPU, fornecer desempenho básico da CPU e progredir quando necessário. O compartilhamento de recursos de CPU, também conhecido como excesso de assinatura, é o que permite que um único host dedicado T3 ofereça suporte a até quatro vezes mais instâncias do que hosts dedicados de uso geral comparáveis.

Os hosts dedicados T3 geram um TCO menor ao oferecer maior densidade de instâncias do que qualquer outro host dedicado do Amazon EC2. As instâncias T3 com capacidade de intermitência permitem consolidar um número maior de instâncias com utilização low-to-moderate média da CPU em menos hosts do que nunca. Os hosts dedicados T3 também oferecem tamanhos de instância menores em um número maior de combinações de vCPU e memória do que outros hosts dedicados do Amazon EC2. Instâncias menores podem contribuir para reduzir o TCO e ajudar a fornecer taxas de consolidação equivalentes ou superiores às dos hosts locais.

Os hosts dedicados T3 são mais adequados para executar software BYOL com utilização de low-to-moderate CPU e licenças de software elegíveis por soquete, por núcleo ou por VM, incluindo Microsoft Windows desktop, Windows Server, SQL Server e bancos de dados Oracle.

Use hosts dedicados T3 para reduzir as licenças do Windows Server Datacenter (por núcleo)

Em ambientes locais, você está aproveitando o fato de poder facilmente sobrescrever suas CPUs físicas em hosts VMware e alcançar altos níveis de consolidação.

Considere o seguinte exemplo. No momento, você está usando hosts VMware de 10x36 núcleos e 384 GB de RAM em um ambiente local. Além disso, cada host está executando máquinas virtuais Windows Server de 96x2 vCPU e 4 GB de RAM com baixa média de utilização da CPU.

Agora você pode alcançar níveis muito mais altos de consolidação movendo suas máquinas virtuais para hosts dedicados T3, que têm o dobro da quantidade de RAM em comparação com seus hosts VMware locais atuais. Você pode executar o mesmo número de servidores em hosts dedicados T3 com 50% menos custo de host. Isso pode ajudá-lo a reduzir os custos de licenciamento do Windows Server em 33%. A tabela a seguir destaca a economia do uso de hosts dedicados T3.

	Hosts VMware locais	Hosts dedicados T3	Economia
Servidores físicos	10	5	
Núcleos físicos por host	36	48	
RAM por host (GB)	384	768	
2 vCPUs, 4 GB de RAM VMs por host	96	192	
Número total de VMs	960	960	
Total de licenças do Windows Server Datacenter (por núcleo) = (Número de servidores * Contagem física de núcleos)	$10 * 36 = 360$	$5 * 48 = 240$	33%

Comparando locação compartilhada com hosts dedicados com SQL Server BYOL

Considere um exemplo prático para demonstrar o valor dos hosts dedicados do Amazon EC2. Nesse cenário, uma organização executa uma carga de trabalho do SQL Server em um ambiente local com 240 núcleos e deseja implantar a mesma carga de trabalho de forma econômica em AWS. Se essa organização trazer suas próprias licenças (BYOL), ela continuará pagando pelo SA e a redução do número de núcleos afetará diretamente seus custos.

O diagrama a seguir compara a AWS economia entre direitos da Microsoft e do SQL Server.

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

Ao dimensionar corretamente as instâncias na locação AWS compartilhada, você pode reduzir as licenças do SQL Server para 140 núcleos. Isso resulta em custos de SA de \$197.000.

Os hosts dedicados do Amazon EC2 permitem que você licencie o SQL Server no nível do núcleo físico. Isso não é possível na locação compartilhada em que o licenciamento do SQL Server é baseado no número de vCPUs alocados para a instância. Consequentemente, ao usar dois hosts dedicados R5 com 48 núcleos cada, você só precisa cobrir 96 núcleos em vez dos 140 vCPUs necessários na locação compartilhada. Ao implantar hosts dedicados R5 e licenciar as cargas de trabalho no nível físico, você pode reduzir o número necessário de licenças da edição SQL Server Enterprise para 96 núcleos. Isso significa que você pode implantar até 192 núcleos (contabilizando o hyper-threading) de cargas de trabalho do SQL Server, sem deixar de atender aos requisitos de licenciamento e obter economias de custo significativas.

Nesse caso, a organização paga aproximadamente \$341.000 anualmente em custos de SA. Após o dimensionamento correto da locação compartilhada, eles reduzem os custos para USD 197.000 com 140 vCPUs. Os hosts dedicados do Amazon EC2 reduzem ainda mais os custos para 151.000 USD (uma redução de aproximadamente 56%).

Implantações altamente disponíveis do SQL Server

Este exemplo analisa como o custo pode influenciar uma implantação do SQL Server AWS com várias considerações de licenciamento. Suponha que uma organização precise implantar seis servidores SQL Server Enterprise AWS para oferecer suporte a três aplicativos. Esses servidores exigem alta disponibilidade e têm 16 vCPUs e 256 GB de RAM cada. Veja os detalhes do cenário a seguir:

- Servidor — SQL Server
- Edição do sistema operacional — Windows Server Datacenter 2019
- Edição SQL Server — SQL Server Enterprise 2019
- vCPU — 16
- Memória (GB) — 256

- Quantidade — 6

Para otimizar os custos AWS sem sacrificar o desempenho, recomendamos que você dimensione corretamente as instâncias com base na utilização de CPU, memória, rede e disco (IOPS/BW). Depois de dimensionar corretamente as cargas de trabalho, coloque-as no tipo de instância x2iedn.4xlarge, que oferece 16 vCPUs. No entanto, esse tipo de instância também inclui o dobro da memória necessária para as cargas de trabalho. Ainda é possível otimizar ainda mais.

Cenário 1

Uma organização implanta seis servidores SQL Server Enterprise em locação AWS compartilhada usando a opção de licença incluída para Windows e SQL Server. Com essa opção, o custo das licenças do Windows e do SQL Server é incorporado ao preço da instância. Veja os detalhes do cenário a seguir:

- Locação compartilhada (instância) — x2iedn.4xlarge
- Custo por hora (USD) — \$10,0705
- Custo mensal por unidade (USD) — \$7.351,47
- Número de servidores — 6
- CPU — 16
- Memória — 512
- Custo mensal para 6 servidores — \$44.108

Cenário 2

Uma organização tem SA e BYOL para SQL Server em locação compartilhada. Isso significa que a organização usa a opção de licença incluída para Windows, mas fornece suas próprias licenças do SQL Server com base no número de vCPUs alocados para a instância. Como a organização tem seis servidores SQL Server Enterprise com 16 vCPUs cada, é necessário um total de 96 vCPUs. Veja os detalhes do cenário a seguir:

- Locação compartilhada (instância) — x2iedn.4xlarge
- Custo por hora (USD) — \$4.0705
- Custo mensal por unidade (USD) — \$2971,47
- Número de servidores — 6
- CPU — 16

- Memória — 512
- Núcleos BYOL — 96
- Custo mensal para 6 servidores — \$17.828

Ao trazer suas próprias licenças do SQL Server com o SA, a organização nesse cenário pode obter economia de custos em comparação com o uso da opção de licença incluída para o SQL Server. A economia precisa de custos depende dos preços e dos termos específicos do contrato de licenciamento. Nesse cenário, AWS os custos diminuem em USD 26.280 por mês ao trazer licenças do SQL Server Enterprise para o. AWS

Cenário 3

Uma organização tem BYOL para Windows e SQL Server em hosts dedicados do Amazon EC2. Isso significa que a organização atribuirá licenças no nível do núcleo físico, permitindo que licenciem somente os núcleos físicos do host. O licenciamento no nível do núcleo físico permite que você implante o número máximo de instâncias sem afetar as licenças necessárias. Esse modelo de licenciamento é comumente usado com o Windows Server Datacenter e a edição SQL Server Enterprise.

Esse cenário usa dois hosts dedicados X2iezn Amazon EC2. Cada host tem 24 núcleos físicos e 48 vCPUs. Isso fornece capacidade adequada para os seis servidores SQL Server Enterprise com 16 vCPUs e 256 GB de RAM cada. Veja os detalhes do cenário a seguir:

- Número de hosts dedicados — 2
- Família de instâncias — x2iezn
- Custo por hora (USD) — \$11.009
- Custo mensal por unidade (USD) — \$8.036
- Núcleo físico — 48
- vCPU disponível — 96
- São necessárias licenças básicas do Windows Server — 24
- Licenças necessárias para os núcleos do SQL Server Enterprise — 24
- Custo mensal — 16.073

O custo total de dois hosts dedicados Amazon EC2 da família X2iEZN é de 16.073 USD por mês. Para obter mais informações sobre preços, consulte a AWS Pricing Calculator [estimativa](#) para esse

cenário. Nesse cenário, a organização pode economizar 1.755,65 USD por mês trazendo suas licenças do Windows. Se eles usarem hosts dedicados do Amazon EC2, também poderão reduzir o número de licenças necessárias do SQL Server. Em locação compartilhada, eles precisariam de 96 licenças do SQL Server Enterprise para cobrir os seis servidores SQL Server Enterprise com 16 vCPUs cada. No entanto, usando os hosts dedicados do Amazon EC2 e o licenciamento no nível do núcleo físico, eles podem reduzir o número de licenças necessárias para 48 núcleos.

Os detalhes a seguir comparam os custos do exemplo 3 e mostram quanto você pode economizar implantando cargas de trabalho em hosts dedicados do Amazon EC2 com a opção BYOL em comparação com outros cenários.

- Servidor local — SQL Server
- vCPU — 16
- Memória — 256
- Número de servidores — 6
- Custo mensal do cenário 1: Windows (LI) + SQL Server Enterprise (LI) — \$44.108
- Custo mensal do cenário 2: Windows (LI) + SQL Server Enterprise (BYOL) — \$17.828
- Custo mensal para o cenário 3: Windows (LI) + SQL Server Enterprise (BYOL) no host dedicado Amazon EC2 — USD 16.073

Note

O custo é baseado em preços sob demanda. Você pode reduzir ainda mais os custos usando Savings Plans ou Instâncias Reservadas Dedicadas. Essas opções oferecem um modelo de preços flexível com economias de custo significativas em comparação com preços sob demanda. Com esses planos, você pode se comprometer com um prazo de um ou três anos. Para obter mais informações, consulte a seção [Otimizar gastos para Windows no Amazon EC2](#) deste guia.

Considere as seguintes opções de pagamento para hosts dedicados do Amazon EC2:

- [Hosts dedicados](#) (documentação do Amazon EC2)
- [Reservas de anfitriões dedicados](#) (documentação do Amazon EC2)
- [Savings Plans](#) (documentação do Amazon EC2)

O [AWS Pricing Calculator](#) agora oferece suporte aos preços de host dedicado. Isso pode ajudar você a escolher o host dedicado subjacente apropriado.

Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando o AWS Cost Explorer:

1. [Ative o Cost Explorer](#).
2. Use o Cost Explorer para [visualizar e analisar os custos e o uso](#) de suas implantações de host dedicado do Amazon EC2.
3. Valide que você está executando o BYOL. Você pode exibir os seguintes detalhes da plataforma e valores de operação de uso nas instâncias ou nas páginas da AMI no console do Amazon EC2 ou na resposta retornada pelo comando `describe-images` ou `ordescribe-instances`.
 - Detalhes da plataforma: Windows, Operação de uso ::0002 RunInstances (Licença incluída)
 - Detalhes da plataforma: Windows BYOL, Operação de uso: :0800 RunInstances

Recursos adicionais

- [Tipos de licença elegíveis para conversão do tipo de licença](#) (AWS License Manager documentação)
- [AWS License Manager e workshop dedicado para anfitriões](#) (AWS License Manager Workshop)
- [Perguntas frequentes sobre hosts dedicados do Amazon EC2 \(documentação da AWS\)](#)
- [Como criar AMIs Bring-Your-Own-License do Windows Server a partir do local com o VM Import/Export \(Microsoft Workloads no blog\)](#) AWS
- [VM Import/Export \(documentação\)](#) AWS
- [Amazon Web Services e Microsoft: perguntas frequentes](#) (AWS documentação)
- [Conversões de tipo de licença no License Manager](#) (AWS License Manager documentação)
- [Implantação de SQL Server de alta disponibilidade em hosts dedicados do Amazon EC2](#) (blog de operações e migrações AWS na nuvem)

Otimize os gastos com Windows no Amazon EC2

Visão geral

Uma das principais preocupações sobre a migração de servidores para lá AWS são os custos de infraestrutura. É verdade que um dos benefícios da nuvem é pagar pelos recursos sob demanda, mas há cargas de trabalho de produção que precisam estar disponíveis 24 horas por dia, 7 dias por semana, 365 dias por ano. Os [Savings Plans](#) foram projetados para economizar dinheiro em seu AWS uso estável em todas as instâncias do EC2, e. AWS Lambda AWS Fargate

Os Savings Plans oferecem um modelo de preços flexível e podem ajudá-lo a reduzir os preços do uso do Amazon EC2, Fargate, Lambda e SageMaker Amazon em troca do compromisso com uma quantidade consistente de uso (por exemplo, \$10/hora). Você se compromete com uma quantidade consistente de gastos de computação por hora em um ou três anos e, em troca, recebe um desconto por esse uso.

Você pode escolher entre três opções de pagamento diferentes com Savings Plans:

- A opção Sem pagamento adiantado não exige nenhum pagamento adiantado, e seu compromisso é cobrado apenas mensalmente.
- A opção Partial Upfront oferece preços mais baixos em Savings Plans. Você paga antecipadamente pelo menos metade do seu compromisso e o restante é cobrado mensalmente.
- A opção All Upfront oferece os preços mais baixos e todo o seu compromisso é cobrado em um único pagamento.

Você pode acompanhar as expirações dos seus Savings Plans e os próximos Savings Plans em fila. AWS Cost Explorer Você pode usar os alertas do Savings Plans para receber alertas antecipados por e-mail 1, 7, 30 ou 60 dias antes da data de expiração do seu plano ou quando um compromisso estiver na fila para compra. Essas notificações também alertam você sobre a data de expiração. Você pode enviar notificações para até 10 destinatários de e-mail.

Conceitos básicos do Savings Plans

Todo tipo de uso de computação tem uma taxa sob demanda e uma taxa de Savings Plans. Se você se comprometer com \$10/hora de uso de computação, obterá preços de Savings Plans em todo o uso de até \$10 à taxa de Savings Plans. Qualquer uso além do compromisso de gastos com computação é cobrado de acordo com taxas regulares sob demanda. Você pode começar a usar o Savings Plans usando o Cost Explorer no AWS Management Console.

Você pode facilmente se comprometer com os Savings Plans usando as recomendações fornecidas no [Cost Explorer](#) para obter a maior economia. O compromisso horário recomendado é baseado em seu histórico de uso sob demanda e em sua escolha de tipo de plano, duração do prazo e opção de pagamento. O Savings Plans é aplicado primeiro à conta que comprou o plano e, em seguida, é compartilhado com outras contas da família de faturamento consolidado.

Note

A opção de compartilhamento do Savings Plans AWS Organizations está ativada por padrão. Você pode recusar essa opção no AWS Billing console da conta do pagador. Você pode visitar sua página de [Recomendações](#) para ver os Savings Plans que AWS recomendam ajudá-lo a economizar no uso qualificado. Essas recomendações podem ser atualizadas a qualquer momento para facilitar a compra dos melhores Savings Plans.

Savings Plans para computação

Os Compute Savings Plans oferecem a maior flexibilidade e ajudam a reduzir seus custos. Esses planos se aplicam automaticamente ao uso da instância EC2, independentemente da família, tamanho, zona de disponibilidade, região, sistema operacional ou localização da instância. Eles também se aplicam ao uso do Fargate ou do Lambda. Por exemplo, com o Compute Savings Plans, você pode mudar de instâncias C4 para M5, transferir uma carga de trabalho da UE (Irlanda) para a UE (Londres) ou mover uma carga de trabalho do EC2 para Fargate ou Lambda a qualquer momento. Você continua pagando automaticamente o preço do Savings Plans.

Savings Plans para instâncias do EC2

Os EC2 Instance Savings Plans oferecem os maiores descontos em troca do compromisso com o uso de famílias de instâncias individuais em uma região (por exemplo, comprometendo-se com um nível consistente de uso de M5 no Norte da Virgínia). Isso fornece automaticamente descontos no preço sob demanda da família de instâncias selecionada nessa região, independentemente da zona de disponibilidade, tamanho, sistema operacional ou localização. Os EC2 Instance Savings Plans permitem que você altere seu uso entre instâncias dentro de uma família nessa região. Por exemplo, você pode passar de c5.xlarge executando Windows para c5.2xlarge executando Linux e se beneficiar automaticamente dos preços do Savings Plans.

Os planos de economia de instância de computação e EC2 se aplicam às instâncias do EC2 que fazem parte dos clusters Amazon EMR, Amazon Elastic Kubernetes Service (Amazon EKS) e

Amazon Elastic Container Service (Amazon ECS). As cobranças do Amazon EMR, do Amazon EKS e do Amazon ECS não são cobertas pelos Savings Plans, mas as instâncias EC2 subjacentes sim. Os Savings Plans para instâncias do EC2 são aplicados antes dos Savings Plans para computação porque estes têm uso mais amplo.

Note

Você não pode alterar um Savings Plan facilmente depois de assumir um compromisso. Recomendamos que você planeje cuidadosamente antes de se comprometer com qualquer uma das opções do Savings Plans. Os Savings Plans oferecem preços mais baixos em comparação aos preços sob demanda em troca de um compromisso e não podem ser cancelados durante o período.

Exemplo de compromisso por hora

Se você comprar um Savings Plan, você assume um compromisso monetário de hora em hora com a vigência do plano. Se você se comprometer com \$10/hora de uso de computação, o preço do Savings Plan será aplicado automaticamente a todo o uso, até \$10 dólares por hora. Qualquer uso além do compromisso é cobrado de acordo com as tarifas regulares sob demanda. Você pode usar a ferramenta de recomendações de compra do Savings Plans no Cost Explorer para obter compromissos recomendados que possam maximizar sua economia. O compromisso financeiro por hora de um plano específico não pode ser modificado durante a vigência do plano. Se você quiser um maior comprometimento após analisar o uso, poderá adquirir um Savings Plan adicional para cobrir o uso excessivo.

Benefícios dos Savings Plans

Em comparação com as Instâncias Reservadas, os Savings Plans oferecem um modelo de preços mais flexível que pode economizar dinheiro enquanto você aproveita a seleção mais ampla de opções de computação oferecidas pelos Savings Plans. Os Savings Plans oferecem descontos, mesmo quando suas necessidades de computação mudam. Isso pode ajudá-lo a acompanhar seu ambiente dinâmico em constante mudança sem incorrer em nenhuma sobrecarga adicional de gerenciamento. Aqui estão alguns outros benefícios de usar o Savings Plans:

- Fácil de usar — Receba descontos automáticos em troca de um compromisso monetário.
- Flexibilidade — um compromisso único que se aplica a vários tipos de uso.
- Economia potencial — Há várias maneiras de economizar. Considere os seguintes exemplos:

- 60% de economia nas cargas de trabalho do Windows Server usando Compute Savings Plans ([d2.8xlarge, 3 anos, tudo adiantado, windows, locação compartilhada, us-east-2](#))
- 73% de economia nas cargas de trabalho do Windows Server usando EC2 Instance Savings Plans ([d2.8xlarge, 3 anos, all upfront, windows, shared tenancy, us-east-2](#))
- Economia de 28 a 41% em tipos de instância não exóticos ([família t3, 3 anos, tudo adiantado, windows, locação compartilhada, us-east-2](#))
- Economia média de 25 a 40% para servidores Windows

Note

Os EC2 Instance Savings Plans oferecem um desconto maior do que os Compute Savings Plans devido à flexibilidade reduzida. Você se compromete com o uso por um preço com desconto.

Todo tipo de uso de computação tem uma taxa do Savings Plan e uma taxa sob demanda. A tabela a seguir mostra os Savings Plans e as taxas sob demanda para cada tipo de sistema operacional. Você paga as tarifas do Savings Plans sobre o uso comprometido e qualquer uso além do compromisso é cobrado de acordo com as tarifas regulares sob demanda.

Nome da instância	Taxa de Savings Plans	Economia sob demanda	Tarifa sob demanda	Sistema operacional	Região	Opção de pagamento	Duração do mandato
x2iedn.xlarge	\$0,32	61%	\$0,83	Linux	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3
x2iedn.xlarge	\$2,01	50%	\$1,02	Windows	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3

Nome da instância	Taxa de Savings Plans	Economia sob demanda	Tarifa sob demanda	Sistema operacional	Região	Opção de pagamento	Duração do mandato
x2iedn.xlarge	\$1,02	20%	\$2,52	Licença Windows incluída + edição SQL Server Enterprise	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3
x2iedn.xlarge	\$0,32	61%	\$0,83	BYOL	Leste dos EUA (Norte da Virgínia)	Sem taxas iniciais	3

Os Savings Plans incluem o sistema operacional e têm um desconto separado para BYOL. Todos eles estão detalhados na calculadora do [Compute Savings Plans](#).

Modelo de preços de instância reservada

AWS tem outro modelo de preços baseado no compromisso conhecido como Instâncias Reservadas. Esse modelo pode ser problemático se sua computação mudar depois que você já se comprometeu, fazendo com que as instâncias reservadas não sejam usadas. Os Savings Plans foram projetados para oferecer reduções de custo semelhantes às [instâncias reservadas padrão e conversíveis](#), mas com muito mais flexibilidade. Os Compute Savings Plans oferecem preços mais baixos no uso de instâncias EC2, independentemente da família, tamanho, sistema operacional, localização ou região da instância. Eles também permitem a máxima flexibilidade.

A tabela a seguir pode ajudar você a escolher entre Savings Plans ou Reserved Instances.

	Reserved Instance	Savings Plans para instâncias do EC2	Savings Plans para computação
Desconto médio de 1 ano	Até 38%	Até 29%	Até 29%
Desconto médio de 3 anos	Até 58%	Até 73%	Até 60%
Família de instâncias	Fixed	Fixed	Flexível
Tamanho da instância	Fixo (não Linux)	Flexível	Flexível
Geografia	1 região	1 região	Flexível
Sistema operacional	Fixed	Flexível	Flexível
Serviço	Amazon EC2 ou Amazon RDS	Amazon EC2	Amazon EC2, Fargate, Lambda
Opções de pagamento	Tudo, parcial, sem adiantamento	Tudo, parcial, sem adiantamento	Tudo, parcial, sem adiantamento
Limites de instâncias	20 por zona de disponibilidade	Sem limite	Sem limite

Note

Os Savings Plans funcionam oferecendo um desconto com base em um compromisso monetário por hora. O compromisso financeiro por hora não pode ser cancelado ou alterado durante a vigência do seu plano, mas você pode comprar Savings Plans adicionais para cobrir o uso adicional. Isso permite que você mantenha um compromisso horário consistente à medida que sua frota cresce.

Você pode usar ferramentas como [AWS Cost Explorer](#) e [Nuvem AWS Intelligence Dashboards](#) para monitorar seu compromisso. O Cost Explorer fornece uma meta de cobertura que pode ajudar sua organização a planejar sua estratégia de cobertura do Savings Plans. Se 75 por cento da sua

carga de trabalho estiver estável, então 75 por cento é uma boa meta. Isso deixa 25% dos gastos sob demanda/variáveis com base em cargas de trabalho dinâmicas. Se precisar aumentar essa cobertura para 85 por cento, você pode comprar outro compromisso da Savings Plans para aumentar o compromisso monetário por hora.

Note

Recomendamos que você compre Savings Plans em vez de Instâncias Reservadas, mas os dois modelos de compromisso podem funcionar juntos se você já comprou Instâncias Reservadas.

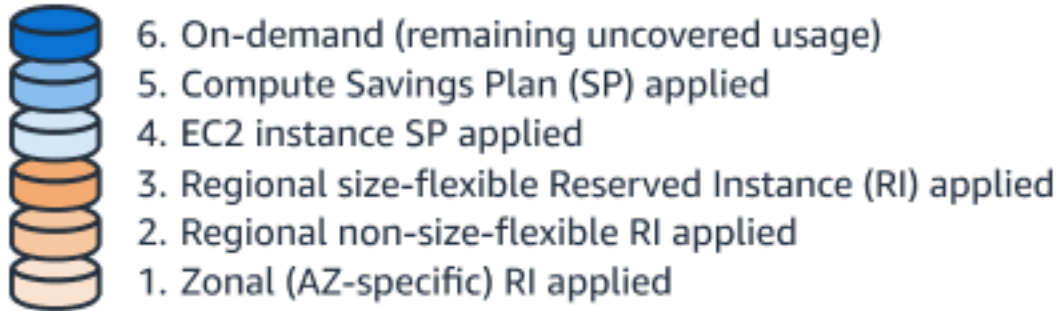
Considere um exemplo em que você comprou uma Instância Reservada, mas quer começar a experimentar a opção Savings Plans. Há uma lógica para que essa combinação se aplique ao seu faturamento final. Aqui está uma hierarquia que você pode aplicar ao seu Contas da AWS:

1. A Instância Reservada Zonal se aplica à conta que a possui. Se uma instância reservada tiver horas restantes, ela se aplicará ao resto da organização.
2. As instâncias reservadas regionais não flexíveis para Windows se aplicam ao uso correspondente na conta proprietária. Tudo o que resta é transferido para o resto da organização.
3. As instâncias reservadas regionais com tamanho flexível se aplicam à conta proprietária (primeiro a menor instância dentro da família e subindo para instâncias maiores) e depois para o resto da organização.
4. As instâncias reservadas regionais se aplicam a qualquer reserva de capacidade sob demanda não utilizada.
5. Os EC2 Instance Savings Plans se aplicam à conta que os comprou.
6. Os Compute Savings Plans se aplicam à conta que os comprou.

Note

Os descontos começam com o uso que resulta no maior desconto e depois descem até o menor desconto. As instâncias do Windows tradicionalmente têm um potencial de desconto menor do que o Linux para os tipos de instância mais comuns (por exemplo, T3, M6 e C5). Isso significa que as instâncias Linux se beneficiam mais do que as instâncias do Windows na maioria dos casos.

O gráfico a seguir mostra o preço depois de dividir as Instâncias Reservadas das Savings Plans. Tanto o Compute quanto o EC2 Instance Savings Plans se aplicam primeiro às instâncias em execução e depois às reservas de capacidade sob demanda não utilizadas.



Cenários de otimização de custos

Esta seção aborda cenários de otimização de custos para hosts dedicados do Amazon EC2 e instâncias do Amazon EC2 que usam um modelo de cobrança com licença incluída.

Hosts dedicados do Amazon EC2

Considere um cenário para o qual você migrará suas cargas de trabalho locais do Windows. AWS Seu data center tem os seguintes servidores:

- Dois servidores com 16 vCPU e 128 GB de RAM
- Dois servidores com 32 vCPU e 164 GB de RAM
- Um servidor com 8 vCPUs e 64 GB de RAM
- 16 servidores com vCPU e 32 GB de RAM

Além disso, suponha que você possa trazer sua própria licença AWS porque tem licenças suficientes para trazer. A tabela a seguir mostra as instâncias do servidor que você pode usar em AWS.

Tipo de instância	CPU	RAM	Valor
r5.4xlarge	16	128	2
r5.8xlarge	32	256	2
r5.2xlarge	8	64	1

Tipo de instância	CPU	RAM	Valor
r5.xlarge	4	32	16
			21

Uma análise mostra que essas 21 máquinas virtuais podem ser distribuídas em dois hosts dedicados com um host da família de instâncias R5. A tabela a seguir mostra o custo desses dois hosts dedicados.

Cenário de host dedicado sob demanda	Pagamento adiantado	1 mês	1 ano	3 anos	AWS Pricing Calculator
Sob demanda	Nenhum	\$10.123	\$121.475	\$364.392	AWS Pricing Calculator estimativa
Savings Plan de 1 ano	Nenhum	\$7.447	\$89.362	–	AWS Pricing Calculator estimativa
Savings Plan de 3 anos	Nenhum	\$5.476	\$65.712	\$197.128	AWS Pricing Calculator estimativa
Savings Plan de 3 anos com pagamento adiantado	\$84.438	\$2.755	\$117.499	\$183.618	AWS Pricing Calculator estimativa

Se você tiver os servidores para os quais deseja migrar AWS, o preço final de um Savings Plan de 1 ano é de \$89.362 em vez de \$121.475 para um preço sob demanda. Isso representa um desconto de 26,5 por cento após um ano. Se você está pensando em ficar em casa AWS por um período mais

longo, pode escolher o Savings Plan de 3 anos para economizar ainda mais. Ao final de três anos, você paga \$197.128 em vez de \$364.392. Isso resulta em uma economia de 46% do valor total após três anos.

Instâncias do Amazon EC2 com licenças incluídas

Considere um cenário em que você migrará um único aplicativo de três camadas para AWS e desejará usar as licenças fornecidas pelo. AWS Além disso, suponha que seu aplicativo funcione com os seguintes servidores:

- Dois servidores web com duas vCPUs e 4 GB de RAM
- Dois servidores de aplicativos com oito vCPUs e 16 GB de RAM
- Dois servidores de bancos de dados com 16 vCPUs e 64 GB de RAM (usando a edição SQL Server Standard)

A tabela a seguir mostra as instâncias do servidor que você pode usar em AWS.

Tipo de instância	CPU	RAM	Valor
c5.large	2	4	2
c5.2xlarge	8	16	2
r5.2xlarge	8	64	2
			6 servidores

A tabela a seguir mostra o custo desses servidores em AWS.

Licença incluída por AWS	Pagamento adiantado	1 mês	1 ano	3 anos	AWS Pricing Calculator
Sob demanda	Nenhum	\$3.912	\$46.950	\$140.849	AWS Pricing Calculator estimativa

Licença incluída por AWS	Pagamento adiantado	1 mês	1 ano	3 anos	AWS Pricing Calculator
Savings Plan de 1 ano	Nenhum	\$3.466	\$41.952		AWS Pricing Calculator estimativa
Savings Plan de 3 anos sem pagamento adiantado	Nenhum	\$3.189	\$38.264	\$114.804	AWS Pricing Calculator estimativa
Savings Plan de 3 anos com pagamento adiantado	\$112.110	Nenhum	Nenhum	Nenhum	AWS Pricing Calculator estimativa

Se você quiser executar esses servidores em ambientes de produção (24 horas por dia, 7 dias por semana) com preços sob demanda, pagará um custo mensal de \$3.912. Pagar esse custo mensal equivale a \$46.950 após um ano e um total de \$140.849 após três anos.

Se você escolher o Savings Plan de 1 ano sem pagamento adiantado, o custo mensal diminuirá para \$3.466. No final do primeiro ano, você paga \$41.952. Isso é um desconto total de 11 por cento. Se você escolher o Savings Plan de 3 anos sem pagamento adiantado, o custo mensal diminuirá para \$3.189. Ao final de três anos, você paga \$114.804. Isso proporciona uma economia de 18,5 por cento.

Recomendações de otimização de custos

Ambos os cenários ajudam você a economizar dinheiro ao planejar e prever suas cargas de trabalho. AWSÉ importante reconhecer que o desconto no segundo cenário é menor em comparação com o primeiro cenário. No segundo cenário, o preço da licença está incluído no preço do servidor em nuvem. AWS não oferece desconto no preço da licença, mas você sempre pode trazer suas licenças (em cenários específicos) e AWS sempre garantir o melhor preço de computação/instância.

Recomendamos que você faça o seguinte para controlar seus AWS gastos com recursos de computação e instância:

- Recomendações de acesso
- Personalize as recomendações de acordo com suas necessidades
- Revise o compromisso por hora

Recomendações de acesso

Você pode usar o [console do Amazon EC2](#) para acessar recomendações para seu Savings Plan. Você pode até mesmo baixar suas recomendações para revisar posteriormente no formato CSV. Para obter mais informações, consulte [Monitorando seus planos de poupança](#) na documentação do Savings Plans.

Personalize as recomendações de acordo com suas necessidades

Abra o [console do Amazon EC2](#), expanda a seção Instâncias e escolha Savings Plans. Esta página mostra os preços de instâncias e computadores antes e depois de fazer uma recomendação. Você também pode ajustar os seguintes fatores para sua recomendação:

- Prazo — Por exemplo, de 1 a 3 anos
- Opção de pagamento — por exemplo, adiantado, adiantado parcial ou sem adiantamento
- Histórico — Por exemplo, os últimos 7, 30 ou 60 dias

Revise o compromisso por hora

Usando o mesmo exemplo, suponha que você tenha uma instância em execução 24 horas por dia, 7 dias por semana. A recomendação é usar um Savings Plan. De acordo com o tamanho, você tem um preço sob demanda de \$120/hora. Você tem a opção de comprometer \$90/hora, mas isso pode variar dependendo da região, instância e opção de compra. Neste exemplo, você pode economizar 25% em comparação com o custo sob demanda. Você também pode monitorar sua utilização e cobertura, se estiverem abaixo do limite definido, e configurar um alerta quando o orçamento terminar.

Revise as recomendações

Recomendamos que você analise cuidadosamente as recomendações do Savings Plan. AWS não mudará nada sem sua permissão. Essas são apenas recomendações e cabe a você aplicá-las ou não.

Adquira um plano

Abra o [console do Amazon EC2](#), expanda a seção Instâncias e escolha Savings Plans. Em seguida, escolha Purchase Savings Plans. Com base em seus requisitos, você pode selecionar as seguintes opções: prazo, região, família de instâncias, compromisso por hora, opção de pagamento e até mesmo data de início. Você pode escolher entre Compute Savings Plans, EC2 Instance Savings Plans e SageMaker Savings Plans. Para obter mais informações, consulte [Purchasing Savings Plans](#) na documentação do Savings Plans.

Obtenha um relatório de utilização

Depois de comprar um Savings Plan, você pode obter um relatório de utilização. O relatório ajuda você a verificar sua utilização, ver se o plano adquirido é suficiente para cobrir e maximizar o desconto e cancelar ou adicionar novos descontos. Esse relatório pode ser exportado para outros formatos, como CSV. Para obter mais informações, consulte [Usando o relatório de utilização](#) na documentação do Savings Plans.

Siga as melhores práticas de compra

Recomendamos que você siga estas melhores práticas antes de comprar Savings Plans:

- Use [AWS Trusted Advisor](#) para remover recursos do EC2 ociosos.
- Faça qualquer dimensionamento correto antes da compra do Savings Plans.
- Estabeleça uma taxa horária que você mantenha de forma consistente por 30 a 60 dias.
- Adquira um compromisso para cobrir a taxa horária consistente com a qual sua organização se sinta confortável. Considere as flutuações na demanda ou na temporada.
- Escolha uma revisão trimestral do orçamento do Savings Plans para manter uma taxa consistente (por exemplo, meta de cobertura de 70% para a cobertura do Savings Plans). Se a tarifa cair abaixo da cobertura desejada, adquira um Savings Plan adicional como ajuste para atingir sua meta de cobertura.

Recursos adicionais

- [Savings Plans para instâncias reservadas do Amazon EC2](#) (AWS whitepapers)
- [Entendendo como os Savings Plans se aplicam ao seu AWS uso](#) (documentação do Savings Plans)
- [Anúncio do faturamento por segundo para instâncias EC2 do Windows Server e do SQL Server](#) (documentação)AWS
- [AWS Série de otimização de custos: Vídeo da Savings Plans | Amazon Web Services](#) () YouTube

Monitore os custos usando AWS ferramentas

Visão geral

A visibilidade dos custos é um fator-chave para otimizar os custos. AWS tem várias ferramentas que você pode usar para visualizar custos e criar alertas em reação a esses custos. Isso inclui ferramentas que ajudam você a monitorar e relatar seus gastos. AWS Budgets Esta seção aborda formas específicas de monitorar os AWS gastos do Windows, para que você possa acompanhar e reagir de acordo com seus requisitos de orçamento. Isso inclui adicionar as tags necessárias aos recursos do Windows EC2. Essas tags permitem que você monitore adequadamente o Windows EC2 e outros serviços da Microsoft usando AWS Budgets.

Ao monitorar gastos e criar alertas com AWS ferramentas, você pode estar mais informado sobre gastos atuais, gastos projetados e anomalias de gastos. Se você usa [Savings Plans](#) para ajudar a reduzir o preço de sua instância EC2 por hora, recomendamos que você veja a utilização geral e a cobertura do Savings Plan. Isso pode ajudá-lo a garantir que você esteja continuamente economizando. Você pode usar AWS Cost Explorer para visualizar o inventário do Savings Plan e obter recomendações de Savings Plans adicionais com base no uso anterior. Você também pode monitorar gastos específicos usando [AWS Budgets](#) e configurando [AWS Cost Anomaly Detection](#).

Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para otimizar seus custos usando AWS Budgets o Cost Explorer e a detecção de anomalias:

- Marcar recursos do Windows EC2
- Configure alertas usando AWS Budgets
- Habilite a detecção de anomalias de custo

- Obtenha uma análise de gastos em tempo real
- Veja os gastos com a licença incluída para Windows usando o Cost Explorer

Marcar recursos do Windows EC2

Para monitorar seus AWS gastos com eficiência, você deve estabelecer uma [estratégia de marcação](#) para as cargas de trabalho que deseja monitorar. Isso é importante para que você possa agrupar recursos categoricamente e ser notificado sobre gastos específicos, em vez de gastos de uso geral. Você pode usar recursos de marcação que não apenas ajudam no custo, mas também podem ser usados para outros fins, como [AWS Systems Manager automação](#). Além disso, recomendamos que você implemente algum gerenciamento para as [tags necessárias](#).

Para monitorar seus gastos no AWS Budgets Cost Explorer e no Cost Anomaly Detection, você deve garantir que as tags adequadas estejam em vigor. Você pode usar tags para configurar um orçamento específico para itens que correspondam a essas tags, para que você seja alertado quando os gastos aumentarem.

Por exemplo, você pode usar uma tag simples como Key=OS Value=Windows. Isso reúne todas as suas instâncias do Windows em um grupo para o qual você pode monitorar os gastos. Você também pode usar tags para outros itens, como Systems Manager. Depois de criar uma tag, você deve ativar a tag para controle de custos. Considere adicionar uma [AWS Config regra que monitore as tags](#) anexadas a determinados recursos. AWS Config pode alertá-lo se houver recursos em execução que não contenham as tags apropriadas, que fornecem uma representação precisa de seus gastos com o Windows EC2.


Depois de colocar suas tags, você pode criar um orçamento personalizado em AWS Billing. Isso fornece visibilidade de seus gastos com o Windows EC2. Você pode definir um orçamento diário ou mensal.

Configure alertas usando AWS Budgets

Neste cenário de exemplo, você cria um orçamento diário para o Windows EC2. É um orçamento recorrente que usa a opção de ajuste automático para monitorar seus gastos e ajustar o orçamento adequadamente. Se você tiver um ambiente estático, poderá usar um orçamento fixo em vez disso. Certifique-se de escolher um intervalo de tempo básico (por exemplo, 30 dias).

1. Faça login no AWS Management Console e abra o [AWS Cost Management console](#).
2. No painel de navegação, selecione Orçamentos.

3. Na parte superior da página, escolha Criar orçamentos.
4. Em Configuração de orçamento, escolha Personalizar (avançado).
5. Em Tipos de orçamento, escolha Orçamento de custo. Em seguida, escolha Próximo.
6. Em Detalhes, em Nome do orçamento, insira o nome do seu orçamento. Por exemplo, gastos com o Windows EC2.
7. Em Definir valor do orçamento, em Período, escolha Diariamente.
8. Em Tipo de renovação do orçamento, escolha Orçamento recorrente para um orçamento que é redefinido após o período orçamentário.
9. Em Data de início, escolha a data ou o período de início para começar a rastrear seu valor orçado.
10. Em Método de orçamento, escolha Ajuste automático (novo).
11. Em Intervalo de tempo da linha de base, escolha Intervalo personalizado e insira 30 dias.
12. Escolha Próximo.
13. Na seção Escopo do orçamento, selecione Filtrar dimensões de AWS custo específicas. É aqui que as tags são usadas para criar as dimensões adequadas. AWS Budgets não suporta o tipo de plataforma como uma opção em seus filtros. Por esse motivo, você deve aplicar as tags do sistema operacional.
14. Escolha Adicionar filtro e, em seguida, selecione a opção Tag em Dimensões.
15. Escolha a tag do sistema operacional e, em seguida, escolha o valor do Windows para criar um orçamento para a tag.
16. Escolha Próximo.
17. Na página Configurar alertas, escolha Adicionar um limite de alerta. Aqui você configura dois alertas: um para um limite de 50 por cento e outro para um limite de 100 por cento. Se o alerta de limite de 50% for violado antes da metade do mês, ele fornecerá um aviso. Dessa forma, você pode verificar se seus gastos estão acima do esperado e reagir antes de chegar ao final do mês.
18. Em Limite, insira 50 e selecione % do valor orçado.
19. Em Trigger, escolha Real.
20. Para destinatários de e-mail, insira um endereço de e-mail. Adicione outro alerta para um limite de 100.

 Note

Este exemplo usa uma simples notificação por e-mail para o alerta, mas você também pode usar o [Amazon Chime ou](#) o Slack.

Habilite a detecção de anomalias de custo

Você pode usar suas etiquetas de custo para configurar alertas de gastos que sejam uma anomalia. Por exemplo, você pode usar [AWS Cost Anomaly Detection](#) para criar monitores para seus gastos e ser alertado quando o sistema detectar gastos anormais em sua conta.

Para configurar um monitor e alertas para as tags Key=OS e Value=Windows que você criou anteriormente, faça o seguinte:

1. Faça login no AWS Management Console e abra o [AWS Cost Management console](#).
2. No painel de navegação, escolha Detecção de anomalias de custo.
3. Escolha a guia Monitores de custo e, em seguida, escolha Criar monitor.
4. Na Etapa 1, escolha a etiqueta de alocação de custos como seu tipo de monitor.
5. Em Chave de etiqueta de alocação de custos, escolha Gastos do Windows EC2.
6. Em Valor da etiqueta de alocação de custos, escolha Windows.
7. Em Nomear seu monitor, insira Gastos do Windows EC2.
8. Escolha Próximo.
9. Para criar uma assinatura para os alertas, selecione Criar uma nova assinatura. Se você tiver assinaturas existentes, selecione Escolher uma assinatura existente.
10. Em Nome da assinatura, insira Windows EC2 spend anomaly.
11. Em Frequência de alertas, escolha Resumos diários.
12. Para destinatários do alerta, insira seu endereço de e-mail.
13. Escolha Adicionar limite. Em Limite, insira 10 e selecione a porcentagem acima da velocidade esperada.
14. Escolha Criar monitor.

Tenha uma visão em tempo real dos gastos

Um alerta é uma ferramenta útil para monitorar seus gastos com o Windows EC2, mas você deve usar o Cost Explorer se quiser uma visão em tempo real dos gastos. Assista a este vídeo para saber como o Cost Explorer permite que você analise e reduza seus custos do EC2. Para obter mais informações, assista ao vídeo [AWS Supports You | Understanding and Reducing Your EC2 Costs](#) em YouTube.

Veja os gastos com licença incluída para Windows

Você pode ver os gastos do EC2 Windows em sua conta usando o Cost Explorer. Para ver os gastos incluídos na licença para Windows, você deve definir os seguintes [filtros](#) corretos no Cost Explorer:

- Em Plataforma, escolha Windows (Amazon VPC). Para operação de API, escolha:0002RunInstance. Esse é o AWS Billing código para instâncias do Windows EC2 incluídas na licença.
- Se você quiser ver seus gastos com a instância BYOL, RunInstance altere:0002 para:0800. RunInstance Esse é o código de cobrança do Windows EC2 BYOL.

Com essa visibilidade no Cost Explorer, você pode filtrar rapidamente seus custos para exatamente o que você está gastando no Windows EC2. Se você quiser se aprofundar ainda mais em seus AWS gastos, use AWS Cost and Usage Report para filtrar os gastos no nível de instância individual. Você também pode gerar relatórios que podem ser visualizados na Amazon QuickSight e criar painéis personalizados.

Para obter mais informações, assista ao [vídeo AWS Apoia você - Visualizando seus relatórios de custo e uso](#) em YouTube.

Recursos adicionais

- [Configurando as tags necessárias com AWS Config](#) (AWS Config documentação)
- [AWS Budgets Tutorial - Configurar alertas para AWS Billing | Amazon Web Services](#) (YouTube)
- [AWS Cost and Usage Report Biblioteca de consultas](#) (AWS Well-Architected Labs)

SQL Server

Os clientes executam cargas de trabalho da Microsoft AWS há mais de 15 anos, mais do que qualquer outro provedor de nuvem. Isso ocorre principalmente porque AWS tem a maior experiência com aplicativos da Microsoft na nuvem e oferece a melhor plataforma para Windows Server e Microsoft SQL Server nas seguintes áreas:

- Maior desempenho e confiabilidade
- Maior segurança e serviços de identidade
- Mais suporte à migração
- Os recursos mais amplos e profundos
- Menor custo total de propriedade (TCO)
- Opções flexíveis de licenciamento

AWS oferece suporte a tudo o que é necessário para criar e executar aplicativos do Windows que dependem do SQL Server, incluindo Active Directory, .NET, SQL Server, área de trabalho do Windows como serviço e todas as versões compatíveis do Windows Server. Com experiência comprovada, AWS pode ajudá-lo a levantar e mudar, refatorar ou até mesmo modernizar suas cargas de trabalho do Windows com facilidade.

Esta seção do guia aborda os seguintes tópicos:

- [Escolha uma solução de alta disponibilidade e recuperação de desastres](#)
- [Entenda o licenciamento do SQL Server](#)
- [Selecione a instância EC2 certa para cargas de trabalho do SQL Server](#)
- [Consolide instâncias](#)
- [Compare as edições do SQL Server](#)
- [Avalie a edição SQL Server Developer](#)
- [Avalie o SQL Server no Linux](#)
- [Otimize as estratégias de backup do SQL Server](#)
- [Modernize os bancos de dados do SQL Server](#)
- [Otimize o armazenamento para SQL Server](#)
- [Otimize o licenciamento do SQL Server usando o Compute Optimizer](#)
- [Otimize o dimensionamento do SQL Server usando o Compute Optimizer](#)

- [Analise Trusted Advisor as recomendações para cargas de trabalho do SQL Server](#)

Escolha uma solução de alta disponibilidade e recuperação de desastres

Visão geral

Recomendamos que você projete uma arquitetura para sua implantação do SQL Server AWS que atenda às suas necessidades de negócios e, ao mesmo tempo, atenda aos [objetivos de recuperação de desastres \(DR\)](#), incluindo seu objetivo de tempo de recuperação (RTO) e objetivo de ponto de recuperação (RPO). As soluções a seguir podem ajudá-lo a projetar a arquitetura certa para o SQL Server na Amazon Elastic Compute Cloud (Amazon EC2) e, ao mesmo tempo, otimizar os custos de suas cargas de trabalho do SQL Server.

- Grupos de disponibilidade do SQL Server Always On — Os grupos de disponibilidade do SQL Server Always On fornecem soluções de alta disponibilidade e recuperação de desastres (HA/DR) para bancos de dados do SQL Server. Um grupo de disponibilidade consiste em um conjunto de bancos de dados de usuários que realizam failover juntos. Os grupos de disponibilidade do Always On também fornecem redundância no nível do banco de dados, mas não exigem armazenamento compartilhado — cada réplica tem seu próprio armazenamento local. Você pode implantar esse recurso como uma solução de HA/DR. Para obter mais informações, consulte [O que é um grupo de disponibilidade Always On?](#) na documentação da Microsoft.
- Instâncias de cluster de failover (FCI) do SQL Server Always On — Os FCIs do SQL Server Always On usam o Windows Server Failover Clustering (WSFC) para fornecer HA no nível da instância do SQL Server. Os FCIs exigem armazenamento compartilhado para hospedar bancos de dados. Você pode usar o armazenamento em bloco compartilhado ou o armazenamento compartilhado de arquivos. Por exemplo, você pode usar o Amazon FSx for Windows File Server ou o Amazon FSx NetApp for ONTAP como uma solução de armazenamento compartilhado com várias zonas de disponibilidade. Para obter mais informações, consulte [Always On Failover Cluster Instances \(SQL Server\)](#) na documentação da Microsoft.
- SIOS DataKeeper — O SIOS DataKeeper pode ajudá-lo a atender aos requisitos de HA e DR ao habilitar uma FCI do SQL Server que abrange as zonas de disponibilidade e. Regiões da AWS O SIOS DataKeeper cria uma SAN virtual em cluster usando volumes locais do Amazon Elastic Block Store (Amazon EBS) e usa a replicação síncrona entre as zonas de disponibilidade para HA, enquanto usa a replicação assíncrona entre regiões e para recuperação de desastres. Para

obter mais informações, consulte [Proteção de alta disponibilidade para aplicativos Windows](#) na documentação do SIOS.

- Grupos de disponibilidade distribuídos — os grupos de disponibilidade distribuídos são um tipo especial de grupo de disponibilidade que se estende por dois grupos de disponibilidade Always On separados. Um grupo de disponibilidade pode residir em duas regiões separadas (por exemplo, us-east-1 e us-west-1). Você pode pensar em um grupo de disponibilidade distribuído como um grupo de disponibilidade de grupos de disponibilidade porque os grupos de disponibilidade Always On subjacentes estão configurados em dois clusters WSFC diferentes. A edição SQL Server Enterprise é necessária para implantar grupos de disponibilidade distribuídos. Para obter mais informações, consulte [Grupos de disponibilidade distribuídos](#) na documentação da Microsoft.
- Envio de registros — Você pode implementar o envio de registros para proteger seus bancos de dados em várias regiões, no caso raro de uma região ser afetada e ficar indisponível. Dependendo da transação e da frequência de envio do registro, você pode obter RPO e RTO em minutos. Para obter mais informações, consulte [Sobre o envio de registros \(SQL Server\)](#) na documentação da Microsoft.
- AWS Elastic Disaster Recovery— O Elastic Disaster Recovery é um aplicativo de software como serviço (SaaS) que gerencia a replicação de servidores de qualquer infraestrutura AWS para fins de DR. Você também pode usar o Elastic Disaster Recovery para replicar o SQL Server em todas as regiões. O Elastic Disaster Recovery é uma solução baseada em agentes que replica máquinas virtuais inteiras, incluindo o sistema operacional, todos os aplicativos instalados e todos os bancos de dados em uma área de armazenamento. Para obter mais informações, consulte [O que é o Elastic Disaster Recovery?](#) na documentação do Elastic Disaster Recovery.
- AWS Database Migration Service (AWS DMS) — AWS DMS suporta a migração ao vivo de dados de e para AWS, incluindo uma região diferente. Você pode usar esse recurso para configurar uma instância separada do SQL Server em uma região diferente para servir como banco de dados de recuperação de desastres. Para obter mais informações, consulte [O que é AWS Database Migration Service?](#) na AWS DMS documentação.

Grupos de disponibilidade do SQL Server Always On

Se você estiver usando a edição SQL Server Enterprise apenas para um [grupo de disponibilidade Always On de alta disponibilidade](#), poderá fazer o downgrade para a edição SQL Server Standard aproveitando os grupos de disponibilidade básicos. Você pode reduzir os custos de 65 a 75 por cento usando grupos de disponibilidade básicos em vez de grupos de disponibilidade Always On.

Note

Para obter informações adicionais sobre diferenças de custo entre diferentes edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

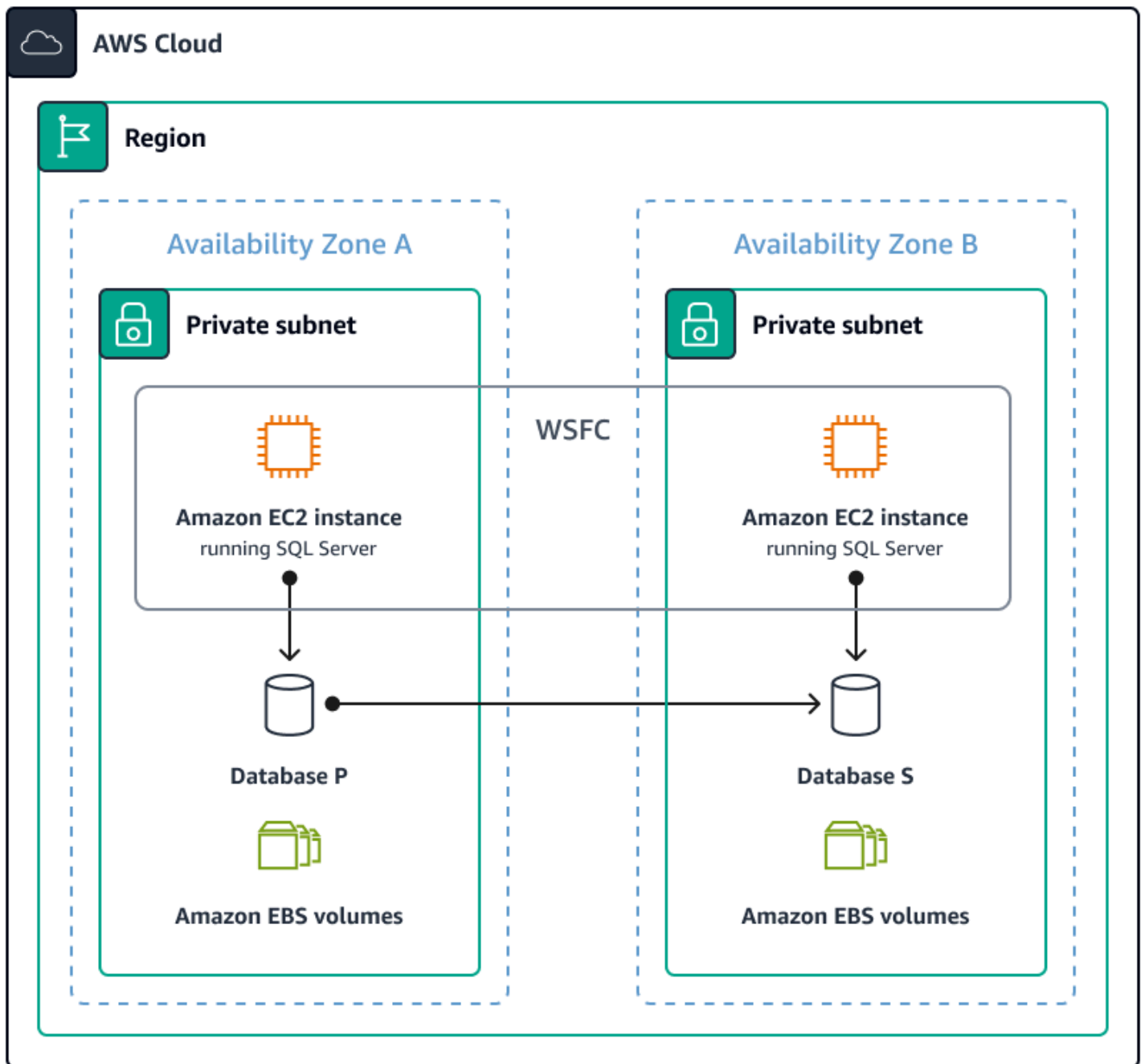
Atributos

- Disponível na edição SQL Server Standard
- Limite de duas réplicas (primária e secundária)
- Sem acesso de leitura na réplica secundária
- Sem verificações de integridade em réplicas secundárias

Limitações

- Support para somente um banco de dados de disponibilidade por grupo de disponibilidade
- Grupos de disponibilidade básica não podem fazer parte de um grupo de disponibilidade distribuído


O diagrama a seguir mostra um exemplo de arquitetura para uma solução de cluster de failover do Windows Server.



Instâncias de cluster de failover do SQL Server Always On

Você pode usar instâncias de cluster de failover (FCIs) para garantir operações contínuas do banco de dados, minimizando o tempo de inatividade e reduzindo o risco de perda de dados. Os FCIs oferecem uma solução confiável se você estiver buscando alta disponibilidade para seu banco de dados SQL Server sem uma configuração de réplica de leitura.

Ao contrário dos grupos de disponibilidade, os FCIs podem fornecer uma solução de failover confiável sem exigir a edição SQL Server Enterprise. Em vez disso, os FCIs exigem apenas o licenciamento da edição SQL Server Standard. Você pode usar os FCIs para reduzir os custos de licenciamento do SQL Server em 65 a 75 por cento.

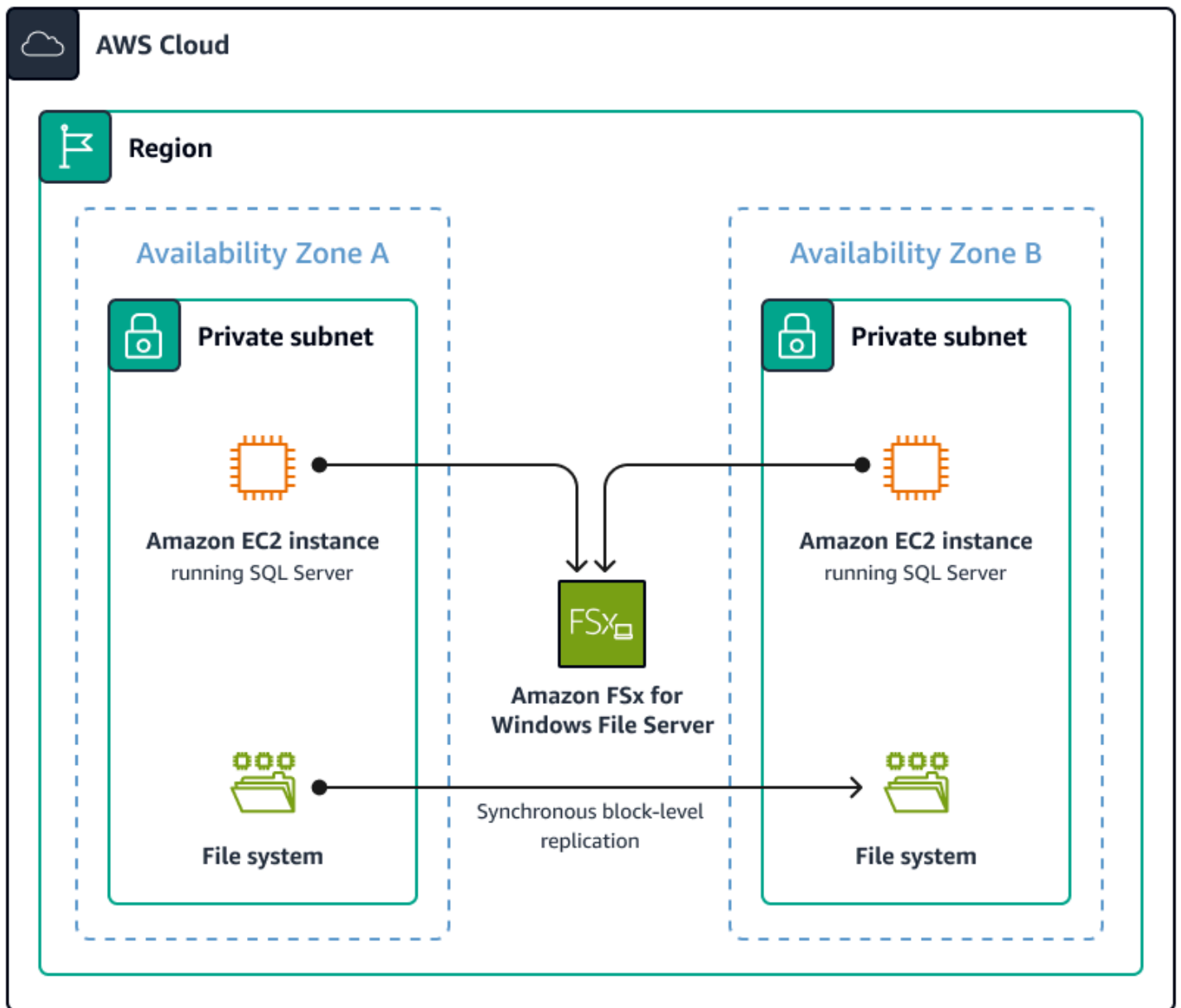
 Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Considere o seguinte:

- O Amazon FSx for Windows File Server oferece uma solução poderosa para atender aos requisitos de armazenamento compartilhado FCI do SQL Server. Você pode usar o FSx for Windows File Server para evitar a necessidade de comprar uma licença para uma solução de replicação de armazenamento e gerenciar o armazenamento compartilhado por conta própria. Isso pode resultar em economias significativas de 30 a 40 por cento. Para obter mais informações, consulte a publicação [Simplifique suas implantações de alta disponibilidade do Microsoft SQL Server usando o Amazon FSx for Windows File Server AWS](#) no blog de armazenamento.
- Com o [resumo dos benefícios do Software Assurance](#) (PDF disponível para download) e o modelo Bring Your Own License (BYOL), você pode aproveitar os benefícios do failover passivo, desde que o servidor secundário seja passivo. Isso resulta em economia de custos para o licenciamento de SQL porque você não precisa fornecer licenças para o nó passivo do cluster.

O diagrama a seguir mostra um exemplo de arquitetura para uma FCI do SQL Server usando FSx for Windows File Server.




SIOS DataKeeper

Recomendamos que você considere os requisitos de armazenamento compartilhado se estiver planejando implantar o SQL Server FCIs no AWS. As instalações locais tradicionais geralmente usam uma rede de área de armazenamento (SAN) para atender aos requisitos de armazenamento compartilhado, mas essa não é uma opção viável. AWS O Amazon FSx for Windows File Server é a solução de armazenamento recomendada para o SQL Server FCI AWS on, mas tem limitações que impedem a adição de servidores de cluster em diferentes. Regiões da AWS

Você pode usar o [SIOS DataKeeper](#) para criar uma FCI do SQL Server que cubra zonas e regiões de disponibilidade e reduza os custos em 58—71 por cento. O SIOS DataKeeper pode ajudar você a obter os benefícios de alta disponibilidade do FCI. Isso torna o SIOS DataKeeper uma solução econômica e confiável para organizações.

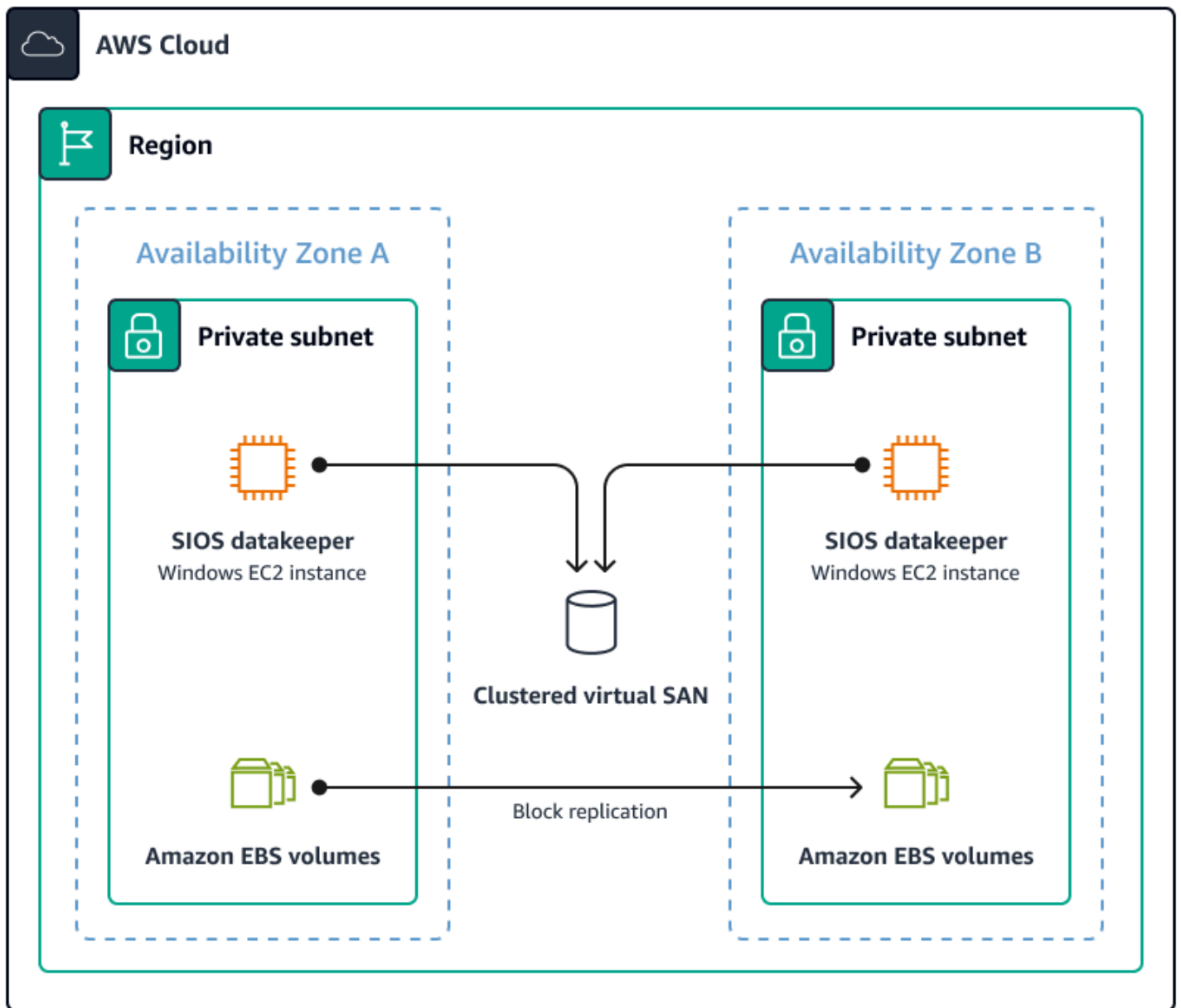
Considere os seguintes benefícios adicionais do uso do SIOS DataKeeper:

- O SIOS DataKeeper cria uma SAN virtual em cluster usando volumes locais do EBS e usa replicação síncrona entre zonas de disponibilidade para alta disponibilidade. Para recuperação de desastres, o SIOS DataKeeper usa replicação assíncrona entre regiões.
- O SIOS DataKeeper fornece recursos de clustering de classe empresarial usando a edição SQL Server Standard. Isso reduz os custos de licenciamento do SQL Server entre 65 e 75 por cento em comparação com a implementação de alta disponibilidade com grupos de disponibilidade do SQL Server Always On que usam a edição SQL Server Enterprise. Com o SIOS DataKeeper, você pode criar um ambiente SQL Server altamente disponível, flexível e econômico que atenda às necessidades da sua organização.

 Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

O diagrama a seguir mostra um exemplo de arquitetura para uma FCI do SQL Server usando uma solução SAN virtual em cluster.

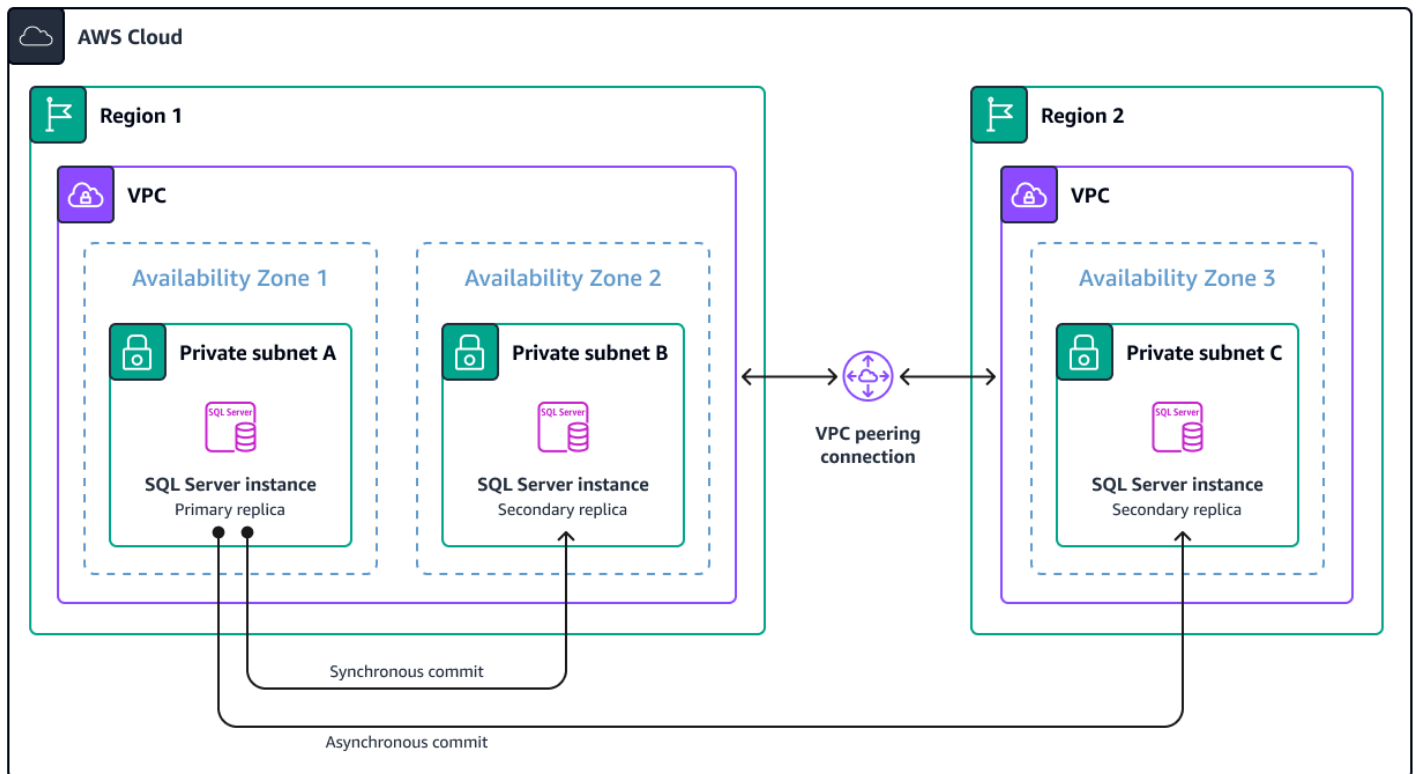


Grupos de Disponibilidade Always On

Você pode usar grupos de disponibilidade Always On para fins de alta disponibilidade e recuperação de desastres. Você pode obter alta disponibilidade implantando o SQL Server em duas zonas de disponibilidade em uma região. Você pode obter a recuperação de desastres estendendo os grupos de disponibilidade em todas as regiões.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada em grupos de disponibilidade Always On. As réplicas na Região 1 do diagrama estão usando uma confirmação

síncrona, que fornece um failover automático do grupo de disponibilidade. A réplica na Região 2 está usando uma confirmação assíncrona, que exigirá um failover manual do grupo de disponibilidade.



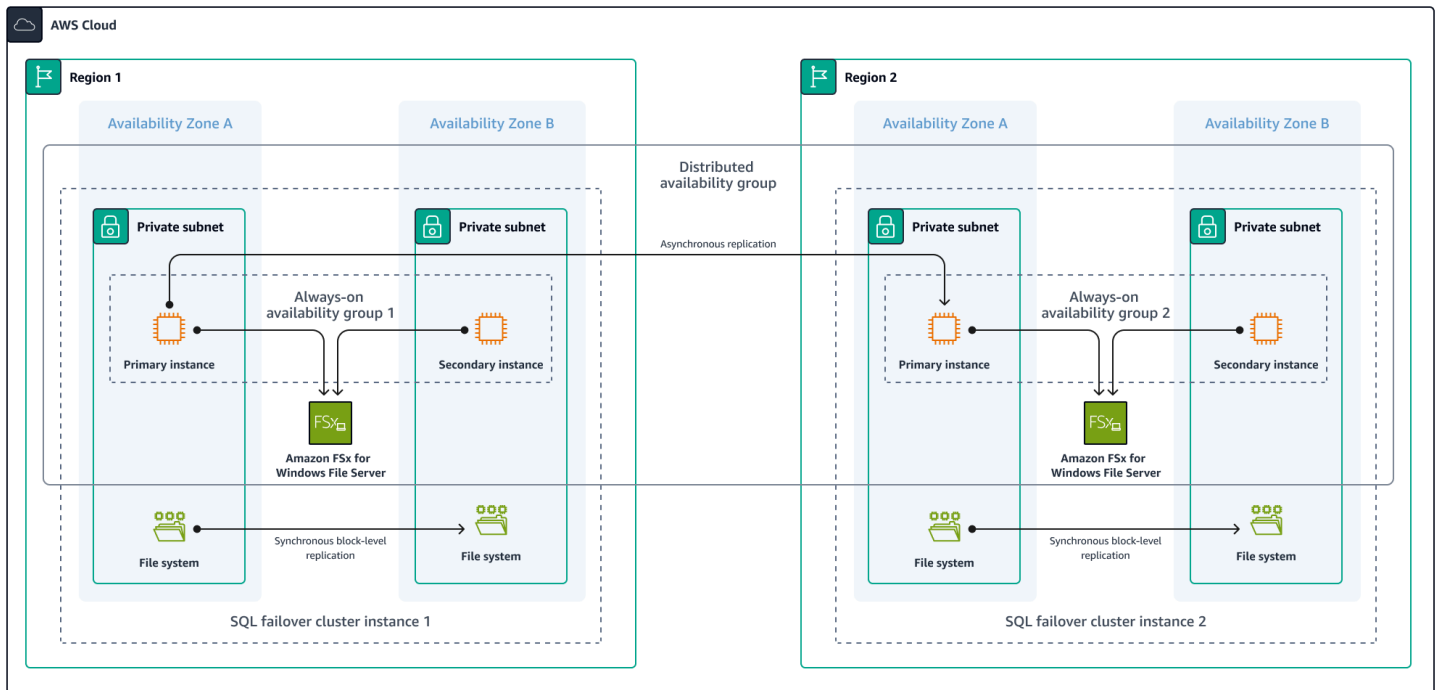
Grupos de disponibilidade distribuídos

Para implantações essenciais do SQL Server nas quais você não pode comprometer a confiabilidade ou a recuperação de desastres, recomendamos uma abordagem multirregional. Distribuir seus grupos de disponibilidade em várias regiões é a solução mais resiliente para manter a continuidade dos negócios e minimizar o tempo de inatividade.

Essa arquitetura aproveita ao máximo os recursos do Amazon FSx para Windows File Server, incluindo armazenamento compartilhado, replicação síncrona em nível de bloco e FCIs do SQL Server. Esses recursos possibilitam a criação de um ambiente SQL Server altamente disponível que abrange várias zonas de disponibilidade. Ao replicar essa configuração em outra região, você obtém um sistema totalmente redundante que pode lidar até mesmo com as interrupções mais graves. O que diferencia essa solução é o nível de flexibilidade e segurança que ela oferece. A arquitetura independente de domínio dos grupos de disponibilidade distribuídos permite que os servidores de cluster Windows subjacentes se unam a diferentes domínios do Active Directory, enquanto a autenticação baseada em certificados garante a máxima proteção para seus ambientes SQL Server e fornece altos requisitos de RTO e RPO para uma estratégia de DR multirregional. Para obter

informações sobre como criar uma arquitetura multirregional, consulte [Notas de campo: Criando uma arquitetura multirregional para o SQL Server usando FCI e grupos de disponibilidade distribuídos no AWS blog](#) de arquitetura.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução multirregional usando grupos de disponibilidade distribuídos.



Envio de logs

O envio de registros é um método comprovado, confiável e econômico para proteger seus bancos de dados em todas as regiões no caso de uma interrupção inesperada. As organizações usam o envio de registros para proteger seus dados há décadas.

Se você implementar o envio de registros em AWS, poderá obter RPO e RTO em minutos, dependendo da frequência das transações e dos trabalhos de envio de registros. No caso improvável de uma região ficar inacessível, o envio de registros mantém seus dados seguros e recuperáveis.

Considere os seguintes benefícios adicionais de usar o envio de toras:

- Reduza custos e atenda aos requisitos de seus negócios usando o envio de registros para resiliência de recuperação de desastres em todas as regiões. O envio de registros reduz seu TCO porque você só precisa de licenças do SQL Server Standard Edition ou do SQL Server Web Edition.

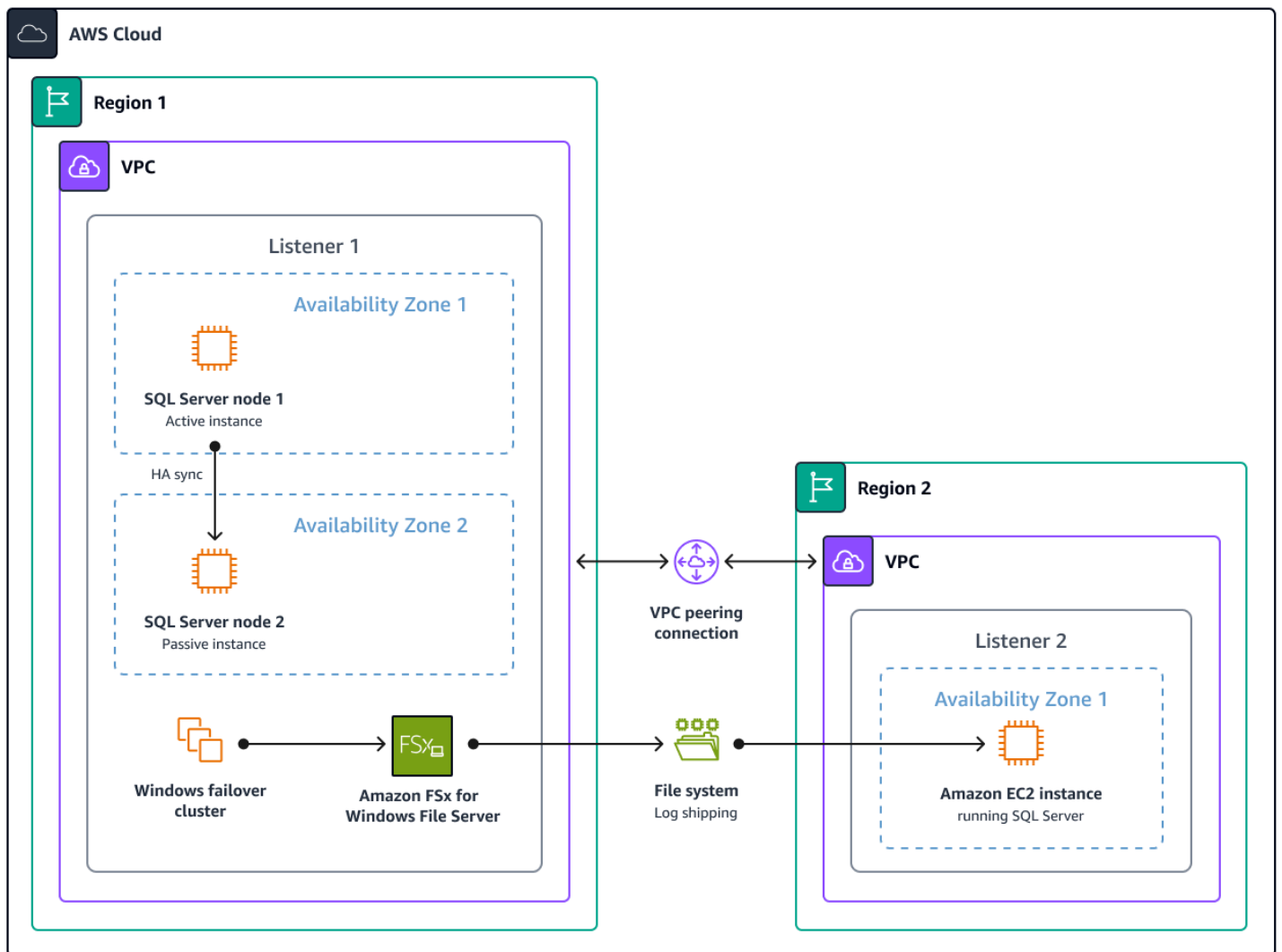
- [Remova os custos de licenciamento de um servidor passivo/de recuperação de desastres usando o envio de registros com o Software Assurance ativo.](#) Somente o SQL Server primário/ativo precisa ser licenciado quando você usa o envio de registros com o Software Assurance.
- Reduza os custos de licenciamento do SQL Server em 65 a 75 por cento eliminando a necessidade do SQL Server Enterprise Edition para configurar grupos de disponibilidade distribuídos entre as regiões. Você pode fazer isso usando o SQL Server Standard Edition e o SQL Server FCIs combinados com o envio de registros para atender aos seus requisitos de recuperação de desastres.

Note

Para obter informações adicionais sobre as diferenças de custo entre as edições do SQL Server, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Para obter mais informações, consulte [Estender o SQL Server DR usando o envio de registros para SQL Server FCI com a configuração do Amazon FSx para Windows](#) AWS no blog de arquitetura.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução de envio de toras.

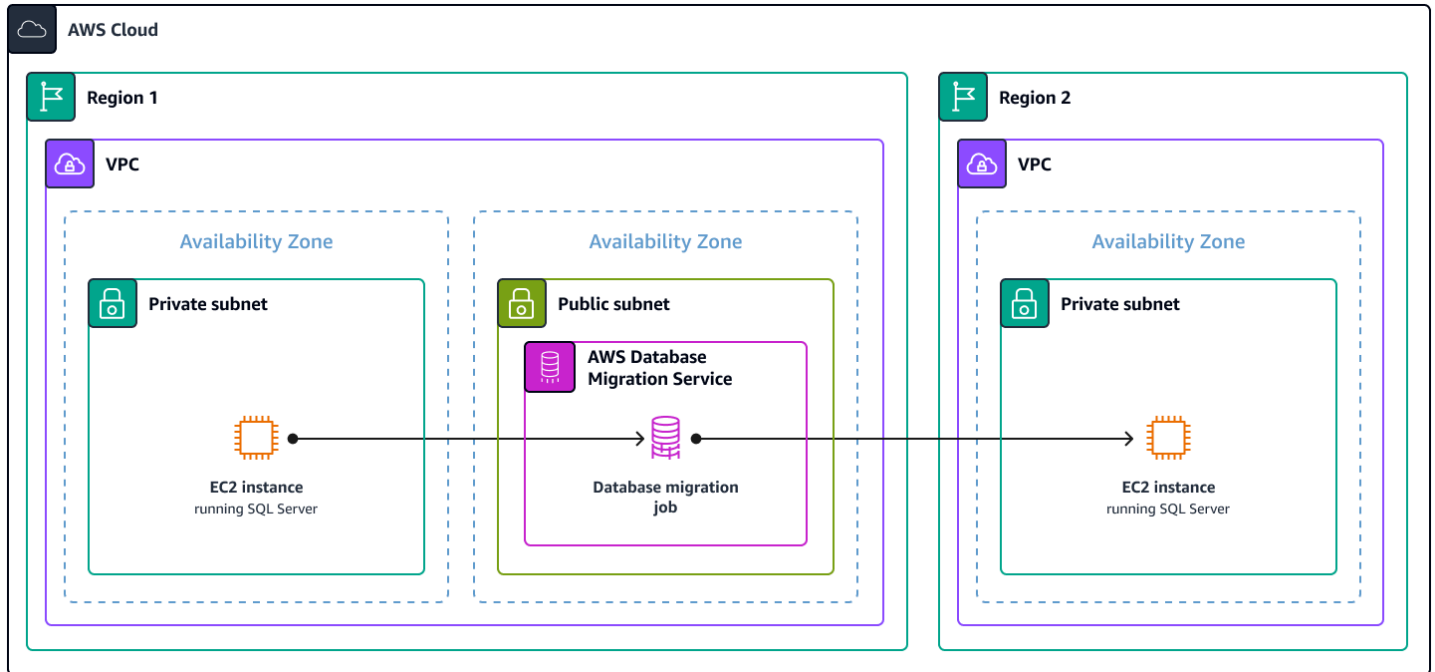


AWS Database Migration Service

Você pode usar AWS Database Migration Service (AWS DMS) para projetar uma solução de HA/DR com base nas necessidades do seu aplicativo. AWS DMS permite que você copie dados facilmente para um banco de dados secundário do SQL Server na mesma região (HA) ou entre regiões (DR). Essa abordagem é tecnicamente sólida e permite que você maximize seu investimento em AWS infraestrutura e, ao mesmo tempo, otimize o uso de recursos.

AWS DMS é um serviço econômico. Você é cobrado somente pelos recursos de CPU usados durante o processo de transferência e por qualquer armazenamento adicional de registros. Isso significa que você pode se beneficiar dessa solução sem incorrer em custos adicionais significativos. Você pode usar AWS DMS para garantir que seus dados estejam disponíveis e acessíveis, minimizando os custos associados ao licenciamento e ao uso de recursos.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada em AWS DMS.



AWS Elastic Disaster Recovery

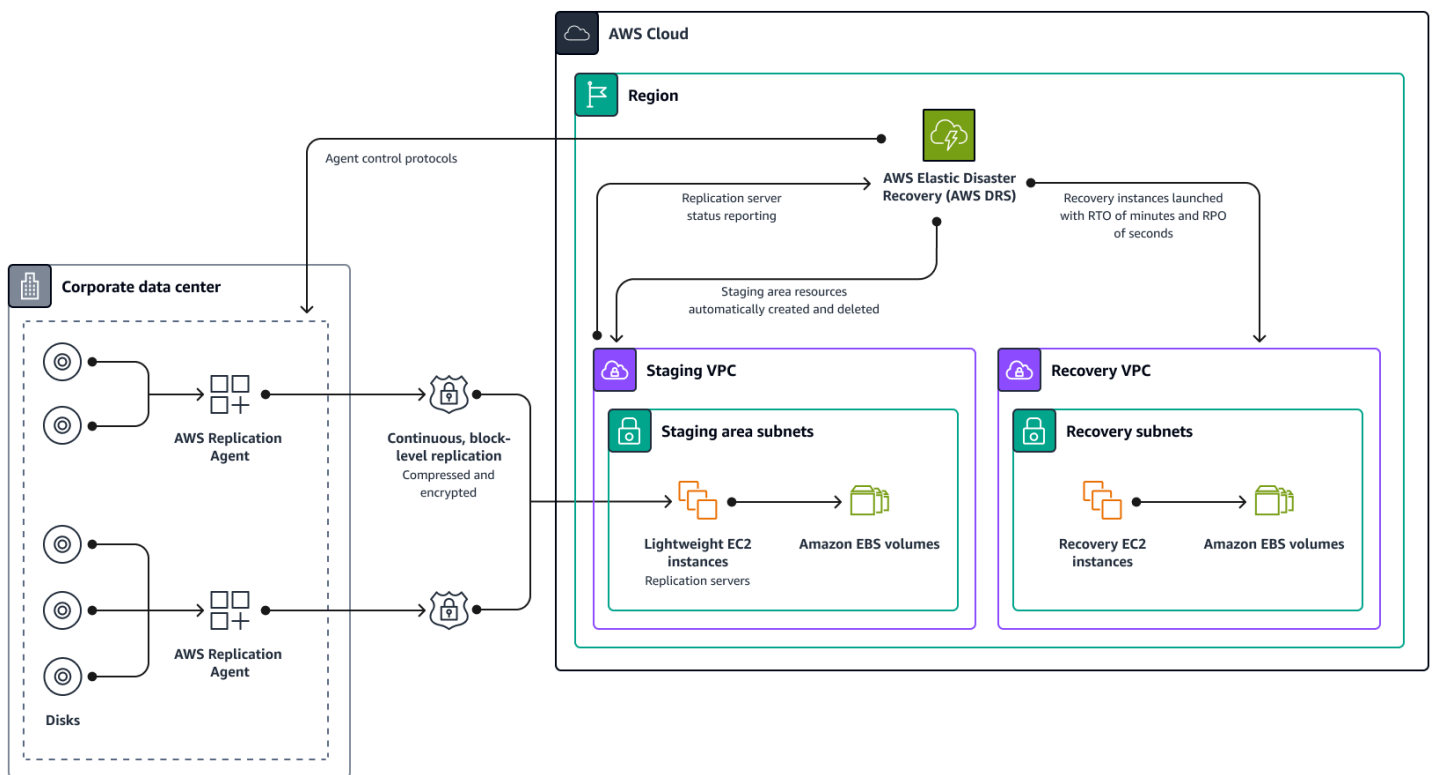
Algumas organizações devem garantir que todos os aplicativos comerciais essenciais tenham um plano de recuperação de desastres em vigor. No passado, muitas dessas organizações investiam pesadamente em soluções tradicionais de recuperação de desastres, que exigem que você pré-construa e mantenha toda uma infraestrutura duplicada. Essa abordagem é cara, demorada e difícil de escalar.

Agora, você pode usar AWS Elastic Disaster Recovery para eliminar a necessidade de pré-construir uma infraestrutura de recuperação de desastres. As máquinas de recuperação de desastres não são iniciadas no Elastic Disaster Recovery até que sejam necessárias, então você paga somente pelo que usar quando precisar. Isso significa que você pode reduzir significativamente o licenciamento de software e os custos de computação de alto desempenho.

Além disso, a área de preparação da solução de recuperação de desastres contém volumes de baixo custo do Amazon Elastic Block Store (Amazon EBS). Os volumes do EBS reduzem ainda mais o custo do provisionamento de recursos duplicados. Isso permite que você reduza seus custos gerais de recuperação de desastres e, ao mesmo tempo, mantenha uma solução de recuperação de desastres robusta e confiável que atenda aos requisitos de sua empresa. Você pode usar o Elastic Disaster Recovery para se concentrar em suas principais atividades comerciais, enquanto AWS cuida da infraestrutura subjacente da sua solução de recuperação de desastres.

Para o SQL Server, você pode usar o Elastic Disaster Recovery como uma opção econômica de recuperação de desastres. O licenciamento do nó passivo em uma arquitetura SQL Server altamente disponível e tolerante a falhas é coberto se você usar o Software Assurance ativo. No entanto, você ainda está pagando os custos de computação para que o servidor passivo esteja on-line. Com o Elastic Disaster Recovery, o servidor primário pode se replicar para o ambiente de DR sem a necessidade de manter o Software Assurance ativo e sem ter que pagar pelos custos computacionais da recuperação de desastres. Essa combinação de economias pode reduzir seus custos de recuperação de desastres do SQL Server em 50% ou mais.

O diagrama a seguir mostra um exemplo de arquitetura para uma solução baseada no Elastic Disaster Recovery.



Para obter mais informações, consulte [Como configurar a alta disponibilidade do SQL Server no site de DR que foi restaurado usando AWS Elastic Disaster Recovery](#) o Microsoft Workloads on AWS Blog.

Comparação de custos

A tabela a seguir compara os custos das soluções de HA/DR abordadas nesta seção. As seguintes suposições são feitas para fins dessa comparação:

- Tipo de instância — r5d.xlarge
- Tipo de licença — Licença incluída para Windows e SQL Server
- Região: us-east-1

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Envio de logs	Não	Sim	Sim	Sim	Edição SQL Server Enterprise: \$32.674,8 (2 nós) Edição SQL Server Standard: \$14.804,4 (2 nós)
Grupos de Disponibilidade Always On	Sim	Sim	Sim	Sim, mas grupos de disponibilidade básica (2 nós)	Edição SQL Server Enterprise: \$32.674,8 (2 nós) Edição SQL Server Standard: \$14.804,4 (2 nós)
FCIs sempre ativos	Sim	Não	Sim	Sim (2 nós)	Edição SQL Server Standard: \$14.804,4

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Grupos de disponibilidade distribuídos	Sim	Sim	Sim	Não	Edição SQL Server Enterprise: \$65.349,6 (4 nós)

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
Recuperação de desastres do Elastic	Não	Sim	Sim	Sim	<p>Aproximadamente 107,48 USD/mês para replicação de 1 instância e 1 TB de armazenamento</p> <p>Observação: o Elastic Disaster Recovery é cobrado por hora, por servidor replicante. O custo é o mesmo, independentemente do número de discos, do tamanho do armazenamento, do número de lançamentos de perfuração ou recuperação ou da região que</p>

Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
					<p>you are replicating.</p>
Protetor de dados SIOS	Sim	Sim	Sim	Sim	<p>Grupos de disponibilidade sempre ativos com o Software Assurance (2 nós, 24 núcleos): \$213.480</p> <p>Cluster SQL Server de 2 nós em execução na edição SQL Server Standard com SIOS DataKeeper e Software Assurance : \$61.530 (2 nós)</p>


Solução	Alta disponibilidade	Recuperação de desastres	Enterprise Edition	Standard Edition	Custo
AWS DMS	Não	Sim	Sim	Sim	745,38 USD/mês para instância r5.xlarge e 1 TB de armazenamento

Recomendações de otimização de custos

Recomendamos que você execute as próximas etapas a seguir para escolher uma solução de HA/DR que atenda aos requisitos da sua organização:

- Consulte a seção [Selecionar a instância EC2 certa para cargas de trabalho do SQL Server](#) deste guia.
- Determine os requisitos de IOPS e taxa de transferência de suas cargas de trabalho executando contadores de desempenho durante picos de carga de trabalho:
 - IOPS = leituras de disco/seg + gravações de disco/seg
 - Taxa de transferência = bytes de leitura de disco/seg + bytes de gravação em disco por segundo
- Use os seguintes tipos de volume de armazenamento para obter melhor desempenho e economia de custos:
 - Armazenamento de instâncias NVMe tempdb e extensão do buffer pool
 - volumes io2 para arquivos de banco de dados
- Use [AWS Trusted Advisor](#) para obter recomendações sobre otimização de custos para SQL Server no Amazon EC2. Você não precisa instalar um agente para Trusted Advisor fazer verificações de otimização do SQL Server. Trusted Advisor inspeciona suas configurações de instância incluídas na licença do Amazon EC2 SQL Server, como CPUs virtuais (vCPUs), versão e edição. Em seguida, Trusted Advisor faz recomendações com base nas melhores práticas.
- Use tanto AWS Compute Optimizer para a instância do Amazon EC2 quanto para as recomendações de dimensionamento correto do Amazon EBS.
- Use [AWS Pricing Calculator](#) para projetar sua estratégia de HA/DR para estimativas de custos.

- Para determinar se o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard é uma opção possível, use a exibição de gerenciamento dinâmico [sys dm_db_persisted_sku_features](#) para identificar recursos específicos da edição que estão ativos no banco de dados atual.

 Note

ide-by-side As migrações S são necessárias para alterações na edição do SQL Server ao usar instâncias EC2 com licença incluída.

- Realize exercícios de recuperação de desastres semestrais ou anuais para melhor arquitetar um projeto que possa recuperar o banco de dados com RTO e RPO definidos. Isso também pode ajudá-lo a identificar quaisquer pontos fracos da arquitetura.

Recursos adicionais

- [Simplifique suas implantações de alta disponibilidade do Microsoft SQL Server usando o Amazon FSx for Windows File Server AWS](#) (Storage Blog)
- [Notas de campo: Criando uma arquitetura multirregional para o SQL Server usando FCI e grupos de disponibilidade distribuídos](#) (blog de AWS arquitetura)
- [Arquitete uma recuperação de desastres para o SQL Server em AWS: Parte 1](#) (Blog do AWS banco de dados)
- [Alta disponibilidade do Microsoft SQL com o Amazon FSx para Windows](#) () YouTube
- [Maximizando o desempenho do Microsoft SQL Server com o Amazon EBS](#) (blogAWS de armazenamento)
- [Comparando seus padrões de armazenamento local com os serviços AWS de AWS armazenamento](#) (Storage Blog)
- [Planejando substituir um NAS de data center pelo Amazon FSx File Gateway](#) (AWS Storage Blog)
- [Otimizando o custo de suas implantações de alta disponibilidade do SQL Server no AWS](#) (AWS Storage Blog)
- [Como configurar a recuperação de desastres para grupos de disponibilidade do SQL Server Always On usando AWS Elastic Disaster Recovery](#) (Microsoft Workloads on AWS)
- [Como configurar a alta disponibilidade do SQL Server no local de DR que foi restaurado usando AWS Elastic Disaster Recovery](#) (Microsoft Workloads on AWS)

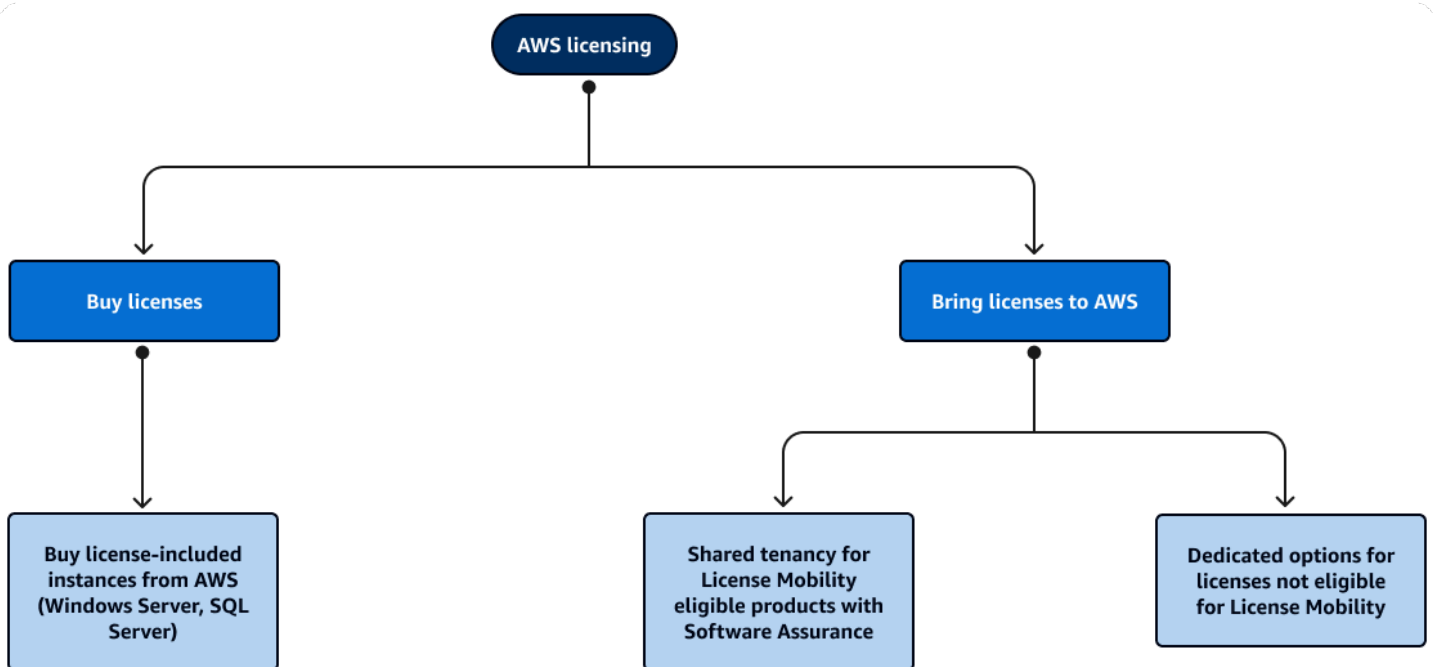
Entenda o licenciamento do SQL Server

Visão geral

À medida que mais e mais empresas transferem suas cargas de trabalho para a nuvem, otimizar os custos nas plataformas de nuvem se tornou uma prioridade máxima. O licenciamento é um dos custos mais significativos associados à execução de cargas de trabalho da Microsoft em AWS. Esta seção explica como otimizar os custos AWS otimizando o licenciamento da Microsoft para o SQL Server.

AWS opções de licenciamento

AWS oferece uma variedade de opções flexíveis de otimização de custos para licenciamento. Essas opções de licenciamento foram projetadas para ajudá-lo a reduzir custos, manter a conformidade e atender às suas necessidades comerciais.



AWS categoriza as licenças em três tipos principais:

1. Licença incluída — Essa opção de licenciamento permite que você compre e use licenças sob demanda, pagando somente pelo que você usa. A opção de licença incluída é ideal para cenários em que você precisa de flexibilidade no uso do licenciamento e deseja evitar custos iniciais. Você pode escolher entre uma variedade de produtos Windows Server, SQL Server e outros produtos da Microsoft.

2. Produtos Bring Your Own License (BYOL) com mobilidade de licenças — Essa opção de licenciamento foi projetada para cenários em que você já tem licenças existentes e deseja usá-las na nuvem. AWS permite que os clientes tragam suas próprias licenças para a nuvem por meio do programa [License Mobility](#) da Microsoft. Você pode levar produtos que tenham mobilidade de licenças, como o SQL Server com Software Assurance (SA), para locação compartilhada ou dedicada para reduzir seus custos de AWS instância.
3. Produtos BYOL sem mobilidade de licenças — Para produtos da Microsoft que não têm mobilidade de licenças, como o Windows Server, AWS oferece opções dedicadas para usar esses produtos na nuvem. Além disso, os anfitriões dedicados oferecem a oportunidade de licenciamento no nível básico físico. Isso pode economizar 50% ou mais nas licenças necessárias para executar suas cargas de trabalho. Os hosts dedicados são uma ótima opção para cargas de trabalho estáveis e previsíveis em execução na maior parte do tempo.

Impacto nos custos de trazer licenças

Trazer licenças pode ter um impacto significativo no custo de execução das cargas de trabalho da Microsoft. AWS Se você trazer suas próprias licenças, não precisará pagar custos adicionais de licenciamento pelas instâncias em execução na nuvem. Isso pode levar a uma economia significativa de custos.

A comparação a seguir mostra o custo mensal sob demanda da execução de uma única instância c5.xlarge 24 horas por dia, 7 dias por semana:

- Edição Windows Server + SQL Server Enterprise: \$1353/mês (licença incluída)
- Edição Windows Server + SQL Server Standard: \$609/mês (licença incluída)
- Somente Windows Server: \$259/mês (licença incluída)
- Somente computação (Linux): \$127/mês

Em última análise, trazer suas próprias licenças pode ter um impacto significativo no custo de execução das cargas de trabalho da Microsoft. AWS Se você usar suas licenças existentes, poderá reduzir os custos de licenciamento e economizar dinheiro em sua fatura geral AWS .

Otimização de licenças

Uma Avaliação AWS de Otimização e Licenciamento (AWS OLA) pode ajudá-lo a otimizar seu licenciamento reduzindo os custos de computação e licenciamento. AWS O OLA foi projetado para

avaliar seus requisitos de licenciamento para cargas de trabalho em execução AWS ou para cargas de trabalho planejadas para migração. AWS O LA fornece recomendações sobre como otimizar o uso da licença.

Uma das principais estratégias para otimizar o uso do licenciamento é o [dimensionamento correto](#) das instâncias. O dimensionamento correto envolve selecionar o tipo de instância certo para sua carga de trabalho com base nos requisitos de CPU, memória e armazenamento. Ao escolher o tamanho adequado da instância, você pode garantir que está usando os recursos de maneira econômica. Isso pode levar a uma economia significativa de custos.

Com o licenciamento de software da Microsoft, o número de núcleos em que o software é executado é um fator crítico na determinação dos custos de licenciamento. Por exemplo, as licenças do Windows Server e do SQL Server geralmente são licenciadas com base no número de núcleos. Ao dimensionar corretamente as instâncias, você pode reduzir o número de núcleos nos quais o software da Microsoft é executado e, por sua vez, reduzir o custo da instância e o número de licenças necessárias.

Recomendações de otimização de custos

A otimização de licenças é um componente essencial da otimização de custos em AWS. Ao implementar as estratégias certas, você pode reduzir os custos de licenciamento, manter a conformidade e obter o melhor valor possível de seu investimento em licenciamento. Esta seção descreve várias estratégias para otimização de licenças.

Traga suas licenças qualificadas do Windows Server

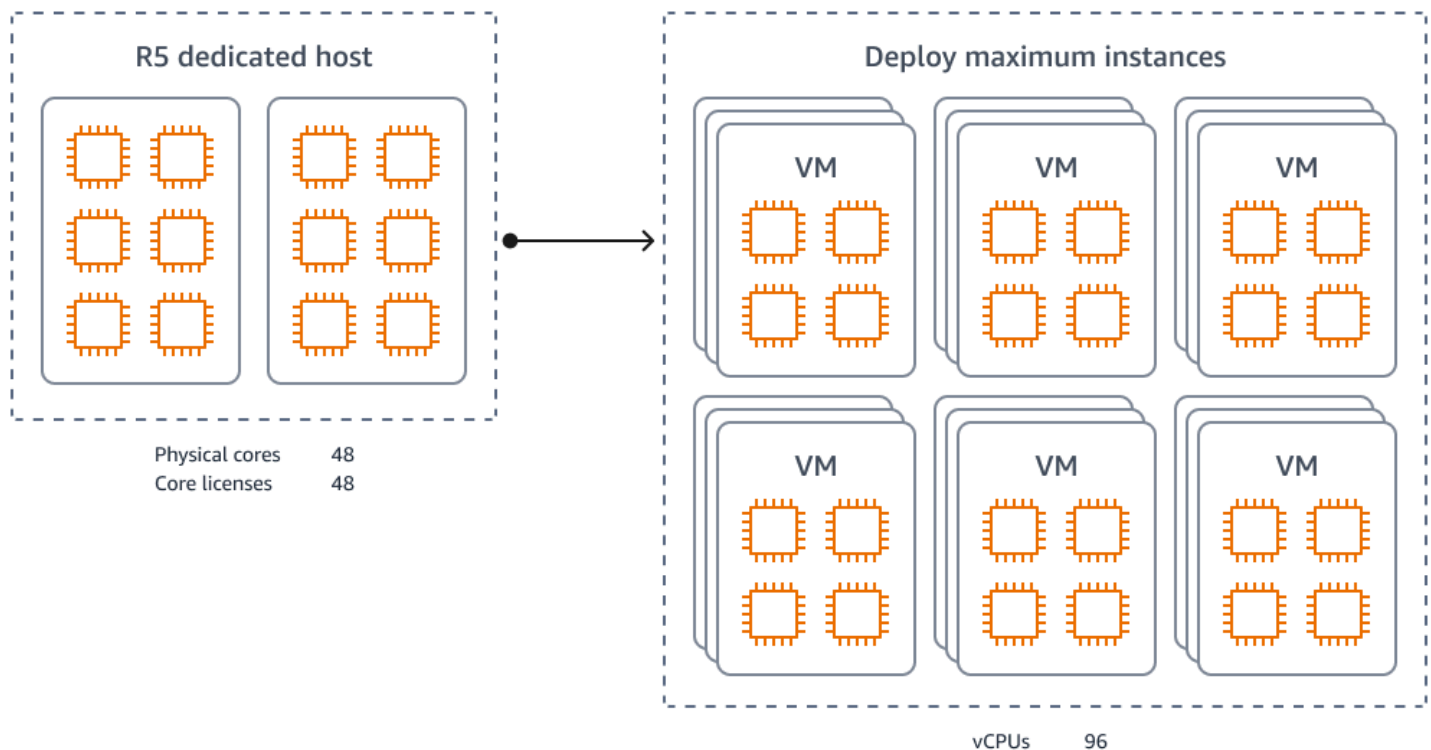
Trazer suas próprias licenças do Windows Server é uma das estratégias mais eficazes para otimização de licenças. Essa estratégia permite que você aproveite seus investimentos existentes para reduzir seus AWS gastos.

Por exemplo, você pode implantar o Windows Server 2019 e versões anteriores em [hosts dedicados do Amazon EC2](#) se tiver comprado as licenças antes de 1/10/2019 ou adquirido as licenças como ajustes sob contratos corporativos ativos assinados antes dessa data. Essa regra se baseia em uma alteração que a Microsoft fez em 2019 em seus termos e condições de licenciamento para produtos sem mobilidade de licenças, como o Windows Server, quando implantados em [fornecedores listados](#) (por exemplo, Alibaba ou Google Cloud AWS). De acordo com os novos termos, você não pode trazer suas próprias licenças do Windows Server, AWS mas deve usar instâncias com licença incluída. No entanto, se você comprou licenças perpétuas antes dessa data, ainda poderá implantar essas licenças do Windows Server nos hosts dedicados do Amazon EC2.

Licenças de nível físico

O licenciamento no nível do núcleo físico permite licenciar apenas os núcleos físicos de um host, para que você possa implantar um número máximo de instâncias sem afetar o número de licenças necessárias. Isso geralmente é feito usando o Windows Server Datacenter e a edição SQL Server Enterprise.

Como exemplo, considere um host dedicado R5 com 48 núcleos, o que se traduz em 96 vCPUs. Se você usa a edição Windows Server Datacenter, você só precisa de 48 licenças. Isso permite que você implante uma combinação de instâncias com até 96 vCPUs, conforme mostra o diagrama a seguir.



Essa abordagem pode ser especialmente econômica se você tiver cargas de trabalho suficientes para maximizar o número de instâncias que podem ser executadas em um host. Ao licenciar no nível do núcleo físico, você pode evitar custos adicionais de licenciamento para cada instância e obter o melhor valor possível para seu investimento em licenciamento.

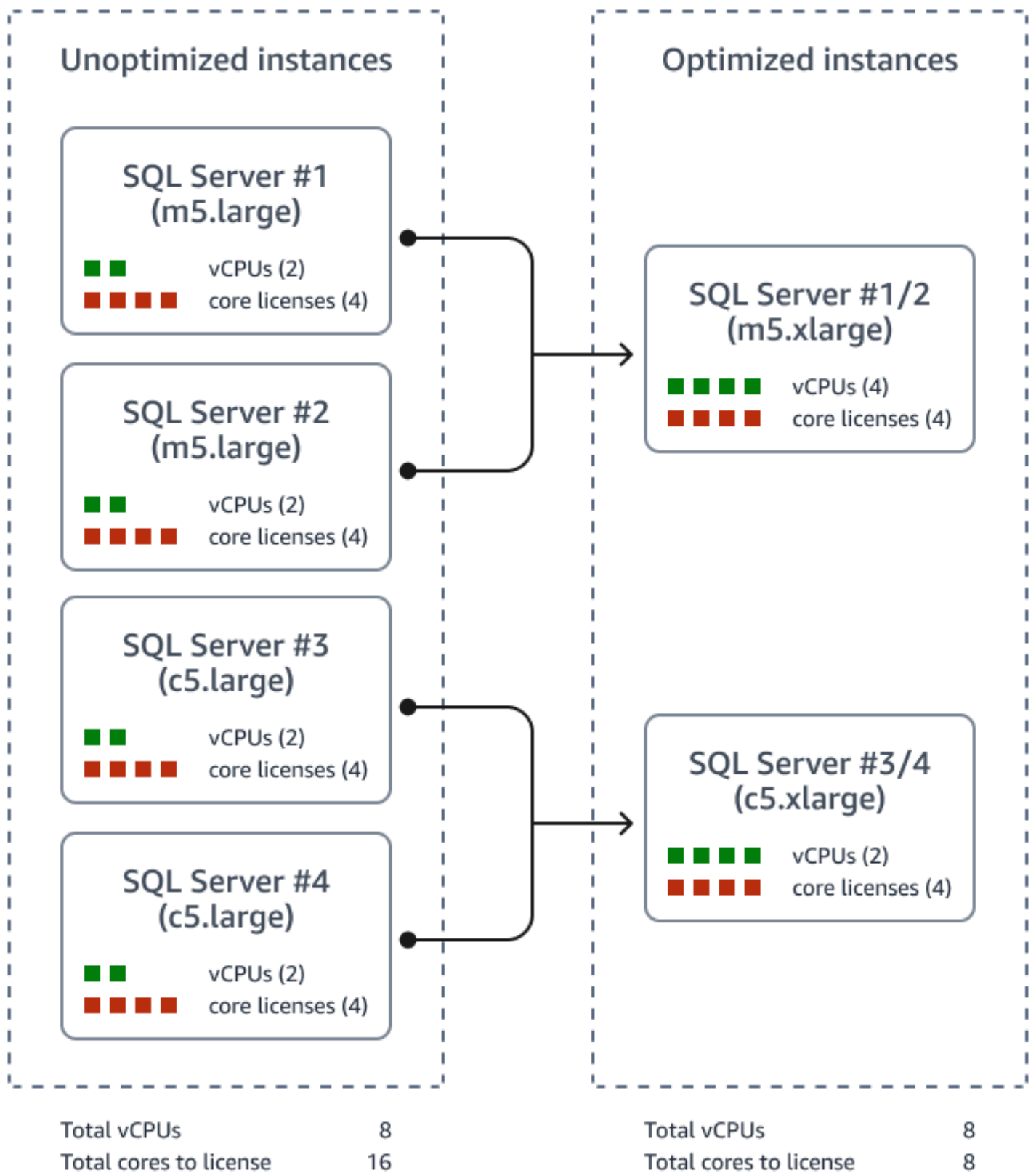
Licença no nível básico físico do SQL Server

Na locação compartilhada, o licenciamento do SQL Server é baseado no número de vCPUs alocados para a instância. Por outro lado, com hosts dedicados, você pode licenciar a edição SQL Server Enterprise no nível do núcleo físico ou no nível da vCPU.

Assim como no exemplo anterior do host dedicado R5, se você licenciar a edição SQL Server Enterprise no nível do núcleo físico, precisará apenas de 48 licenças da edição SQL Server Enterprise para licenciar os hosts. Por outro lado, na locação compartilhada, em que a única opção é licenciar por vCPU, você deve ter 96 licenças da edição SQL Server Enterprise para a mesma carga de trabalho. Portanto, os hosts dedicados podem economizar até 50% nos custos de licenciamento do SQL Server em comparação com a locação compartilhada. Isso além de economizar nos custos de instância ao trazer licenças qualificadas do Windows.

Consolide instâncias do SQL Server

A [consolidação do SQL Server](#) é o processo de combinar várias instâncias do SQL Server em um servidor. O SQL Server exige um mínimo de quatro licenças principais por instância, mesmo que a instância tenha apenas duas vCPUs. Isso significa que executar o SQL Server em servidores com menos de quatro núcleos pode fazer com que você licencie demais essas instâncias e use mais licenças do que o necessário.



Por exemplo, consolidar duas instâncias com duas vCPUs cada em uma única instância com quatro vCPUs pode reduzir a exigência de licenciamento em 50%. Isso ocorre porque são necessárias apenas quatro licenças principais, em vez de oito.

Para obter mais informações sobre consolidação, consulte a seção de [consolidação do SQL Server](#) deste guia.

Faça o downgrade das edições do SQL Server

[Alterar as edições do SQL Server](#) pode ser uma estratégia fundamental para otimizar o uso do licenciamento e reduzir custos. A edição Enterprise do SQL Server é consideravelmente mais cara do que a edição Standard, portanto, o downgrade pode resultar em economias de custo significativas.

A Criptografia de Dados Transparente (TDE) e os grupos de disponibilidade Always On são dois recursos populares na edição SQL Server Enterprise. No entanto, existem alternativas econômicas para esses recursos que você pode considerar se não precisar do conjunto completo de recursos da edição SQL Server Enterprise. Por exemplo, você pode obter o TDE na edição SQL Server Standard a partir do SQL Server 2019. No lugar dos grupos de disponibilidade Always On, você pode usar o cluster de failover com armazenamento compartilhado no FSx for Windows File Server para obter alta disponibilidade com a edição SQL Server Standard.

Ao fazer o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard, você pode reduzir significativamente seus custos de licenciamento. Para obter mais informações, consulte a postagem [Otimizando o custo de suas implantações de alta disponibilidade do SQL Server no AWS](#) blog AWS de armazenamento.

Além de reduzir os custos de licenciamento, o downgrade das edições do SQL Server pode ajudar a reduzir seus gastos com o Software Assurance e evitar futuras correções. Se você devolver licenças não utilizadas para a prateleira, poderá evitar custos adicionais de licenciamento e obter o melhor valor possível de seu investimento em licenciamento.

É importante avaliar cuidadosamente suas cargas de trabalho do SQL Server e determinar quais recursos são essenciais para suas necessidades comerciais. Para obter mais informações, consulte [Avaliando seu ambiente](#) na Orientação AWS Prescritiva e determine se seu banco de dados Microsoft SQL Server usa recursos específicos da edição SQL Server Enterprise.

Se você escolher a edição certa do SQL Server e usar alternativas aos recursos da edição SQL Server Enterprise, poderá obter economias de custo significativas, mantendo a conformidade e

atendendo às suas necessidades comerciais. Para obter mais informações sobre as opções de downgrade, consulte a seção [Comparar edições do SQL Server](#) deste guia.

Use a edição SQL Server Developer em ambientes que não sejam de produção

Em ambientes de não produção, você pode implantar edições licenciáveis do SQL Server, como a edição Enterprise ou Standard, usando assinaturas do MSDN em ambientes locais. No entanto, a assinatura do MSDN não tem mobilidade de licenças. Portanto, se você migrar para AWS, não poderá trazer essas licenças. Em vez disso, você deve usar a edição SQL Server Developer.

A edição SQL Server Developer é uma edição completa do SQL Server que está disponível gratuitamente. Esta edição está disponível para as versões 2016 e posteriores do SQL Server. Você pode baixá-lo no site da Microsoft. A edição SQL Server Developer foi projetada para ser usada em todos os ambientes que não sejam de produção, como desenvolvimento, teste e preparação, desde que não esteja conectada a dados de produção ativos.

Se você usa a edição SQL Server Developer em ambientes que não são de produção, você pode evitar custos adicionais de licenciamento. Para obter mais informações, consulte a seção [Avaliar a edição SQL Server Developer](#) deste guia.

Otimize a CPU para cargas de trabalho do SQL Server

Em alguns casos, pode ser necessário escolher um tipo de instância com mais CPUs do que o necessário para sua carga de trabalho devido a outros fatores, como RAM ou limites de rede. No entanto, AWS fornece uma solução para ajudá-lo a otimizar seus custos de licenciamento nessas situações.

Você pode, como a maioria dos clientes que trazem licenças básicas do SQL Server, desativar o hyperthreading ou desativar as CPUs na instância do EC2 para limitar o número de CPUs disponíveis para o host. Essa opção permite que você aproveite outros recursos da instância, como RAM, enquanto ainda economiza no custo da compra de licenças extras. Essa opção só está disponível ao usar BYOL e não está disponível se você estiver usando a licença incluída para SQL Server ou Windows Server.

Por exemplo, se você implantar uma instância r5.4xlarge porque sua carga de trabalho exige 128 GB de memória, mas você só precisa de oito núcleos do SQL Server, poderá desativar o hyperthreading na inicialização de uma instância com apenas oito CPUs ativas. Ao fazer isso, você pode economizar 50% nas licenças necessárias do SQL Server, pois você só precisa licenciar os oito núcleos que estão sendo usados ativamente.

Tipo de instância	Total de vCPUs	vCPU ativa com o recurso de otimização de CPUs	Economia de licenças do SQL Server
r5.4xlarge	16	8	50%
r5.12xlarge	48	8	83%

Se você usar o recurso Otimizar CPUs com imagens de máquina da Amazon (AMIs) do Windows incluídas na licença, ainda precisará pagar pelas licenças do Windows em todos os núcleos. É por isso que é importante dimensionar corretamente suas instâncias e reavaliar periodicamente seus tipos de instância.

Se você dimensionar corretamente suas instâncias, poderá garantir que está usando os tipos de instância mais econômicos para suas cargas de trabalho. À medida que a AWS introduz novos tipos de instância, é importante avaliar se essas novas instâncias podem atender aos requisitos de carga de trabalho com menos núcleos.

Recursos adicionais

- [Amazon Web Services e Microsoft: perguntas frequentes](#) (AWS documentação)

Selecione a instância EC2 certa para cargas de trabalho do SQL Server

Important

Antes de ler esta seção, recomendamos que você leia primeiro as seções [Compreender o licenciamento do SQL Server](#) e [Selecionar o tipo de instância certo para cargas de trabalho do Windows](#) deste guia.

Visão geral

O Microsoft SQL Server está sendo executado nas instâncias do Amazon Elastic Compute Cloud (Amazon EC2) há mais de 15 anos. AWS pegou essa experiência e a usou para ajudar a

desenvolver instâncias do Amazon EC2 para se adequar às cargas de trabalho do SQL Server, executadas desde especificações mínimas até clusters multirregionais de alto desempenho.

A escolha da instância EC2 correta para o SQL Server depende muito da sua carga de trabalho. Entender como o SQL Server é licenciado, como ele usa a memória e como os recursos do SQL Server se alinham às ofertas do Amazon EC2 pode ajudar a orientá-lo até a melhor instância do EC2 para seu aplicativo.

Esta seção aborda uma variedade de cargas de trabalho do SQL Server e como elas podem ser combinadas com determinadas instâncias do EC2 para reduzir ao mínimo seus custos de licenciamento e computação.


Comparação de custos

O Amazon EC2 permite que você traga sua própria licença (BYOL) ou pague conforme o uso com o licenciamento do Windows Server e do SQL Server. Para pay-as-you-go licenciamento, os custos de licenciamento das licenças do Windows Server e do SQL Server são incorporados ao custo por hora da instância EC2. Por exemplo, você pode ter AMIs diferentes com preços diferentes. O preço da AMI depende da edição do SQL Server na qual a AMI é executada.

Os preços do Windows Server e do SQL Server não estão detalhados. Você não encontrará preços detalhados em ferramentas como o [AWS Pricing Calculator](#). Se você selecionar combinações diferentes de ofertas com licença incluída, os custos de licenciamento poderão ser deduzidos, conforme mostra a tabela a seguir.

EC2 instance (Instância do EC2)	AMI	Preço de computação	Preço da licença do Windows	Preço da licença SQL	Preço total
r5.xlarge	Linux (preços de computação)	\$183,96	-	-	\$183,96
r5.xlarge	Desenvolvedor Linux + SQL	\$183,96	\$0	\$0	\$183,96
r5.xlarge	Servidor Windows (LI)	\$183,96	\$134,32	-	\$318,28

EC2 instance (Instância do EC2)	AMI	Preço de computação	Preço da licença do Windows	Preço da licença SQL	Preço total
r5.xlarge	Desenvolvedor Windows + SQL	\$183,96	\$134,32	\$0	\$318,28
r5.xlarge	Windows + SQL Web (LI)	\$183,96	\$134,32	\$49,64	\$367,92
r5.xlarge	Padrão Windows + SQL (LI)	\$183,96	\$134,32	\$350,4	\$668,68
r5.xlarge	Windows + SQL Enterprise (LI)	\$183,96	\$134,32	\$1095	\$1413,28

 Note

Os preços na tabela anterior são baseados nos preços sob demanda na us-east-1 região.

O método mais econômico para executar o SQL Server é permanecer em uma edição de nível inferior até que você precise de um recurso de uma edição de nível superior. Para obter mais informações, consulte a seção [Comparar edições do SQL Server](#) deste guia. A atualização da edição SQL Server Web para a edição SQL Server Standard custa mais de sete vezes o custo de licenciamento do SQL Server e mais de três vezes o custo da mudança da edição Standard para a edição Enterprise. A disparidade nos custos de licenciamento é um fator importante a ser considerado e é explorada no restante desta seção.

Cenário de otimização de custos

Considere um exemplo de cenário em que uma empresa de análise que rastreia veículos de entrega está buscando melhorar o desempenho do SQL Server. Depois que um especialista em MACO analisa os gargalos de desempenho da empresa, a empresa faz a transição das instâncias x1e.2xlarge para as instâncias x2iedn.xlarge. Embora o tamanho da instância seja menor, os aprimoramentos nas instâncias x2 melhoram o desempenho e a otimização do SQL Server usando extensões de buffer pool. Isso permitiu que a empresa fizesse o downgrade da edição SQL Server Enterprise para a edição SQL Server Standard e reduzisse seu licenciamento do SQL Server de 8 vCPUs para 4 vCPUs.

Antes da otimização:

Servidor	EC2 instance (Instância do EC2)	Edição do SQL Server	Custo mensal
ProdDB1	x1e.2xlarge	Enterprise	\$3.918,64
ProdDB2	x1e.2xlarge	Enterprise	\$3.918,64
Total			\$7.837,28

Após a otimização:

Servidor	EC2 instance (Instância do EC2)	Edição do SQL Server	Custo mensal
ProdDB1	x2iedn.xlarge	Padrão	\$1.215,00
ProdDB2	x2iedn.xlarge	Padrão	\$1.215,00
Total			\$2.430,00

As mudanças combinadas de instâncias x1e.2xlarge para instâncias x2iedn.xlarge permitiram que o cliente exemplo economizasse 5.407 USD por mês em seus servidores de banco de dados de produção. Isso reduziu o custo total da carga de trabalho em 69%.

Note

Os preços na tabela anterior são baseados nos preços sob demanda na us-east-1 região.

Recomendações de otimização de custos

Instâncias otimizadas para memória

Um dos aspectos mais importantes do SQL Server é entender sua dependência da memória. O SQL Server tenta usar toda a RAM disponível que não está sendo usada pelo sistema operacional (até 2 TB para uma instalação padrão). Ele faz isso por motivos de desempenho. Trabalhar com dados na memória tem muito mais desempenho do que ter que extrair dados constantemente do disco, fazer alterações e depois gravá-los de volta no disco. Em vez disso, o SQL Server tenta carregar o máximo possível de dados dos bancos de dados anexados e mantém esses dados na RAM. As alterações feitas nos dados acontecem na memória e são posteriormente armazenadas no disco.

Note

Para obter uma explicação detalhada de como o SQL Server grava alterações, consulte [Escrevendo páginas](#) na documentação da Microsoft.

Como o SQL Server funciona melhor com grandes quantidades de RAM, geralmente recomendamos começar com os tipos de instância [otimizados para memória do Amazon EC2](#). As instâncias otimizadas para memória são versáteis e oferecem uma variedade de opções diferentes. A família R tem uma proporção de 1 para 8 vCPU/RAM e tem opções para processadores Intel, processadores AMD, rede aprimorada, desempenho aprimorado do EBS, armazenamento de instâncias e velocidade aprimorada do processador. Para cargas de trabalho com muita memória, há também uma família X que combina muitas das mesmas opções e estende a proporção de VCPU/RAM para 1 a 32. Devido à versatilidade das instâncias otimizadas para memória, você pode aplicá-las a cargas de trabalho do SQL Server de todas as formas e tamanhos.

Cargas de trabalho abaixo do mínimo de recursos (menos de 4 vCPUs)

Embora alguns casos de uso funcionem bem com instâncias intermitentes (T3), recomendamos que você geralmente evite usar instâncias com capacidade de intermitência para cargas de trabalho do SQL Server. O licenciamento do SQL Server é baseado no número de vCPUs atribuídas a

uma instância. Se o SQL Server ficar inativo a maior parte do dia e estiver adquirindo créditos intermitentes, você paga pelas licenças do SQL que não está utilizando totalmente. Além disso, o SQL Server tem um requisito mínimo de licença de 4 núcleos por servidor. Isso significa que, se você tem uma carga de trabalho do SQL Server que não exige 4 vCPUs de poder computacional, você está pagando uma licença do SQL Server que não está usando. Nesses cenários, seria melhor [consolidar várias instâncias do SQL Server](#) em um servidor maior.

Cargas de trabalho usando recursos mínimos (menos de 64 GB de RAM)

Muitas cargas de trabalho do SQL Server com menos de 64 GB de RAM não priorizam o alto desempenho ou a alta disponibilidade. Para esses tipos de cargas de trabalho, o SQL Server Web Edition pode ser uma boa opção se o aplicativo estiver coberto pelas restrições de licenciamento da Microsoft.

Important

O SQL Server Web Edition tem um caso de uso restrito com base nos termos de licenciamento da Microsoft. A edição SQL Server Web pode ser usada apenas para fornecer suporte a páginas da Web, sites, aplicativos da Web e serviços da Web públicos e acessíveis pela Internet. Ele não pode ser usado para oferecer suporte a line-of-business aplicativos (por exemplo, gerenciamento de relacionamento com clientes, gerenciamento de recursos corporativos e outros aplicativos similares).

O SQL Server Web Edition pode ser expandido para até 32 vCPUs e 64 GB de RAM e é 86% mais barato do que o SQL Server Standard Edition. Para cargas de trabalho com poucos recursos, usar uma instância otimizada para memória AMD, como a r6a, que tem um preço de computação 10% mais barato do que sua equivalente da Intel, também é uma boa maneira de reduzir ao mínimo os custos de computação e licenciamento de SQL.

Cargas de trabalho com recursos médios (menos de 128 GB de RAM)

A edição SQL Server Standard é usada na maioria das cargas de trabalho do SQL Server com até 128 GB de RAM. A edição SQL Server Standard é 65 a 75% mais barata do que a edição SQL Server Enterprise e pode ser expandida para até 48 vCPUs e 128 GB de RAM. Como a limitação de 128 GB de RAM geralmente é atingida antes da limitação de 48 vCPUs, esse é o foco da maioria dos clientes que desejam evitar a atualização para a edição SQL Server Enterprise.

O SQL Server tem um recurso chamado [extensão do buffer pool](#). Esse recurso permite que o SQL Server use uma parte de um disco para atuar como uma extensão da RAM. [A extensão do buffer pool funciona bem quando combinada com armazenamento ultrarrápido, como os SSDs NVMe usados no armazenamento de instâncias do Amazon EC2](#). As instâncias do Amazon EC2 que contêm armazenamento de instâncias são indicadas com um “d” no nome da instância (por exemplo, r5d, r6id e x2iedn).

As extensões do buffer pool não substituem a RAM normal. No entanto, se você precisar de mais de 128 GB de RAM, poderá usar extensões de buffer pool com instâncias do EC2, como r6id.4xlarge e x2iedn.xlarge, para atrasar a atualização para o licenciamento da edição Enterprise.

Cargas de trabalho de alto desempenho (menos 128 GB de RAM)

As cargas de trabalho do SQL Server que exigem alto desempenho são um desafio para a otimização de custos devido à sua dependência de muitos recursos. No entanto, entender as diferenças nas instâncias do EC2 pode impedir que você faça a escolha errada.

A tabela a seguir mostra uma variedade de instâncias do EC2 otimizadas para memória e seus limites de desempenho.

	r5b	r6idn	7riz	x2iedn	x2iezn
Processador	3,1 GHz	3,5 GHz	3,9 GHz	3,5 GHz	4,5 GHz
	Processador Intel Xeon de 2ª geração	Processador Intel Xeon de 3ª geração	Processador escalável Intel Xeon de 4ª geração	Processador Intel Xeon de 3ª geração	Processador Intel Xeon de 2ª geração
Proporção de CPU: RAM	1:8	1:8	1:8	1:32	1:32
Máximo de vCPU	96	128	128	128	48
RAM máxima	768 GB	1.024 GB	1.024 GB	4.096 GB	1.536 GB
Armazenamento de instância	–	SSD de NVMe	–	SSD de NVMe	–

	r5b	r6idn	7riz	x2iedn	x2iezn
		(4 x 1900 GB)		(2 x 1900 GB)	
io2 Block Express	Compatível	Compatível	Compatível	Compatível	–
IOPS máximo do EBS	260.000	350.000	160.000	260.000	80.000
Taxa de transferência máxima do EBS	60 Gbps	80 Gbps	40 Gbps	80 Gbps	19 Gbps
Largura de banda máxima da rede	25 Gbps	200 Gbps	50 Gbps	100 Gbps	100 Gbps

Cada instância é usada para uma finalidade diferente. Compreender sua carga de trabalho do SQL Server pode ajudá-lo a escolher o tipo de instância mais adequado para você.

Detalhes sobre atributos:

- r5b — O atributo “b” em r5b significa que esse tipo de instância está focado no alto desempenho do EBS. Na quinta geração de instâncias otimizadas para memória, o r5b foi a escolha preferida. Foi o primeiro tipo de instância a utilizar volumes io2 Block Express e atingir o máximo de IOPS de armazenamento de 260.000. O tipo de instância r5b ainda é uma alternativa econômica para as necessidades de alto desempenho do EBS.
- r6idn — A sexta geração de instâncias otimizadas para memória ofereceu melhorias consideráveis em relação à geração anterior. Os aprimoramentos de desempenho do EBS do r5b são levados um passo adiante com o r6idn, aumentando o IOPS máximo para 350.000. O r6idn também tem um volume de armazenamento de instâncias para extensões tempdb e buffer pool para aumentar ainda mais o desempenho do SQL Server.
- x2iedn — O x2iedn é semelhante ao r6idn. Ele oferece níveis semelhantes de EBS aprimorado, rede aprimorada e armazenamento de instância SSD NVMe, mas com uma proporção de 1:32

de vCPU/RAM para altas cargas de trabalho de memória e baixa quantidade de CPU (menores custos de licenciamento do SQL Server).

- **x2iezn** — O atributo “z” em x2iezn indica que esse tipo de instância está focado no alto desempenho do processador. O processador Cascade Lake tem uma frequência turbo de todos os núcleos de até 4,5 GHz. Recomendamos que você use essa instância do EC2, juntamente com uma proporção de 1:32 de vCPU/RAM, em um cenário em que você deseja manter a quantidade de vCPU baixa. Isso, por sua vez, pode manter baixos os custos de licenciamento do SQL Server.
- **r7iz** — O atributo “z” em r7iz indica que esse tipo de instância está focado no alto desempenho do processador. O processador Sapphire Rapids tem uma frequência turbo de todos os núcleos de até 3,9 GHz. Assim como as instâncias x2iezn, o r7iz prioriza o desempenho do processador de alta frequência, mas com uma proporção de 1:8, vCPU/RAM.

Recursos adicionais

- [Instâncias do Amazon EC2 de uso geral \(documentação\)](#) AWS
- [Ferramenta de comparação](#) (Vantage)
- [Licenciamento — SQL Server](#) (AWS documentação)

Consolide instâncias

Esta seção se concentra na técnica de otimização de custos de combinar várias instâncias do SQL Server no mesmo servidor para minimizar os custos de licenciamento e maximizar a utilização dos recursos.

Visão geral

A criação de uma instância faz parte do processo de instalação do Mecanismo de Banco de Dados do SQL Server. A instância do SQL Server é uma instalação completa, contendo seus próprios arquivos de servidor, logons de segurança e bancos de dados do sistema (master, model, msdb e tempdb). Como uma instância tem todos os seus próprios arquivos e serviços, você pode instalar várias instâncias do SQL Server no mesmo sistema operacional sem que as instâncias interfiram umas nas outras. No entanto, como as instâncias estão todas instaladas no mesmo servidor, todas compartilham os mesmos recursos de hardware, como computação, memória e rede.

É comum usar apenas uma única instância do SQL Server por servidor em ambientes de produção para que uma instância “ocupada” não use demais os recursos de hardware compartilhados. Dar

a cada instância do SQL Server seu próprio sistema operacional, com seus próprios recursos, é um limite melhor do que confiar na governança de recursos. Isso é especialmente verdadeiro para cargas de trabalho de alto desempenho do SQL Server que exigem grandes quantidades de recursos de RAM e CPU.

No entanto, nem todas as cargas de trabalho do SQL Server usam uma grande quantidade de recursos. Por exemplo, algumas organizações atribuem a cada um de seus clientes sua própria instância dedicada do SQL Server para fins de conformidade ou segurança. Para clientes menores ou clientes que normalmente não estão ativos, isso significa executar as instâncias do SQL Server com recursos mínimos.

Conforme observado no [guia Microsoft SQL Server 2019: Licenciamento](#), cada servidor que executa o SQL Server deve contabilizar um mínimo de quatro licenças de CPU. Isso significa que, mesmo que você execute um servidor com apenas duas vCPUs, você ainda deverá licenciar o SQL Server para quatro vCPUs. Com base nos [preços públicos do SQL Server da Microsoft](#), essa é uma diferença de \$3.945 se você usar a edição SQL Server Standard. Para organizações que executam vários servidores com instâncias únicas do SQL Server usando recursos mínimos, o custo combinado de ter que licenciar recursos não utilizados pode ser substancial.

Cenário de otimização de custos

Esta seção explora um exemplo de cenário que compara a diferença entre executar quatro servidores Windows Server, cada um com uma única instância do SQL Server, com um único servidor Windows Server maior executando várias instâncias do SQL Server simultaneamente.

Se cada instância do SQL Server precisar apenas de duas vCPUs e 8 GB de RAM, o custo total por servidor será de 7.890 USD para a licença do SQL Server, além de um custo de computação por hora de 0,096 USD.

EC2 instance (Instância do EC2)	vCPUs	RAM	Preço	vCPUs para licenciar	Custo total de licenciamento do SQL Server
m6i.large	2	8	0,096	4	\$7.890

Expandindo isso para quatro servidores, o custo total é de \$31.560 para a licença do SQL Server com um custo de computação por hora de \$0,384.

EC2 instance (Instância do EC2)	vCPUs	RAM	Preço	vCPUs para licenciar	Custo total de licenciamento do SQL Server
4x 6i.large	2	32	0,384	16	\$31.560

Se você combinar todas as quatro instâncias do SQL Server em uma única instância do EC2, a quantidade total de recursos computacionais e computação permanecerá a mesma. No entanto, ao remover os custos desnecessários de licenciamento do SQL Server, você pode reduzir o custo total de execução da carga de trabalho em \$15.780.

EC2 instance (Instância do EC2)	vCPUs	RAM	Preço	vCPUs para licenciar	Custo total de licenciamento do SQL Server
m6i.2xlarge	8	32	0,384	8	\$15.780

Note

Nas tabelas anteriores, os custos computacionais mostram preços por hora sob demanda para servidores Amazon EC2 que executam o Windows Server na região. us-east-1 Os custos de licenciamento do SQL Server Standard Edition se referem aos [preços públicos do SQL Server da Microsoft](#).

Recomendações de otimização de custos

Se você está pensando em consolidar instâncias do SQL Server, a maior preocupação é o consumo de recursos para cada uma das instâncias que você deseja consolidar. É importante obter métricas de desempenho por longos períodos para entender melhor os padrões de carga de trabalho em cada servidor. Algumas ferramentas comuns para monitoramento do consumo de recursos são [Amazon CloudWatch](#), [Windows Performance Monitor](#) (perfmon) e as [ferramentas de monitoramento nativas](#) do SQL Server.

Recomendamos que você considere as seguintes questões ao analisar se suas cargas de trabalho do SQL Server poderiam ser combinadas para usar os mesmos recursos do servidor sem que elas interfiram umas nas outras:

- Quais recursos (CPU, memória e largura de banda de rede) são consumidos durante seu estado estável?
- Quais recursos (CPU, memória e largura de banda de rede) são consumidos durante os picos?
- Com que frequência ocorrem picos? Os picos são consistentes?
- Os picos de recursos de um servidor coincidem com os picos de recursos de outro servidor?
- Quais são as [IOPS e a taxa de transferência de](#) armazenamento usadas pelo SQL Server?

Se você quiser seguir em frente com um plano para combinar instâncias do SQL Server, consulte a postagem [Executar várias instâncias do SQL Server em uma instância do Amazon EC2](#) no blog AWS Cloud Operations & Migrations. Esta postagem fornece instruções sobre como fazer as alterações de configuração no SQL Server para adicionar instâncias adicionais. Antes de começar, considere as pequenas diferenças quando várias instâncias são instaladas no mesmo servidor:

- A instância padrão do banco de dados SQL Server é nomeada MSSQLSERVER e usa a porta 1433.
- Cada instância adicional instalada no mesmo servidor é uma instância de banco de dados “nomeada”.
- Cada instância nomeada tem um nome de instância e uma porta exclusivos.
- O [Navegador do SQL Server](#) deve ser executado para coordenar o tráfego para as instâncias nomeadas.
- Cada instância pode usar locais separados para arquivos de dados do banco de dados e logins separados.
- As [configurações de memória máxima do servidor](#) do SQL Server devem ser definidas de acordo com as necessidades de desempenho de cada instância, com seu total combinado também deixando memória suficiente para o sistema operacional subjacente.
- Você pode usar os recursos [nativos de backup e restauração do](#) SQL Server ou [AWS DMS](#) para migração ou consolidação.

Recursos adicionais

- [Ficha técnica de licenciamento do SQL Server](#) (blog de operações e migrações AWS na nuvem)

- Publicação no blog sobre [configuração de várias instâncias do SQL Server \(blog de operações e migrações AWS na nuvem\)](#)
- [Guia de melhores práticas do SQL Server](#) (documentação de orientação AWS prescritiva)

Compare as edições do SQL Server

Visão geral

O licenciamento do Microsoft SQL Server é uma das maiores despesas de um ambiente de carga de trabalho do Windows. Os custos de licenciamento do SQL Server podem facilmente ir além dos custos de computação para executar a carga de trabalho. Se você escolher a edição errada, poderá pagar por recursos que não está usando ou nem precisa. Esta seção compara as seguintes edições do SQL Server, incluindo seus recursos e custos relativos:

- Enterprise — A edição SQL Server Enterprise fornece recursos de data center com alto desempenho, virtualização ilimitada e várias ferramentas de business intelligence (BI).
- Standard — A edição SQL Server Standard fornece gerenciamento básico de dados e inteligência comercial para organizações e departamentos menores.
- Web — A edição Web do SQL Server é adequada para empresas que são hospedeiras na web ou provedores de valor agregado da web (VAPs). Esta edição oferece um baixo custo total de propriedade e fornece recursos de escalabilidade e capacidade de gerenciamento para propriedades da web de pequena a grande escala.

Important

Você pode usar o SQL Server Web Edition para oferecer suporte somente a páginas da Web, sites, aplicativos Web e serviços Web públicos e acessíveis pela Internet. Você não pode usar o SQL Server Web Edition para oferecer suporte a line-of-business aplicativos (como aplicativos de gerenciamento de relacionamento com clientes ou aplicativos de gerenciamento de recursos corporativos).

- Desenvolvedor — A edição SQL Server Developer inclui todas as funcionalidades da edição Enterprise, mas é destinada apenas para fins de desenvolvimento.
- Express — A edição SQL Server Express é um banco de dados gratuito e pode ser usada para aprender ou criar aplicativos de desktop. Você pode atualizar a edição Express para outras edições.

Note

A edição SQL Server Evaluation está disponível por um período de teste de 180 dias.

Impacto nos custos

Você pode comprar licenças do SQL Server de um revendedor da Microsoft e trazê-las AWS com o Software Assurance. Como alternativa, você pode usar licenças do SQL Server com um pay-as-you-go modelo que tenha AMIs do Amazon EC2 incluídas na licença.

Se você comprar licenças do SQL Server de revendedores da Microsoft, as licenças principais serão vendidas em pacotes de duas e você deverá licenciar no mínimo quatro núcleos por servidor. A tabela a seguir mostra uma comparação de custos entre as edições Enterprise e Standard.

Version (Versão)	Edição SQL Server Enterprise (pacote de 2 núcleos)	Edição SQL Server Standard (pacote de 2 núcleos)	Economia
2022	\$15.123	\$3.945	74%
2019	\$13.748	\$3.586	74%

Note

Os preços na tabela anterior são baseados nos preços públicos da Microsoft para o [SQL Server 2022](#) e o [SQL Server 2019](#).

A comparação de custos a seguir mostra a hospedagem de diferentes edições do SQL Server com AMIs do Amazon EC2 incluídas na licença. Nessa comparação, o SQL Server está hospedado em r6i.xlarge (4 vCPUs) na região. us-east-1

Instância	Custo de computação	Custo da licença do Windows	Custo da licença do SQL Server	Total
R6l.xlarge (Linux)	\$183,96	–	–	\$183,96
R6l.xlarge + Windows	\$183,96	\$134,32	–	\$318,28
R6l.xlarge + Edição Web do SQL Server	\$183,96	\$134,32	\$49,35	\$367,63
R6l.xlarge + Edição padrão do SQL Server	\$183,96	\$134,32	\$350,4	\$668,68
Edição R6l.xlarge + SQL Enterprise	\$183,96	\$134,32	\$1.095	\$1.413,28

Você pode economizar até 95% nos custos de licenciamento do SQL Server selecionando a edição certa do SQL Server para sua carga de trabalho. A tabela a seguir compara o custo das licenças do SQL Server em instâncias r6i.xlarge.

Edição	% de economia
Padrão comparado ao Enterprise	68%
Web em comparação com o Standard	86%
Web em comparação com Enterprise	95%

Na maioria dos cenários, as organizações mudam da edição Enterprise para a edição Standard, mas há alguns casos em que é possível mudar da edição Standard ou Enterprise para a edição Web.

Recomendações de otimização de custos

Você pode escolher a melhor edição para sua carga de trabalho com base em limites de escalabilidade, alta disponibilidade, desempenho e segurança. A tabela a seguir mostra os recursos que são compatíveis com as edições do SQL Server. Isso pode ajudar você a decidir qual edição usar. Essa comparação se aplica ao [SQL Server 2016 SP1 e versões posteriores](#).

Limites de escala

A tabela a seguir compara os limites de escala das diferentes edições do SQL Server.

Atributo	Enterprise Edition	Standard Edition	Edição web	Edição Express
Capacidade máxima de computação usada por uma única instância do SQL Server Database Engine, SQL Server Analysis Services (SSAS) ou SQL Server Reporting Services (SSRS)	Sistema operacional máximo	Limitado a menos de 4 soquetes ou 24 núcleos	Limitado a menos de 4 soquetes ou 16 núcleos	Limitado a menos de 4 soquetes ou 4 núcleos
Memória máxima para buffer pool por instância do SQL Server Database Engine	Sistema operacional máximo	128 GB	64 GB	1410 MB
Capacidade máxima para extensão do buffer pool por	32 vezes a memória máxima configurada	4 vezes a memória máxima configurada	N/D	N/D

Atributo	Enterprise Edition	Standard Edition	Edição web	Edição Express
instância do SQL Server Database Engine				
Tamanho máximo do banco de dados relacional	524 LIBRAS	524 LIBRAS	524 LIBRAS	10 GB
Memória máxima para caches do Columnstore ou dados otimizados para memória	Sistema operacional máximo	32 GB	16 GB	352 MB

Se seu aplicativo exigir menos de 16 núcleos (32 vCPUs) e 64 GB de RAM, você poderá começar a avaliar a partir da edição Web do SQL Server. Se sua carga de trabalho exigir mais de 64 GB de memória ou outras opções de alta disponibilidade, você deverá atualizar para a edição SQL Server Standard.

Você pode usar o SQL Server Web Edition para oferecer suporte a páginas da Web, sites, aplicativos Web e serviços Web públicos e acessíveis pela Internet, mas não pode usar o SQL Server Web Edition para oferecer suporte a aplicativos de linha de negócios. Para obter mais informações sobre casos de uso do SQL Server Web Edition, entre em contato com o [Microsoft Licensing Support](#) ou com seu revendedor da Microsoft.

Você pode usar a edição SQL Server Standard para cargas de trabalho de até 24 núcleos (48 vCPUs) e 128 GB de memória. No entanto, você pode usar [extensões de buffer pool](#) para permitir que o SQL Server Standard Edition utilize o [armazenamento de instâncias locais](#), como as presentes nas instâncias r6id EC2. Isso estende a memória até o tamanho de quatro vezes a configuração máxima de memória. Essa combinação de recursos pode atrasar a atualização do servidor para a edição Enterprise quando os requisitos de memória começarem a aumentar.

Você pode identificar a utilização da memória localizando as páginas dos bancos de dados no pool de buffers e nos contadores de [expectativa de vida útil da página](#). A expectativa de vida da página

indica quanto tempo a página fica na memória antes de ser transferida de volta para o disco. Esse valor padrão do contador é 300. Se as páginas permanecerem na memória por horas ou dias, há uma chance de reduzir a memória alocada.

Alta disponibilidade

A tabela a seguir compara os recursos de alta disponibilidade das diferentes edições do SQL Server.

Atributo	Enterprise Edition	Standard Edition	Edição web	Edição Express
Suporte básico do servidor 1	Sim	Sim	Sim	Sim
Envio de logs	Sim	Sim	Sim	Não
Espelhamento de banco de dados	Sim	Modo de segurança COMPLETO	Somente como testemunha	Somente como testemunha
Compressão de backup	Sim	Sim	Não	Não
Instâncias de cluster de failover sempre ativas	16 nós	2 nós	Não	Não
Grupos de Disponibilidade Always On	Até 8 réplicas secundárias, incluindo 2 réplicas secundárias síncronas	Não	Não	Não
Grupos de disponibilidade básica	Não	2 nós	Não	Não

Atributo	Enterprise Edition	Standard Edition	Edição web	Edição Express
Restauração online de páginas e arquivos	Sim	Não	Não	Não
Indexação online	Sim	Não	Não	Não
Alteração do esquema online	Sim	Não	Não	Não
Recuperação rápida	Sim	Não	Não	Não
Backups espelhados	Sim	Não	Não	Não
Adicione memória e CPU a quente	Sim	Não	Não	Não
Backup criptografado	Sim	Sim	Não	Não
Backup híbrido para o Microsoft Azure (backup para URL)	Sim	Sim	Não	Não
Servidor de failover para recuperação de desastres	Sim	Sim	Não	Não
Servidores de failover para alta disponibilidade	Sim	Sim	Não	Não

Outras características comuns

A tabela a seguir compara os recursos mais comuns das diferentes edições do SQL Server. Para obter uma lista extensa de recursos, consulte [Edições e recursos compatíveis do SQL Server 2019](#) na documentação da Microsoft.

Atributo	Enterprise Edition	Standard Edition	Edição web	Edição Express
(Desempenho) Administrador de recursos	Sim	Não	Não	Não
(Segurança) Criptografia de banco de dados transparente (TDE)	Sim	Sim	Sim	Não
(Segurança) Gerenciamento extensível de chaves (EKM)	Sim	Não	Não	Não
(Replicação) Publicação da Oracle	Sim	Não	Não	Não
(Replicação) Replicação transacional ponto a ponto	Sim	Não	Não	Não
Captura de dados de alteração	Sim	Sim	Não	Não

Edição SQL Server Developer

Todas as cargas de trabalho que não sejam de produção, como ambientes de desenvolvimento, controle de qualidade, testes, preparação e UAT, podem usar a edição SQL Server Developer para economizar 100% nos custos de licenciamento do SQL Server. Depois de [baixar o SQL Server](#), você pode instalar o SQL Server Developer Edition em instâncias do EC2 usando locação compartilhada. A infraestrutura dedicada não é necessária para a edição SQL Server Developer. Para obter mais informações, consulte a recomendação deste guia para a [edição SQL Server Developer](#).

Trocando edições

Para cargas de trabalho existentes, a mudança de uma edição para outra requer testes extensivos. É uma prática recomendada verificar as cargas de trabalho em execução nas edições Enterprise ou Standard para ver se os recursos específicos da edição são usados e se há alguma solução alternativa para esses recursos. Por exemplo, se você quiser ver se seus bancos de dados estão usando algum recurso de nível corporativo, você pode executar [exibições de gerenciamento dinâmico \(DMV\)](#) em todos os bancos de dados, conforme mostra o comando de exemplo a seguir.

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

Há alguns recursos da edição Enterprise que não podem ser capturados no T-SQL, como a reindexação on-line como parte dos trabalhos de manutenção do SQL. Eles devem ser verificados manualmente.

Considerações sobre a migração

A forma como você licencia o SQL Server determinará suas opções de troca de edições. As AMIs, incluindo as AMIs do SQL Server, têm o custo de licenciamento incluído no preço da instância do EC2 — o custo do licenciamento está vinculado à AMI. Você pode usar [códigos AWS de cobrança](#) para verificar a versão do SQL Server incluída na sua AMI. Para instâncias AWS com licença incluída, alterar a edição do SQL Server dentro do sistema operacional não alterará o faturamento associado à AMI. Você deve migrar seus bancos de dados para uma nova instância do EC2 com uma AMI executando a nova edição do SQL Server.

Se você estiver trazendo sua própria licença, terá mais flexibilidade. Geralmente, ainda é recomendável migrar para outra instância do EC2 que esteja executando a nova versão. Isso permite um fácil retorno de falha se algo não sair conforme o planejado. No entanto, se você precisar usar o servidor existente, ainda poderá fazer uma side-by-side instalação do SQL Server e migrar os bancos de dados entre as instâncias. Para obter etapas mais detalhadas sobre rebaixamentos

de side-by-side edições, consulte [Atualização e rebaixamento de edições no SQL Server no site MSSQLTips](#).

Recursos adicionais

- [Edições e recursos compatíveis do SQL Server 2022](#) (Microsoft Learn)
- [sys.dm_db_persisted_sku_features \(Transact-SQL\)](#) (Microsoft Learn)
- [Qual versão do SQL Server você deve usar?](#) (Brent Ozar Ilimitado)
- [AWS Pricing Calculator](#) (AWS)

Avalie a edição SQL Server Developer

Visão geral

A [edição SQL Server Developer](#) é uma edição gratuita do SQL Server que contém todos os recursos da edição Enterprise e pode ser usada em qualquer ambiente que não seja de produção. Na nuvem, onde o licenciamento da Microsoft Developer Network (MSDN) não pode ser usado, a edição SQL Server Developer é uma boa maneira de economizar custos sem precisar fornecer licenças para cargas de trabalho de desenvolvimento e teste. Isso é especialmente verdadeiro para equipes que administram grandes ambientes de desenvolvimento e teste e buscam reduzir custos desnecessários.

Um ambiente de produção é definido como um ambiente que é acessado pelos usuários finais de um aplicativo (como um site da Internet) e é usado para mais do que coletar feedback ou testar a aceitação desse aplicativo. Outros cenários que constituem ambientes de produção incluem:

- Ambientes que se conectam a um banco de dados de produção
- Ambientes que oferecem suporte à recuperação de desastres ou ao backup para um ambiente de produção
- Ambientes que são usados para produção pelo menos em parte do tempo, como um servidor que é colocado em produção durante os períodos de pico de atividade

Para obter mais informações sobre licenciamento, consulte [Amazon Web Services e Microsoft: perguntas frequentes](#) na AWS documentação.

Impacto nos custos

Se você usa a edição SQL Server Developer para cargas de trabalho que não são de produção, você pode economizar 100% dos seus custos atuais de licenciamento do SQL Server para ambientes de desenvolvimento e teste.

Versão do SQL Server	Edição SQL Server Enterprise (pacote de 2 núcleos)	Edição SQL Server Standard (pacote de 2 núcleos)	Edição SQL Server Developer
2022	\$15.123	\$3.945	Gratuito
2019	\$13.748	\$3.586	Gratuito

Note

Os preços na tabela anterior são baseados nos preços públicos da Microsoft para o [SQL Server 2022](#) e o [SQL Server 2019](#).

A tabela a seguir compara o custo de diferentes edições do SQL Server executadas com 4 vCPUs e usando preços sob demanda na região. us-east-2 Isso se aplica a cenários que dependem de instâncias com licença incluída de AWS

EC2 instance (Instância do EC2)	AMI	Preço de computação	Preço da licença do Windows	Preço da licença do SQL Server	Preço total
r5.xlarge	Linux (preços de computação)	\$183,96	–	–	\$183,96
r5.xlarge	Edição Linux + SQL Server Developer	\$183,96	\$0	\$0	\$183,96

EC2 instance (Instância do EC2)	AMI	Preço de computação	Preço da licença do Windows	Preço da licença do SQL Server	Preço total
r5.xlarge	Servidor Windows (LI)	\$183,96	\$134,32	–	\$318,28
r5.xlarge	Edição Windows + SQL Server Developer	\$183,96	\$134,32	\$0	\$318,28
r5.xlarge	Edição Web (LI) do Windows + SQL Server	\$183,96	\$134,32	\$49,64	\$367,92
r5.xlarge	Windows + SQL Server Edição Standard (LI)	\$183,96	\$134,32	\$350,4	\$668,68
r5.xlarge	Edição Windows + SQL Server Enterprise (LI)	\$183,96	\$134,32	\$1095	\$1413,28

Cenário de otimização de custos

Depois que uma empresa de integridade de dados fez uma nova aquisição, ela queria migrar a carga de trabalho recém-adquirida de sua localização atual em um provedor de hospedagem gerenciada para consolidar com suas outras cargas de trabalho no. Nuvem AWS O preço inicial mostrou que a carga de trabalho do SQL Server da empresa custaria 60% AWS mais em execução do que no atual provedor de serviços gerenciados. Uma PME da MACO avaliou a estimativa e descobriu que o cliente estava, na verdade, pagando pelas licenças do SQL Server no provedor de hospedagem gerenciada para seus ambientes de desenvolvimento e teste. Ao mudar as cargas de trabalho

não produtivas para a edição SQL Server Developer durante a migração, a empresa reduziu o licenciamento do SQL Server em 40 por cento.

Licença do SQL Server incluída no Amazon EC2

Se você tiver o SQL Server em instâncias do EC2 que usam [AMIs com licença incluída, não é possível](#) fazer uma conversão direta da edição Enterprise para a edição Developer. Os custos de licenciamento para instâncias com licença incluída estão vinculados à AMI. Mesmo que o SQL Server seja desinstalado do sistema operacional, a instância do EC2 ainda será cobrada pelos custos de licenciamento.

Para converter para a edição Developer, você deve [baixar a edição SQL Server Developer](#), instalá-la em uma nova instância do EC2 e, em seguida, migrar seus bancos de dados. Você pode migrar bancos de dados do SQL Server entre instâncias do EC2 usando uma variedade de métodos. Para obter mais informações, consulte [Métodos de migração de banco de dados do SQL Server](#) em Como migrar bancos de dados do Microsoft SQL Server para o Nuvem AWS guia. Você também pode usar a [solução Automated SQL Server Developer](#) para preparar a nova instância para a qual planeja migrar.

SQL Server BYOL no Amazon EC2

Se você tiver instâncias do SQL Server que usam um BYOL, você pode escolher entre as seguintes opções de conversão ou side-by-side downgrade no local:

- Baixe a [edição SQL Server Developer](#) no site da Microsoft. Para obter instruções de instalação manual ou automatizada, consulte a postagem [Automatizando implantações do SQL Server Developer](#) no blog. AWS
- Use o [backup e a restauração nativos do SQL Server](#) para migrar bancos de dados ou desanexar/anexar bancos de dados de uma instância do SQL para outra.
- Use uma [ferramenta de automação](#) para implantações em massa.

Note

A edição SQL Server Developer é somente para ambientes que não são de produção.

Recursos adicionais

- [Automatizando implantações do SQL Server Developer para implantar o SQL Server Developer Edition no EC2 \(Blog\)AWS](#)
- [Preços do SQL 2022 \(Microsoft\)](#)
- [Preços do SQL 2019 \(Microsoft\)](#)
- [Opções de licenciamento \(SQL Server no Amazon EC2\)](#)
- [AWS Pricing Calculator](#)(Documentação do SQL Server no Amazon EC2)
- [Guia de licenciamento do Microsoft SQL Server 2019](#) (baixe da Microsoft)
- [Edição SQL Server 2022 Developer](#) (download da Microsoft)

Avalie o SQL Server no Linux

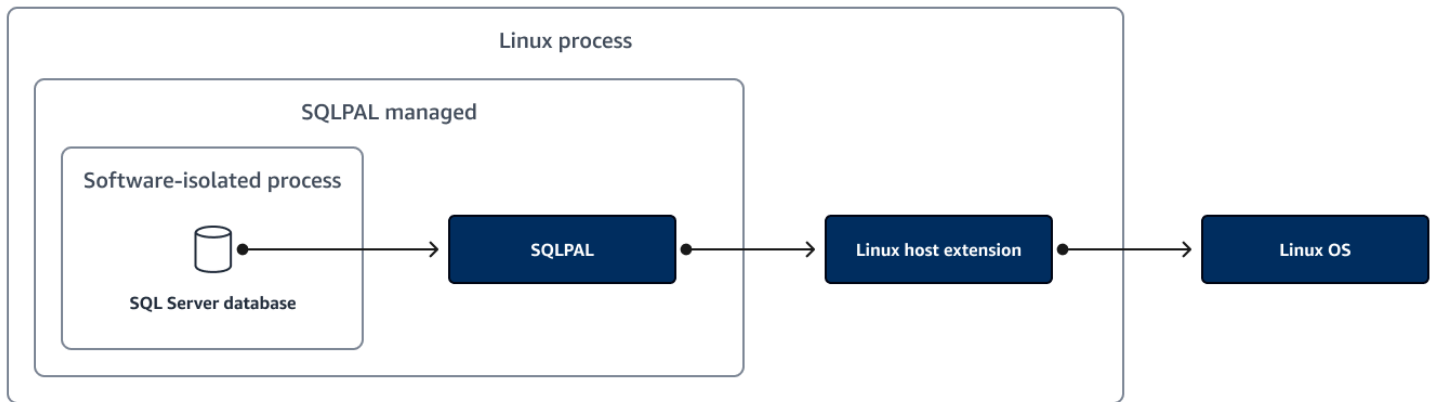
Visão geral

Desde o SQL Server 2017, é possível instalar o SQL Server em sistemas operacionais Linux. O SQL Server no Linux está pronto para uso corporativo e oferece flexibilidade, alto desempenho, recursos de segurança, TCO reduzido, recursos de HA/DR e uma ótima experiência do usuário. Você pode mudar do SQL Server no Windows Server para o SQL Server no Linux para economizar nos custos de licenciamento do Windows Server.

Para Linux, o SQL Server está disponível para implantação no Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), Ubuntu e Amazon Linux 2. O mecanismo de banco de dados do SQL Server é executado da mesma forma no Windows Server e no Linux, mas há algumas mudanças fundamentais em determinadas tarefas ao usar o Linux. Uma diferença fundamental entre a execução de aplicativos SQL Server Always On no Linux e no Windows está relacionada ao cluster de failover. Se você implantar grupos de disponibilidade Always On em um host do Windows Server, poderá aproveitar o [Clustering de Failover do Windows Server \(WSFC\) e o Active Directory como recursos internos que oferecem suporte ao agrupamento de failover](#). No entanto, nem o WSFC nem o Active Directory estão disponíveis para oferecer suporte ao agrupamento de failover no Linux. [Se você quiser iniciar o clustering de failover para o SQL Server no Linux, você pode usá-lo para simplificar AWS Launch Wizard a configuração do cluster e a instalação do SQL em instâncias Linux usando o Pacemaker. ClusterLabs](#)

O SQL Server no Windows e no Linux compartilha uma base de código comum. Ou seja, o mecanismo principal do SQL Server não foi alterado, em absoluto, para ser executado no Linux. O

SQL Server introduziu uma camada de abstração de plataforma (SQLPAL), conforme mostrado no diagrama a seguir.



O SQLPAL é responsável pela abstração de chamadas e comunicação entre o SQL Server e o sistema operacional subjacente. A extensão host é simplesmente um aplicativo Linux nativo. As funções de baixo nível do sistema operacional são chamadas nativas para otimizar o uso de E/S, memória e CPU. Quando a extensão do host é iniciada, ela carrega e inicializa o SQLPAL, que então ativa o SQL Server. O SQLPAL inicia processos de software isolados que fornecem a tradução necessária para o restante do código. Adicionar essa nova camada à arquitetura do SQL Server significa que os mesmos recursos e benefícios básicos de nível corporativo que tornaram o SQL Server tão poderoso no Windows estão disponíveis independentemente do sistema operacional.

Impacto nos custos

Para instâncias r5.2xlarge, a redução do custo de licenciamento do Windows Server é de cerca de \$268 em cada cenário. A redução é uma porcentagem maior do custo total do servidor em comparação com o uso de edições mais baratas do SQL Server. A tabela a seguir mostra a economia de custos.

Instância	Edição	Custo mensal do SQL Server no Windows	Custo mensal do SQL Server no Linux	Economia
r5.2xlarge	Web	\$735	\$466	37%
r5.2xlarge	Padrão	\$1.337	\$1.068	20%
r5.2xlarge	Enterprise	\$2.826	\$2.558	10%

Note

A estimativa de preço na tabela anterior é baseada nos preços sob demanda na us-east-1 região e pode ser visualizada diretamente no [AWS Pricing Calculator](#)

Considere um exemplo de cenário em que um cliente ISV no segmento de pequenas e médias empresas está procurando economizar custos em seu ambiente de desenvolvimento. Eles já estão usando a edição SQL Server Developer em um conjunto de servidores Windows. Ao mudar do Windows com a edição SQL Server Developer para o Linux com a edição SQL Server Developer, o cliente ISV pode economizar 33% em sua carga de trabalho de desenvolvimento. A tabela a seguir mostra os seguintes custos estimados para esse cenário.

Estimativa	Custo mensal
Windows e SQL Server	\$9.307,72
Linux e SQL Server	\$6.218,36
Economia de custos estimada	\$3.089,36 (33%)

Em outro cenário de exemplo, uma empresa migra instâncias EC2 do SQL Server incluídas na licença do Windows para o Linux. A empresa economiza um total de 300 mil dólares por ano em custos de licenciamento do Windows Server — cerca de 20% de sua fatura total. AWS

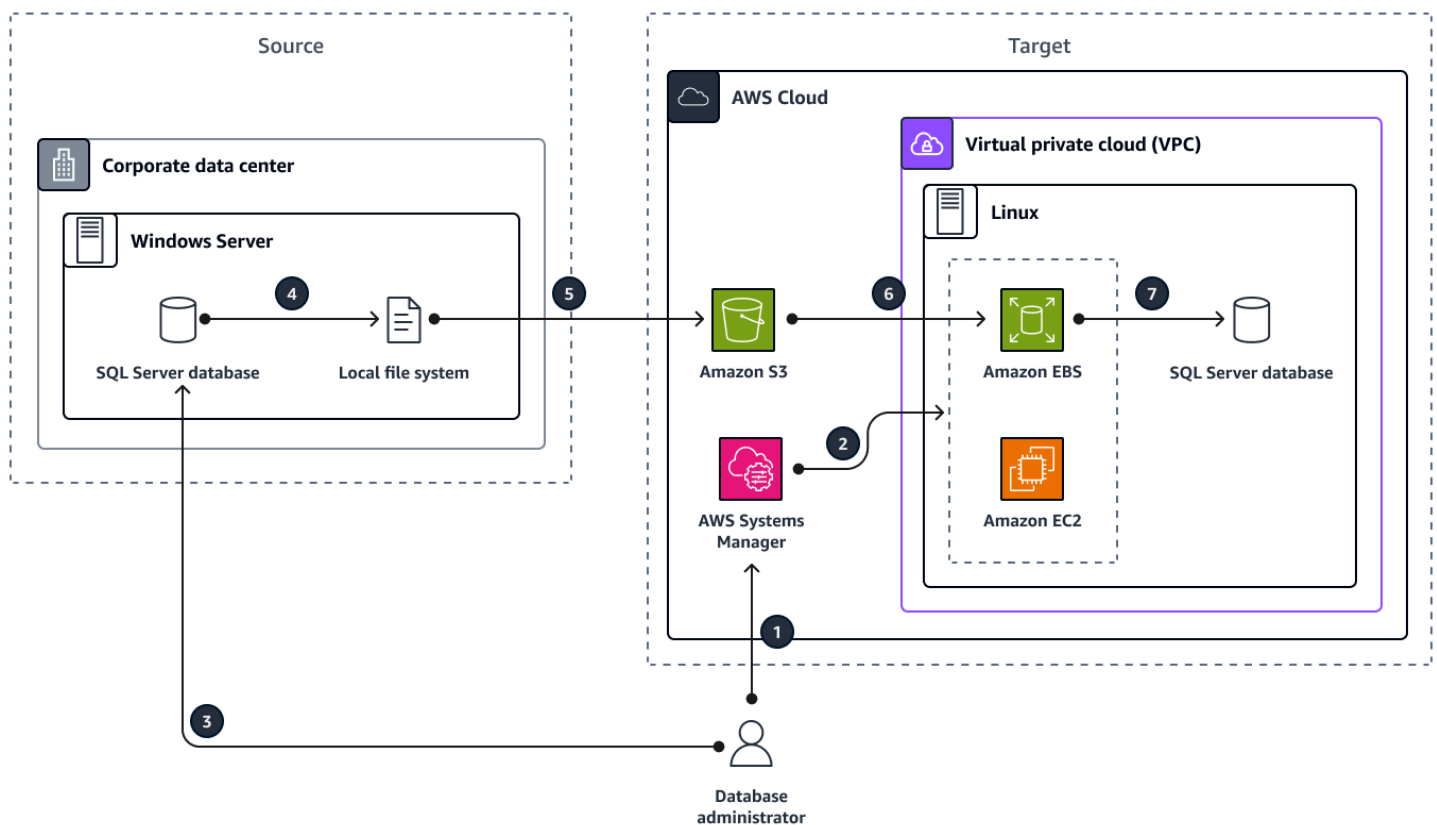
Recomendações de otimização de custos

Recomendamos que você considere o seguinte:

- O SQL Server no Linux é suportado a partir do SQL Server 2017.
- Para ajudar a fazer a mudança, você pode usar o [assistente de replataforma de Windows para Linux para bancos de dados Microsoft SQL Server](#). O assistente de replataforma é uma ferramenta de script que pode ajudá-lo a mover cargas de trabalho existentes do SQL Server dos sistemas operacionais Windows para Linux, verificando incompatibilidades comuns, exportando os bancos de dados do host Windows e importando os bancos de dados para uma instância EC2 executando o Microsoft SQL Server 2017 no Ubuntu 16.04.

- Você também pode usar os recursos de [backup e restauração](#) no SQL Server para mudar do SQL Server no Windows para o Linux.
- Você pode implantar com facilidade e rapidez no SQL Server no Linux ou Ubuntu usando [AWS Launch Wizard](#). O Launch Wizard pode implantar o SQL Server no Linux ou Ubuntu em cenários autônomos e de alta disponibilidade, com base nas necessidades do seu aplicativo. Para obter mais informações, consulte a AWS Launch Wizard postagem [Implantando no SQL Server Always on Linux com](#) no blog Microsoft Workloads on AWS .

O diagrama a seguir mostra a arquitetura de uma solução que usa o assistente de replataforma de Windows para Linux para bancos de dados Microsoft SQL Server.



Recursos adicionais

- [Visão geral do SQL Server no Linux](#) (Microsoft Learn)
- [Guia de instalação do SQL Server no Linux](#) (Microsoft Learn)
- [Implantação no SQL Server Always on Linux com AWS Launch Wizard](#) (Microsoft Workloads on AWS Blog)
- [SQL Server altamente disponível no Linux](#) (blog de código AWS aberto)

Otimize as estratégias de backup do SQL Server

Visão geral

A maioria das organizações está procurando a solução certa para proteger seus dados no SQL Server no [Amazon EC2](#) a fim de atender aos requisitos atuais de objetivo de ponto de recuperação (RPO), o tempo máximo aceitável desde o último backup e o objetivo de tempo de recuperação (RTO), o atraso máximo aceitável entre a interrupção do serviço e a restauração do serviço. Se você estiver executando o SQL Server em instâncias do EC2, você tem várias opções para criar backups de seus dados e restaurá-los. As estratégias de backup para proteger os dados do SQL Server no Amazon EC2 incluem as seguintes:

- Backup em nível de servidor usando snapshots do [Amazon Elastic Block Store \(Amazon EBS\)](#) habilitados para Windows Volume Shadow Copy Service (VSS) ou o [AWS Backup](#)
- Backup em nível de banco de dados usando [backup e restauração nativos](#) no SQL Server

Você tem as seguintes opções de armazenamento para backup nativo em [nível de banco de dados](#):

- Um backup local com um volume [do Amazon EBS](#)
- Um backup do sistema de arquivos de rede com o [Amazon FSx for Windows File Server](#) ou o Amazon FSx for ONTAP NetApp
- Um backup de rede para o Amazon Simple Storage Service (Amazon S3) usando [AWS Storage Gateway](#)
- Backup direto no Amazon S3 para SQL Server 2022

Esta seção faz o seguinte:

- Destaca recursos para ajudar você a economizar espaço de armazenamento
- Compara os custos entre as diferentes opções de armazenamento de back-end
- Fornece links para documentação detalhada para ajudar a implementar essas recomendações

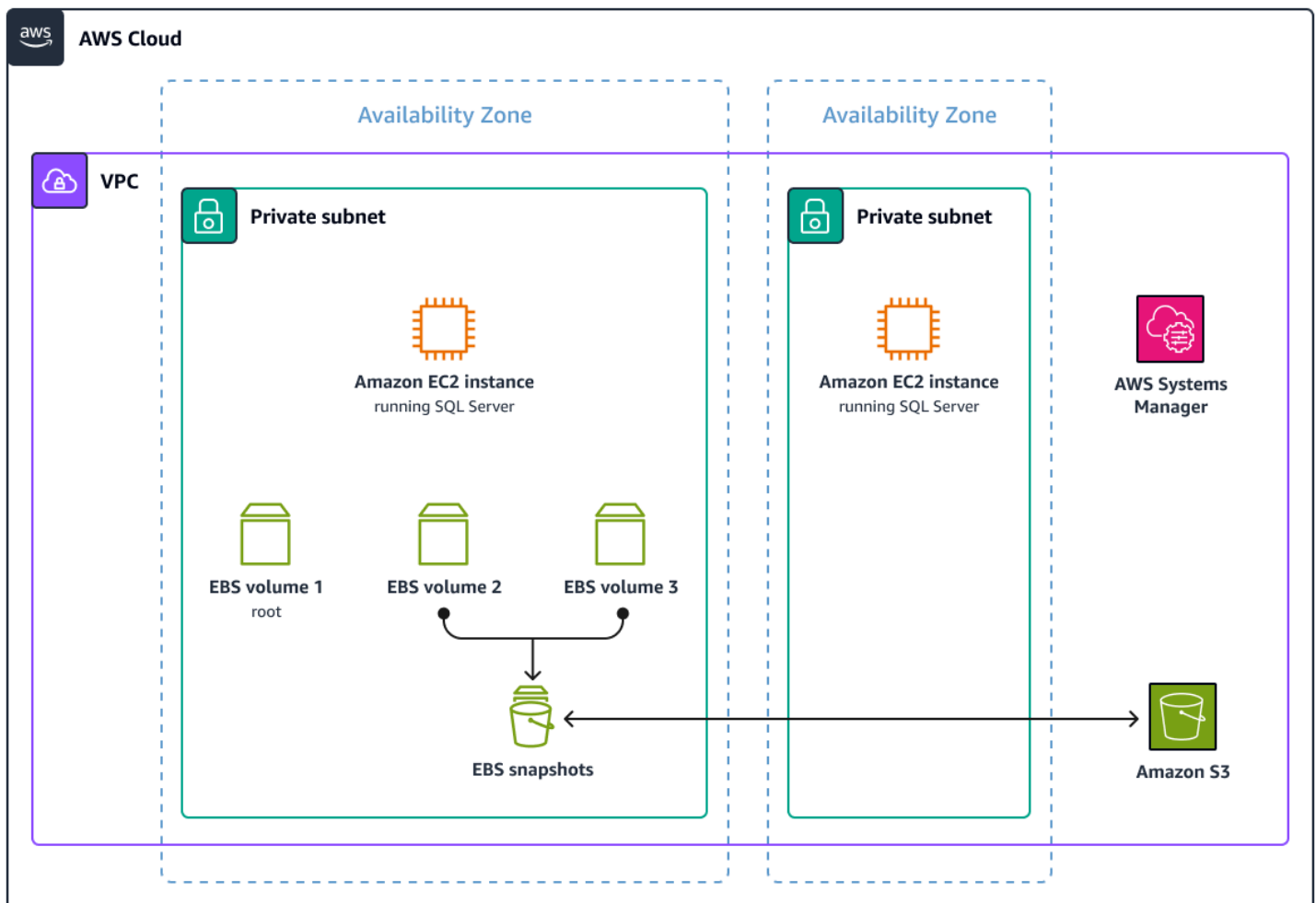
Backup em nível de servidor usando instantâneos habilitados para VSS

Uma arquitetura de instantâneos habilitada para VSS usa o [comando AWS Systems Manager Executar](#) para instalar o agente VSS em suas instâncias do SQL Server. Você também pode usar o comando Executar para invocar todo o fluxo de trabalho de descarga dos buffers do sistema

operacional e do aplicativo no disco, pausar as operações de E/S, tirar um point-in-time instantâneo dos volumes do EBS e, em seguida, retomar a E/S.

O Run Command cria snapshots automatizados de todos os volumes do EBS anexados a uma instância de destino. Você também tem a opção de excluir o volume raiz, pois os arquivos de banco de dados do usuário geralmente são armazenados em outros volumes. Se você fizer o stripe de vários volumes do EBS para criar um único sistema de arquivos para arquivos do SQL Server, o Amazon EBS também oferecerá suporte a snapshots em vários volumes consistentes em caso de falhas usando um único comando de API. Para obter mais informações sobre [snapshots do EBS compatíveis com VSS consistentes com aplicativos, consulte Criar um instantâneo VSS consistente com aplicativos](#) na documentação do Amazon EC2.

O diagrama a seguir mostra uma arquitetura para backup em nível de servidor usando instantâneos habilitados para VSS.



Considere os seguintes benefícios do uso de snapshots habilitados para VSS:

- O primeiro snapshot de uma instância de banco de dados contém os dados da instância de banco de dados completa. Os snapshots subsequentes da mesma instância de banco de dados são [incrementais](#), o que significa que somente os dados que foram alterados depois do snapshot mais recente serão salvos.
- Os instantâneos do EBS fornecem point-in-time recuperação.
- É possível [restaurar para uma nova instância do EC2 do SQL Server desde um snapshot](#).
- Se uma instância for criptografada usando o Amazon EBS ou se um banco de dados for criptografado na instância usando o TDE, essa instância ou banco de dados será restaurado automaticamente com a mesma criptografia.
- É possível copiar seus [backups automatizados entre regiões](#).
- Quando você restaura um volume do EBS a partir de um snapshot, ele fica imediatamente disponível para que os aplicativos o acessem. Isso significa que você pode colocar o SQL Server online imediatamente após restaurar de snapshots um ou mais de seus volumes subjacentes do EBS.
- Por padrão, os volumes restaurados buscam blocos subjacentes do Amazon S3 na primeira vez que uma aplicação tenta lê-los. Isso significa que pode haver um atraso na performance após a restauração de um volume do EBS a partir de um snapshot. O volume em algum momento atingirá a performance nominal. No entanto, é possível evitar esse atraso usando [snapshots de restauração rápida \(FSR\)](#).
- Você pode usar o [gerenciamento do ciclo de vida para snapshots do EBS](#).

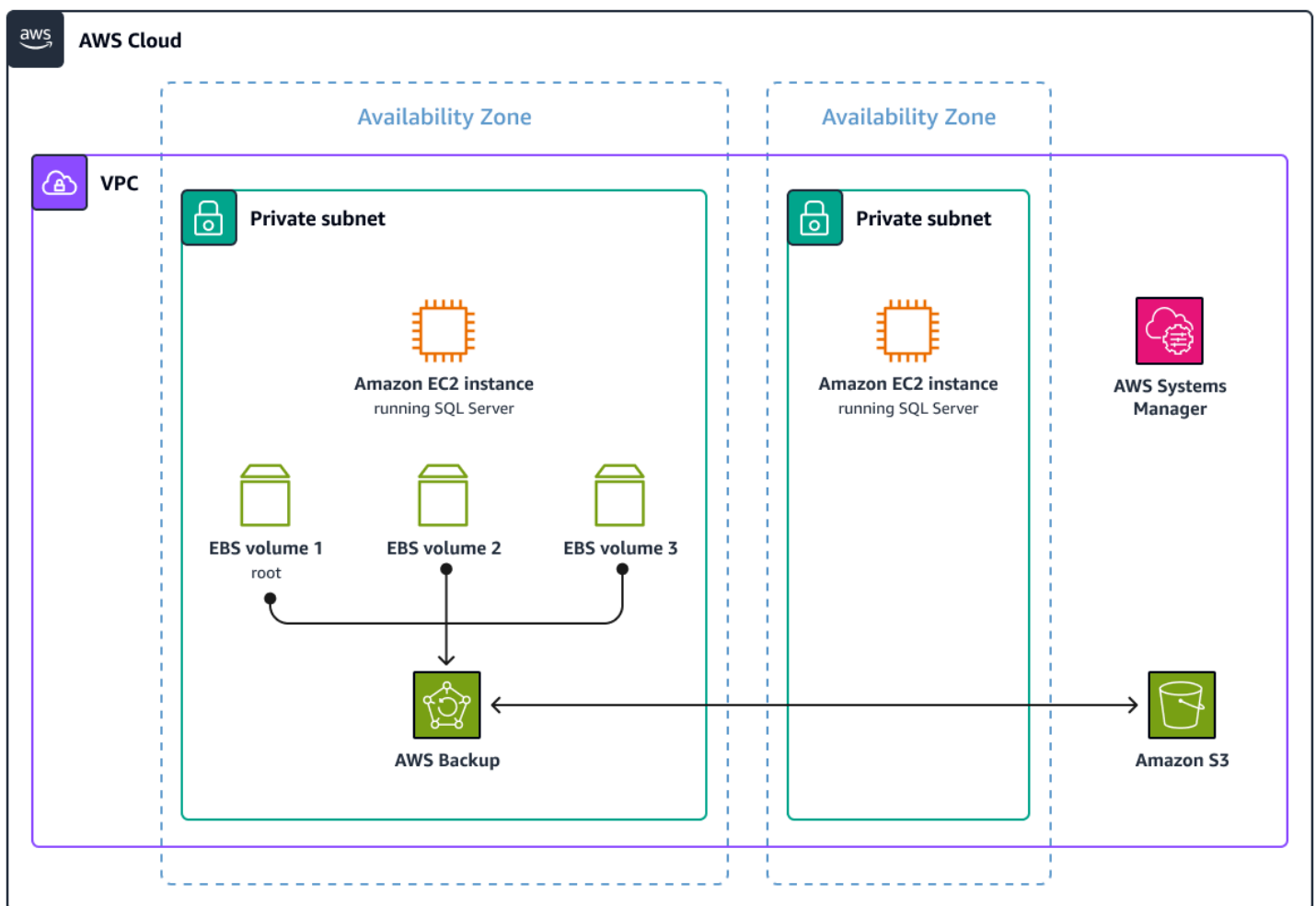
Considere as seguintes limitações do uso de instantâneos habilitados para VSS:

- Você não pode realizar a point-in-time recuperação entre regiões com um instantâneo criptografado para uma instância do SQL Server.
- Você não pode criar um snapshot criptografado de uma instância não criptografada.
- Você não pode restaurar um banco de dados individual porque o snapshot é obtido no nível do volume do EBS.
- Você não pode restaurar a instância para si mesma.
- Um snapshot da instância de banco de dados deve ser criptografado usando a mesma chave AWS Key Management Service (AWS KMS) da instância de banco de dados.
- A E/S de armazenamento é suspensa por uma fração de segundo (aproximadamente 10 milissegundos) durante o processo de backup do snapshot.

Backup do SQL Server usando AWS Backup

Você pode usar [AWS Backup](#) para centralizar e automatizar a proteção de dados em todo o mundo. Serviços da AWS oferece uma solução econômica, totalmente gerenciada e baseada em políticas que simplifica a proteção de dados em grande escala. AWS Backup também ajuda você a apoiar suas obrigações de conformidade regulatória e a cumprir suas metas de continuidade de negócios. Junto com AWS Organizations, AWS Backup permite que você implante centralmente políticas de proteção de dados (backup) para configurar, gerenciar e governar sua atividade de backup em toda a organização Contas da AWS e nos recursos.

O diagrama a seguir mostra a arquitetura de uma solução de backup e restauração para SQL Server no EC2 usando AWS Backup.



Considere os seguintes benefícios de fazer backup do SQL Server usando AWS Backup:

- É possível automatizar os cronogramas de backup, o gerenciamento de retenção e o gerenciamento do ciclo de vida.

- Você pode centralizar sua estratégia de backup em toda a organização, abrangendo várias contas e Regiões da AWS
- Você pode centralizar o monitoramento de sua atividade de backup e os alertas. Serviços da AWS
- Também é possível implementar backups entre regiões para planejar a recuperação de desastres.
- A solução oferece suporte a backups entre várias contas.
- É possível fazer backups seguros usando criptografia de backup secundária.
- Todos os backups oferecem suporte à criptografia usando chaves AWS KMS de criptografia.
- A solução funciona com o TDE.
- Você pode restaurar para um ponto de recuperação específico via console do AWS Backup .
- É possível fazer backup de uma instância do SQL Server inteira, o que inclui todos os bancos de dados do SQL Server.

Backup em nível de banco de dados

Essas abordagens usam a funcionalidade nativa de backup do Microsoft SQL Server. Você pode fazer backups de bancos de dados individuais na instância do SQL Server e restaurar um banco de dados individual.

Cada uma dessas opções para backup e restauração nativas do SQL Server também oferece suporte a:

- Compactação e backup de vários arquivos
- Backups completos, diferenciais e T-log
- Bancos de dados criptografados com TDE

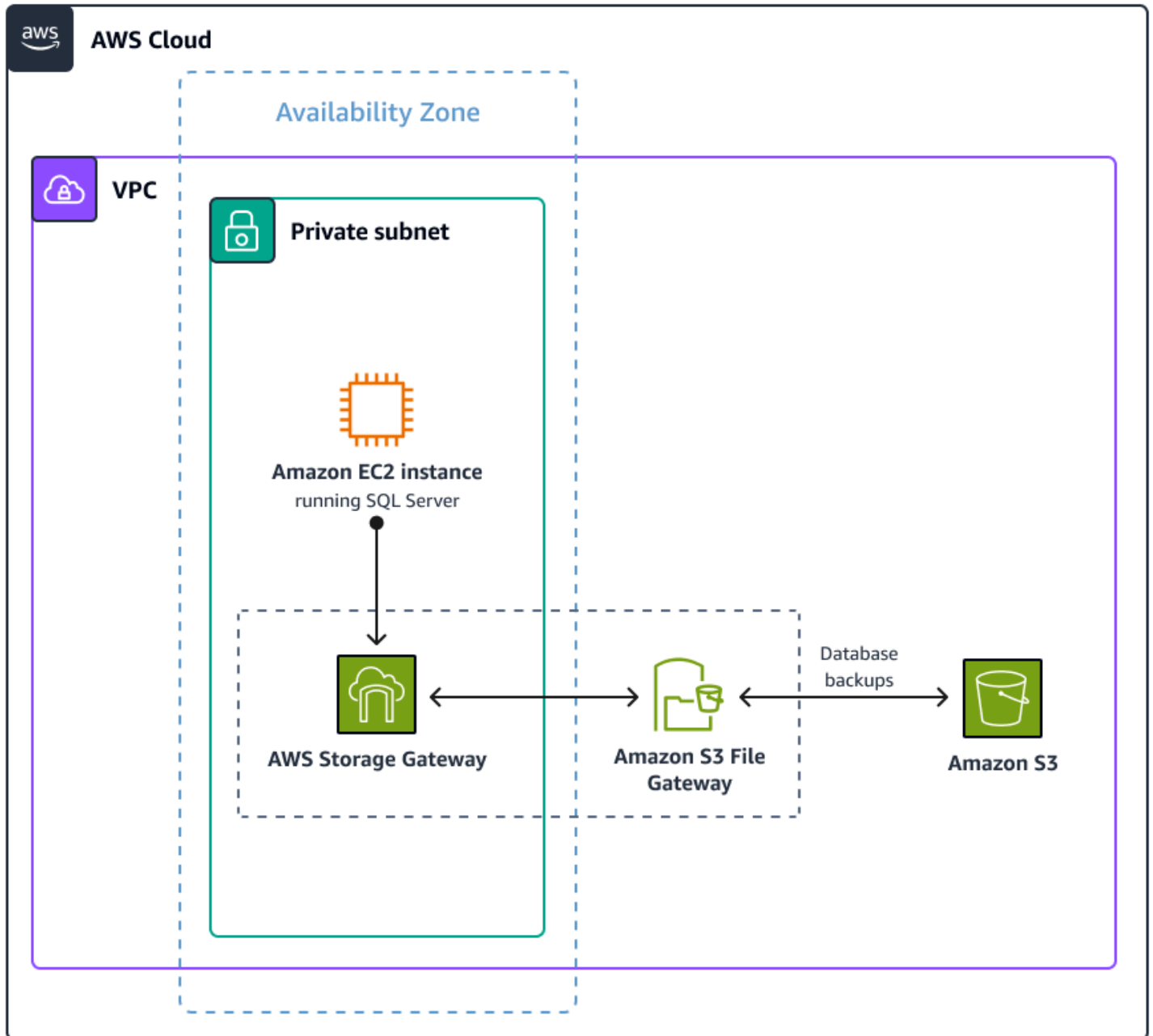
Backup e restauração nativos do SQL Server no Amazon S3

O SQL Server no Amazon EC2 oferece suporte a backup e restauração nativos para bancos de dados SQL Server. É possível fazer um backup do banco de dados do SQL Server e depois restaurar o arquivo de backup em um banco de dados existente ou em uma nova instância do EC2 do SQL Server, Amazon RDS para SQL Server ou em um servidor on-premises.

O Storage Gateway é um serviço de armazenamento em nuvem híbrida que fornece às aplicações on-premises acesso a armazenamento em nuvem praticamente ilimitado. Você pode usar o Storage Gateway para fazer backup de seus bancos de dados Microsoft SQL Server diretamente no Amazon

S3, reduzindo seu espaço de armazenamento local e usando o Amazon S3 para armazenamento durável, escalável e econômico.

O diagrama a seguir mostra a arquitetura de uma solução nativa de backup e restauração que usa o Storage Gateway e o Amazon S3.



Considere os seguintes benefícios de usar o backup nativo do SQL Server com o Storage Gateway:

- É possível mapear um gateway de armazenamento como um compartilhamento de arquivos Server Message Block (SMB) na instância do EC2 e enviar o backup para o Amazon S3.

- O backup vai diretamente para o bucket do S3 ou por meio do cache de arquivos do Storage Gateway.
- Há suporte a backups de vários arquivos.

Considere as seguintes limitações do backup nativo usando o Storage Gateway:

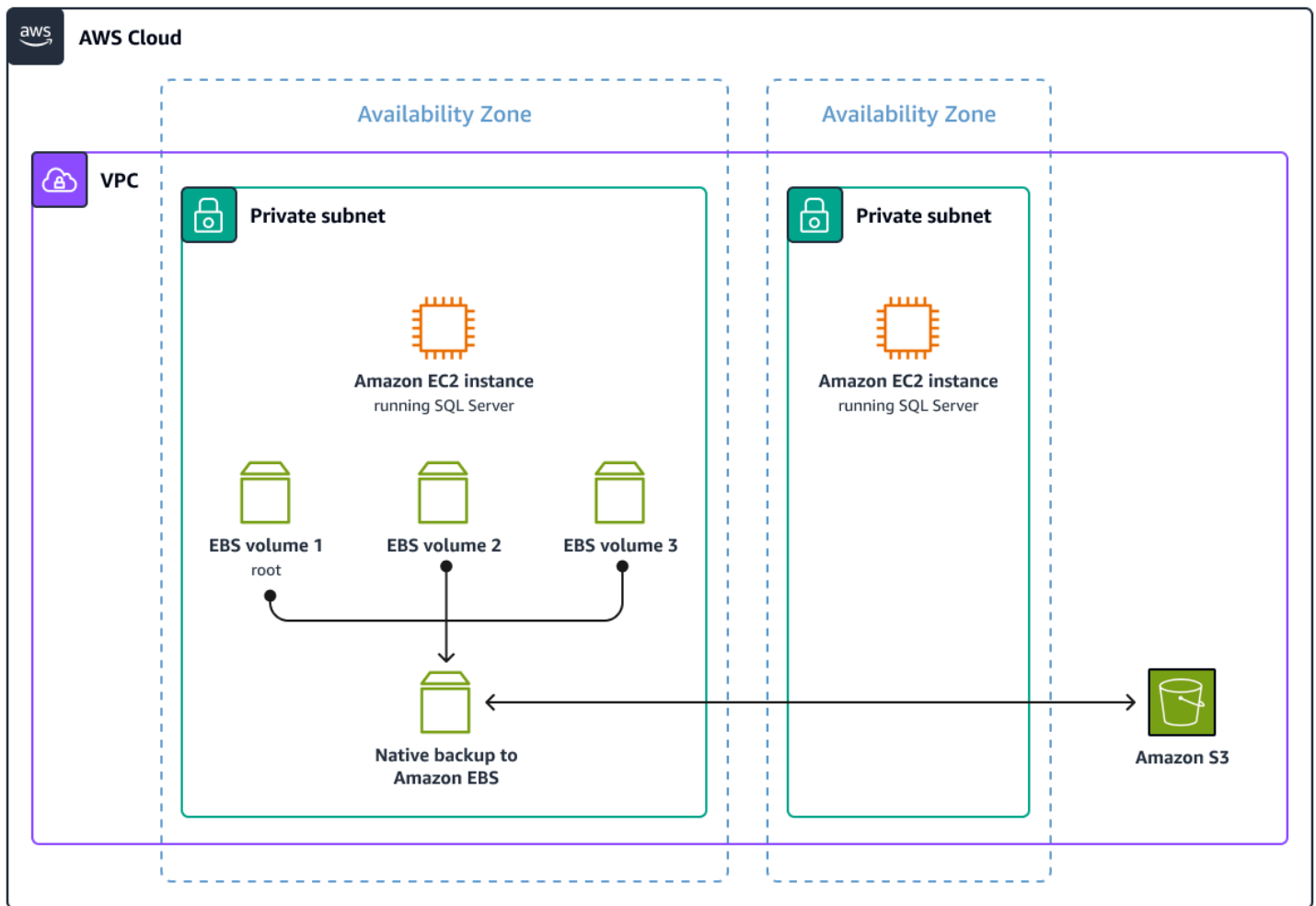
- É necessário configurar o backup e a restauração para cada banco de dados individual.
- É necessário gerenciar a [Política de ciclo de vida do Amazon S3](#) para os arquivos de backup.

Para obter mais informações sobre como configurar o Storage Gateway, consulte a publicação [Armazenar backups do SQL Server no Amazon S3 usando](#) o AWS Storage Gateway AWS blog.

Backup nativo do SQL Server em volumes do EBS

Você pode fazer um backup nativo do seu banco de dados SQL Server e armazenar o arquivo em um volume do Amazon EBS. O Amazon EBS é um serviço de armazenamento em blocos de alta performance. Os volumes do EBS são elásticos, o que suporta criptografia. Eles podem ser desanexados e anexados a uma instância do EC2. Você pode fazer backup do SQL Server em uma instância do EC2 no mesmo tipo de volume do EBS ou em um tipo diferente de volume do EBS. Uma vantagem de fazer backup em um volume diferente do EBS é a economia de custos.

O diagrama a seguir mostra a arquitetura de um backup nativo em um volume do EBS.



Considere os seguintes benefícios de usar o backup nativo do SQL Server para volumes do EBS:

- Você pode fazer backups de bancos de dados individuais em uma instância do EC2 do SQL Server e restaurar um banco de dados individual em vez de precisar restaurar a instância completa.
- Há suporte a backups de vários arquivos.
- É possível programar trabalhos de backup usando o SQL Server Agent e o mecanismo de trabalho do SQL Server.
- Você pode obter benefícios de performance por meio de suas opções de hardware. Por exemplo, você pode usar volumes de armazenamento st1 para obter maior throughput.

Considere as seguintes limitações do uso do backup nativo em volumes do EBS:

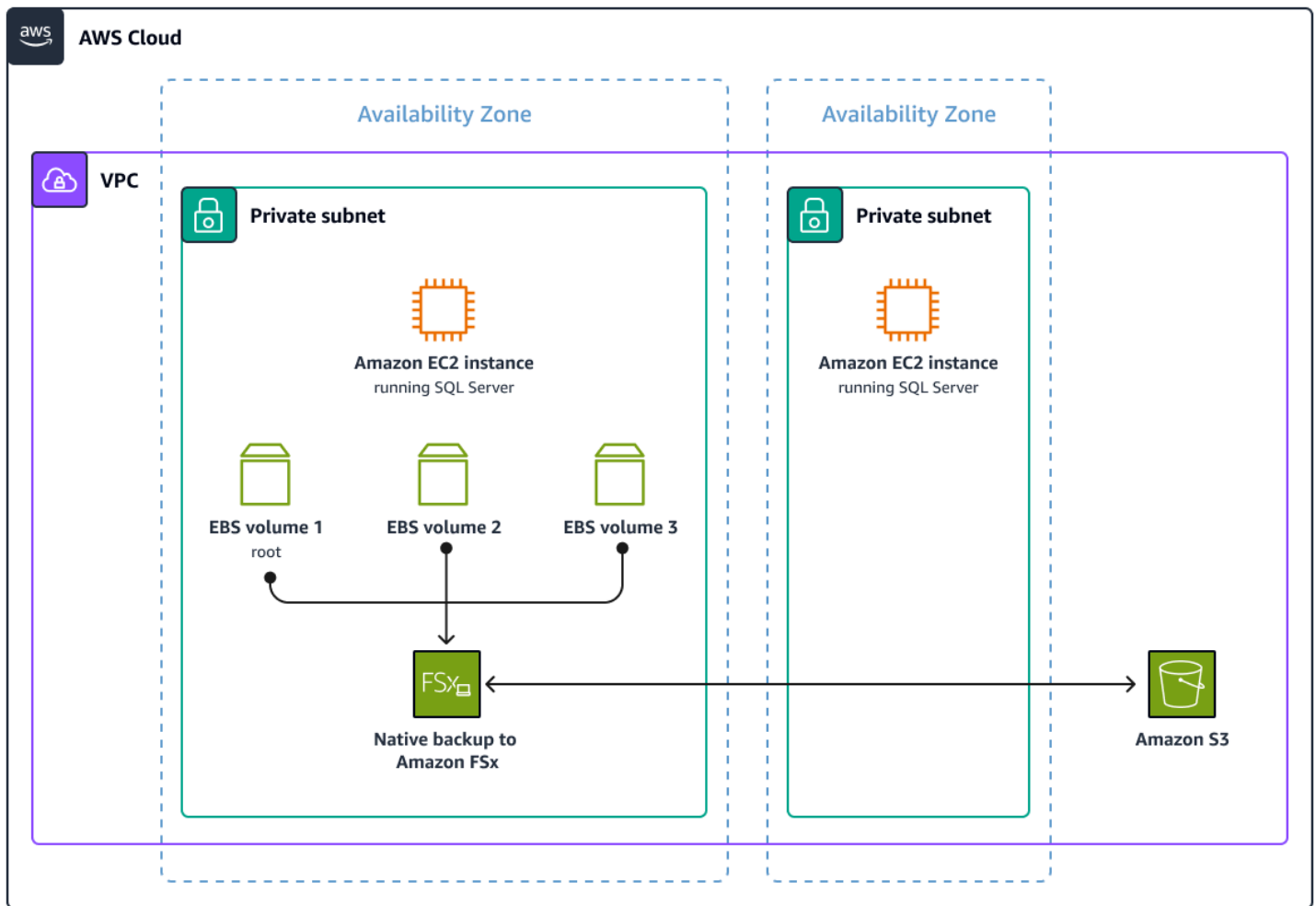
- É necessário mover manualmente os backups do volume do EBS para o Amazon S3.

- Para backups grandes, você deve gerenciar o espaço em disco no Amazon EC2.
- Na instância do EC2, a throughput do Amazon EBS pode ser um gargalo.
- É necessário armazenamento adicional para armazenar backups no Amazon EBS.

Backup nativo do SQL Server no Amazon FSx para Windows File Server

O [Amazon FSx for Windows File Server](#) é um sistema de arquivos Windows nativo totalmente gerenciado que oferece até 64 TB de armazenamento projetado para oferecer desempenho rápido, previsível e consistente. AWS introduziu [suporte nativo para implantações de sistemas de arquivos Multi-AZ](#) no FSx for Windows File Server. O suporte nativo facilita a implantação do armazenamento de arquivos do Windows AWS com alta disponibilidade e redundância em várias zonas de disponibilidade. AWS também introduziu suporte para [compartilhamentos de arquivos SMB Continuously Available \(CA\)](#). É possível usar o FSx para Windows File Server como armazenamento de backup de um banco de dados do SQL Server.

O diagrama a seguir mostra a arquitetura de um backup nativo do SQL Server no FSx para Windows File Server.



Considere os seguintes benefícios de usar o backup nativo do SQL Server no FSx for Windows File Server:

- Você pode fazer backup do seu banco de dados SQL Server em um compartilhamento de arquivos do Amazon FSx.
- Você pode fazer backups de bancos de dados individuais em uma instância do SQL Server e restaurar um banco de dados individual em vez de precisar restaurar a instância completa.
- Há suporte para backups em várias partes.
- É possível programar trabalhos de backup usando o agente e o mecanismo de trabalho do SQL Server.
- As instâncias têm maior largura de banda de rede em comparação com o Amazon EBS.

Considere as seguintes limitações do uso do backup nativo do SQL Server no FSx for Windows File Server:

- Você deve mover manualmente os backups do Amazon FSx para o Amazon S3 usando ou. AWS Backup AWS DataSync
- Backups grandes podem exigir uma sobrecarga adicional para o gerenciamento do espaço em disco no Amazon FSx.
- A throughput da rede da instância do EC2 pode ser um gargalo.
- É necessário armazenamento adicional para armazenar backups no FSx para Windows File Server.

Backup do SQL Server no Amazon FSx para ONTAP NetApp

Os instantâneos com FSx for ONTAP são sempre consistentes com falhas, mas exigem que você desative (ou pause a E/S) do seu banco de dados para criar um instantâneo consistente com o aplicativo. Você pode usar NetApp SnapCenter (uma ferramenta de orquestração com plug-ins para aplicativos específicos, incluindo SQL Server) com o FSx for ONTAP para criar instantâneos consistentes com aplicativos e proteger, replicar e clonar seus bancos de dados sem custo adicional.

NetApp SnapCenter

NetApp SnapCenter é uma plataforma unificada para proteção de dados consistente com aplicativos. SnapCenter refere-se aos instantâneos como backups. Este guia adota a mesma convenção de nomenclatura. SnapCenter fornece um painel único para gerenciar backups, restaurações e clones consistentes com aplicativos. Você adiciona um SnapCenter plug-in para seu aplicativo de banco de dados específico para criar backups consistentes com o aplicativo. O SnapCenter plug-in para SQL Server fornece a seguinte funcionalidade que simplifica seu fluxo de trabalho de proteção de dados.

- Opções de backup e restauração com granularidade para backups completos e de log
- Restauração no local e restauração em um local alternativo

Para obter mais informações sobre isso SnapCenter, consulte a postagem [Proteja suas cargas de trabalho do SQL Server usando o NetApp SnapCenter Amazon FSx NetApp for ONTAP no AWS blog de armazenamento](#).

Otimização de custos para backups

As opções a seguir podem ajudá-lo a reduzir o custo de armazenamento de backups do SQL Server no AWS.

- Ative [a compactação do SQL Server](#) durante a criação do arquivo de backup e envie o menor arquivo possível para o armazenamento. Por exemplo, uma taxa de compactação de 3:1 indica que você está economizando cerca de 66% em espaço em disco. Para consultar essas colunas, você pode usar a seguinte instrução Transact-SQL: `SELECT backup_size/compressed_backup_size FROM msdb..backupset;`
- Para backups que vão para buckets S3, habilite a classe de armazenamento [Amazon S3 Intelligent-Tiering](#) para reduzir os custos de armazenamento em 30%.
- Para backups que vão para FSx for Windows File Server ou FSx for ONTAP, use uma única zona de disponibilidade para economizar 50% (em comparação com o uso de várias zonas de disponibilidade). Para obter informações sobre preços, consulte os preços do [Amazon FSx for Windows File Server e os preços do Amazon FSx NetApp](#) for ONTAP.
- A opção mais eficiente para o SQL Server 2022 é o backup direto para o Amazon S3. Você pode economizar custos adicionais evitando o Storage Gateway.

Resultados do teste de benchmark para backups

Esta seção compara as opções a seguir, do ponto de vista de custo e desempenho, para um banco de dados de amostra de 1 TB, com base nos resultados dos testes de benchmark de desempenho nas soluções de backup abordadas neste guia.

- Especificação de instância EC2 — r5d.8xlarge com Windows Server 2019 e SQL Server 2019 Developer Edition
- Especificação do banco de dados — 1 TB de tamanho com o TDE desativado

Os testes foram realizados com uma instância r5d.8xlarge e um banco de dados SQL Server de 1 TB como fonte. O sistema de origem foi configurado de acordo com as melhores práticas, e o banco de dados de origem continha quatro arquivos de dados (250 GB cada) e um arquivo de log (50 GB) distribuídos em volumes gp3 separados. O BACKUP comando nativo do SQL Server inclui a gravação em 10 arquivos de backup, usando a compactação para otimizar o desempenho do backup e reduzir a quantidade de dados enviados pela rede e gravados no destino. Em todos os casos de teste, o desempenho do armazenamento foi o gargalo.

Há uma variedade quase infinita de configurações possíveis para esses tipos de teste. Esse teste se concentrou na otimização de desempenho, custo, escalabilidade e casos de uso reais. A tabela a seguir mostra as métricas de desempenho que foram capturadas para as opções de destino de backup.

Opções de backup	Nível	Duração da execução (Appx)	Taxa de backup	Custo em USD por mês*
Backup nativo em HDD local EBS st1, 2 TB	Banco de dados	00:30:46 min	554,7 Mbps	\$92,16
Backup nativo para EBS SSD gp3 local, 2 TB	Banco de dados	00:22:00 min	512 Mbps	\$193,84
Backup nativo para FSx for Windows File Server HDD, taxa de transferência de 2 TB @512 Mbps	Banco de dados	00:20:58 min	814,0 Mbps	\$1.146
Backup nativo para FSx for Windows File Server SSD, taxa de transferência de 2 TB @512 Mbps	Banco de dados	00:20:00 min	814,0 Mbps	\$1.326
Backup nativo para o S3 File Gateway m6i.4xlarge (16 vCPU, 64 GB) com 2 TB gp3	Banco de dados	00:23:20 min	731,5 Mbps	\$470,42
Instantâneo do EBS VSS	Volume do EBS	00:00:02 seg 00:00:53 seg	Instantâneo N/A	\$51

Opções de backup	Nível	Duração da execução (Appx)	Taxa de backup	Custo em USD por mês*
AWS Backup (Backup da AMI)	AMI	00:00:04 seg 00:08:00 min	Instantâneo N/A	<u>\$75</u>
Backup nativo do SQL Server diretamente no Amazon S3 (SQL Server 2022)	Banco de dados	00:12:00 min	731,5 Mbps	<u>Primeiros 50 TB/ mês, 0,023 USD por GB 23,55 USD por mês</u>
Backup nativo para FSx for ONTAP (usando SnapCenter)	Banco de dados	–	–	<u>\$440,20</u>

A tabela anterior pressupõe o seguinte:

- Os custos de transferência de dados e do Amazon S3 não estão incluídos.
- O preço do armazenamento está incluído no preço da instância.
- Os custos são baseados na us-east-1 região.
- A taxa de transferência e o IOPS aumentam em 10% com vários backups que têm uma taxa geral de alteração de 10% ao longo do mês.

Os resultados do teste mostram que a opção mais rápida é um backup nativo do banco de dados SQL Server para FSx for Windows File Server. Um backup no Storage Gateway e nos volumes do EBS conectados localmente é a opção mais econômica, mas tem um desempenho mais lento. Para backups em nível de servidor (AMI), recomendamos o uso AWS Backup para otimizar o desempenho, o custo e a capacidade de gerenciamento.

Recomendações de otimização de custos

Compreender as possíveis soluções para fazer backup do SQL Server no Amazon EC2 é fundamental para proteger seus dados, garantir que você atenda às suas necessidades de backup e

implementar um plano para se recuperar de eventos críticos. As diferentes maneiras de fazer backup e restaurar suas instâncias e bancos de dados do SQL Server exploradas nesta seção podem ajudá-lo a criar uma estratégia de backup e restauração que proteja seus dados e atenda aos requisitos da sua organização.

Esta seção aborda as seguintes opções de backup:

- Compactação
- Amazon S3 Intelligent-Tiering
- Zona de disponibilidade única
- Backup em URL

A orientação fornecida para cada uma dessas opções é de alto nível. Se você deseja implementar alguma dessas recomendações em sua organização, recomendamos que entre em contato com a equipe da sua conta. A equipe pode então se envolver com um Microsoft Specialist SA para conduzir a conversa. Você também pode entrar em contato pelo e-mail optimize-microsoft@amazon.com.

Em resumo, recomendamos o seguinte:

- Se você estiver usando o SQL Server 2022, fazer backup no Amazon S3 é a opção mais econômica.
- Se você estiver usando o SQL Server 2019 e edições anteriores do SQL Server, considere fazer backup no Storage Gateway com suporte do Amazon S3 como a opção mais econômica.

Compactação

O objetivo da compactação é ter menos armazenamento consumido por cada backup, o que é benéfico para várias opções de armazenamento. Você deve habilitar a compactação para um backup do SQL Server no nível da [instância do SQL Server](#). O exemplo a seguir mostra como adicionar a palavra-chave de compressão a um banco de dados de backup:

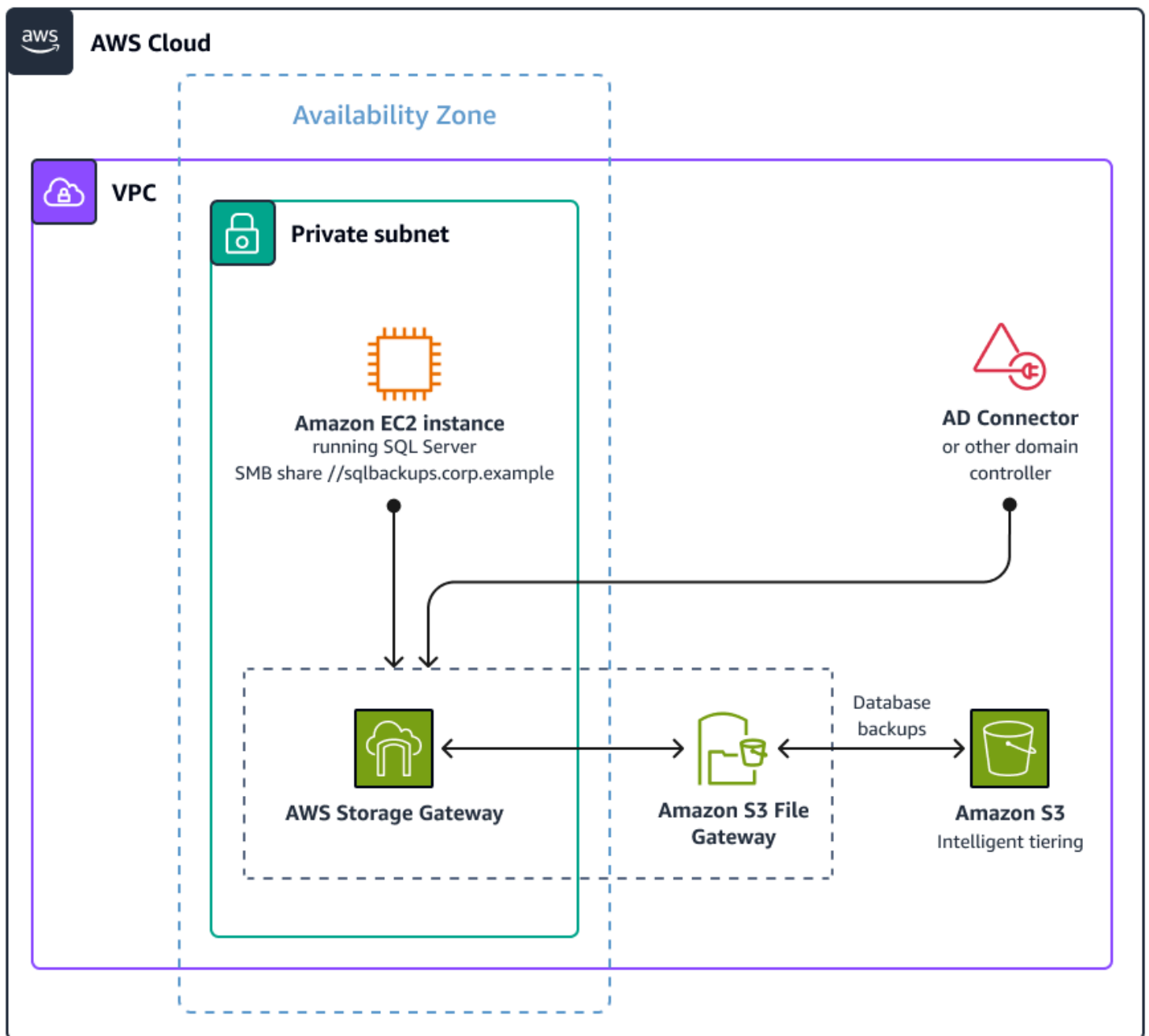
```
BACKUP DATABASE <database_name> TO DISK WITH COMPRESSION (ALGORITHM = QAT_DEFLATE)
```

Amazon S3 Intelligent-Tiering

[Para backups que vão para buckets do Amazon S3, você pode habilitar o Amazon S3 Intelligent-Tiering como sua classe de armazenamento do Amazon S3 File Gateway.](#) Isso pode reduzir os

custos de armazenamento em até 30%. Em seguida, você monta o S3 File Gateway em seus servidores SQL usando um compartilhamento de arquivos SMB que pode ser integrado ao seu domínio do [Active Directory](#). Isso fornece controle de acesso para seu compartilhamento, a capacidade de aproveitar contas de serviço existentes e acesso ao Amazon S3 usando um protocolo de arquivo comum focado na Microsoft. Para contas que talvez não tenham conectividade direta com um controlador de domínio, você pode usar o [Conector do Active Directory](#) para facilitar a comunicação com o Active Directory no local ou na nuvem. Para definir as configurações do Active Directory no gateway, você deve especificar os IPs do conector do Active Directory para que o controlador de domínio faça solicitações de proxy para o Active Directory.

O diagrama a seguir mostra uma arquitetura para uma solução baseada no S3 Intelligent-Tiering.



Por padrão, os arquivos de backup gravados no bucket do S3 usam o nível Standard. Para converter os arquivos de backup do nível Standard para o S3 Intelligent-Tiering, você deve [criar](#) uma regra de ciclo de vida. Você também pode usar o [AWS Management Console](#) para ativar o S3 Intelligent-Tiering. Para obter mais informações, consulte [Conceitos básicos do Amazon S3 Intelligent-Tiering](#) na documentação. AWS

Zona de disponibilidade única

Para criar um sistema de arquivos de zona de disponibilidade única, escolha a opção Single-AZ ao [criar um sistema de arquivos FSx for Windows File Server](#). O Amazon FSx também faz backups altamente duráveis (armazenados no Amazon S3) do seu sistema de arquivos diariamente usando o Windows Volume Shadow Copy Service e permite que você faça backups adicionais a qualquer momento. Lembre-se de alguns dos problemas com o uso de uma única zona de disponibilidade. Por exemplo, o compartilhamento de arquivos SMB fica inacessível se uma zona de disponibilidade afetada na qual o sistema de arquivos é provisionado ficar inativa por horas seguidas. Se você precisar de acesso aos dados, deverá restaurá-los a partir de backups em uma zona de disponibilidade disponível na região de origem. Para obter mais informações, consulte a seção [Usar uma única zona de disponibilidade](#) deste guia.

Backup em URL

Para o SQL Server 2022, o recurso [de backup em URL permite](#) o backup direto para o Amazon S3. Essa é a abordagem de backup ideal para o SQL Server 2022 em execução, AWS pois você obtém o conjunto completo de recursos do Amazon S3 na camada de armazenamento e remove o custo do AWS Storage Gateway dispositivo necessário nas versões anteriores para facilitar essa funcionalidade. Há dois custos principais a serem considerados ao implementar esse recurso: custos de transferência de dados e a classe de armazenamento S3 escolhida. [Se você quiser os recursos nativos de recuperação de desastres do Amazon S3, deve considerar que a replicação entre regiões incorre em custos de saída de dados entre regiões](#). Para saber mais sobre como configurar essa opção, consulte a postagem [Backup de bancos de dados SQL Server no Amazon S3](#) no blog Microsoft Workloads on. AWS

Recursos adicionais

- [Opções de backup e restauração para SQL Server no Amazon EC2 \(orientação AWS prescritiva\)](#)
- [oint-in-time Recuperação de P e backup contínuo para Amazon RDS com AWS Backup](#) (AWS Storage Blog)
- [Proteja suas cargas de trabalho do SQL Server usando o NetApp SnapCenter Amazon FSx NetApp for ONTAP AWS](#) (Storage Blog)
- [Introdução ao uso do Amazon S3 Intelligent-Tiering](#) (Centro de recursos para começar a usar)AWS
- [Estratégias de backup e restauração do Amazon RDS for SQL Server](#)AWS (blog do banco de dados)

- [Migre um banco de dados local do Microsoft SQL Server para o Amazon EC2 \(orientação prescritiva\)](#) AWS
- [Melhores práticas para implantar o Microsoft SQL Server no Amazon EC2](#) AWS (Whitepaper)

Modernize os bancos de dados do SQL Server

Visão geral

Se você está iniciando uma jornada para modernizar bancos de dados legados para escalabilidade, desempenho e otimização de custos, talvez esteja enfrentando desafios com bancos de dados comerciais como o SQL Server. Os bancos de dados comerciais são caros, prendem os clientes e oferecem termos de licenciamento punitivos. Esta seção fornece uma visão geral de alto nível das opções de migração e modernização do SQL Server para bancos de dados de código aberto e informações sobre como escolher a melhor opção para sua carga de trabalho.

Você pode refatorar seus bancos de dados do SQL Server em bancos de dados de código aberto, como o Amazon Aurora PostgreSQL, para economizar nos custos de licenciamento do Windows e do SQL Server. Bancos de dados modernos nativos da nuvem, como o Aurora, combinam a flexibilidade e o baixo custo dos bancos de dados de código aberto com os recursos robustos e de nível corporativo dos bancos de dados comerciais. [Se você tiver cargas de trabalho variáveis ou cargas de trabalho de vários locatários, também poderá migrar para o Aurora serverless V2](#). Isso pode reduzir os custos em até 90%, dependendo das características da carga de trabalho. Além disso, AWS oferece recursos como o [Babelfish para Aurora PostgreSQL](#), ferramentas como [AWS Schema Conversion Tool \(AWS SCT\)](#) e serviços como [AWS Database Migration Service \(AWS DMS\)](#) para simplificar a migração e a modernização dos bancos de dados do SQL Server no AWS.

Ofertas de banco de dados

A migração do SQL Server no Windows para bancos de dados de código aberto como Amazon Aurora, Amazon RDS para MySQL ou Amazon RDS for PostgreSQL pode oferecer economias de custo significativas sem comprometer o desempenho ou os recursos. Considere o seguinte:

- A mudança da edição SQL Server Enterprise no Amazon EC2 para o Amazon RDS for PostgreSQL ou Amazon RDS for MySQL pode resultar em economias de até 80%.
- A mudança da edição SQL Server Enterprise no Amazon EC2 para a edição compatível com o Amazon Aurora PostgreSQL ou a edição compatível com o Amazon Aurora MySQL pode resultar em economias de até 70%.

Para cargas de trabalho de banco de dados tradicionais, o Amazon RDS for PostgreSQL e o Amazon RDS for MySQL atendem aos requisitos e fornecem uma solução econômica para bancos de dados relacionais. O Aurora adiciona vários recursos de disponibilidade e desempenho antes limitados a fornecedores comerciais caros. Os recursos de resiliência do Aurora têm um custo adicional. No entanto, em comparação com recursos semelhantes de outros fornecedores comerciais, os custos de resiliência do Aurora ainda são mais baratos do que os cobrados por softwares comerciais pelo mesmo tipo de recursos. A arquitetura do Aurora é otimizada para oferecer melhorias significativas no desempenho em comparação com as implantações padrão do MySQL e do PostgreSQL.

Como o Aurora é compatível com bancos de dados PostgreSQL e MySQL de código aberto, há o benefício adicional da portabilidade. Se a melhor opção é o Amazon RDS para PostgreSQL, o Amazon RDS para MySQL ou o Aurora, tudo se resume a entender os requisitos de negócios e mapear os recursos necessários para a melhor opção.

Comparação entre Amazon RDS e Aurora

A tabela a seguir resume as principais diferenças entre o Amazon RDS e o Amazon Aurora.

Categoria	Amazon RDS para PostgreSQL ou Amazon RDS para MySQL	Aurora PostgreSQL ou Aurora MySQL
Performance	Bom desempenho	Desempenho 3x ou melhor
Failover	Normalmente de 60 a 120 segundos*	Normalmente 30 segundos
Escalabilidade	Réplica de até 5 leituras Atraso em segundos	Até 15 réplicas de leitura Atraso em milissegundos
Armazenamento	Até 64 TB	Até 128 TB
Armazenamento HA	Multi-AZ com um ou dois modos de espera, cada um com cópia do banco de dados	6 cópias de dados em 3 zonas de disponibilidade por padrão
Backup	Backups diários de instantâneos e registros	Backup contínuo e assíncrono para o Amazon S3

Categoria	Amazon RDS para PostgreSQL ou Amazon RDS para MySQL	Aurora PostgreSQL ou Aurora MySQL
Inovações com a Aurora	N/D	100 GB Clonagem rápida de banco de dados
	Réplicas de leitura com escalabilidade automática	
	Gerenciamento de planos de consultas	
	Aurora Serverless	
	Réplicas entre regiões com banco de dados global	
	Gerenciamento de cache de cluster**	
	Consulta paralela	
	Fluxos de atividades do banco de dados	

*Transações grandes podem aumentar os tempos de failover

**Disponível no Aurora PostgreSQL

A tabela a seguir mostra o custo mensal estimado dos diferentes serviços de banco de dados abordados nesta seção.

Serviço de banco de dados	Custo em USD por mês*	AWS Pricing Calculator (requer Conta da AWS)
Amazon RDS para SQL Server, edição Enterprise	\$3.750	Estimativa
Amazon RDS for SQL Server Edição Standard	\$2.318	Estimativa
Edição SQL Server Enterprise no Amazon EC2	\$2.835	Estimativa
Edição SQL Server Standard no Amazon EC2	\$1.345	Estimativa
Amazon RDS para PostgreSQL	\$742	Estimativa
Amazon RDS para MySQL	\$712	Estimativa
Aurora PostgreSQL	\$1.032	Estimativa
Aurora MySQL	\$1.031	Estimativa

* O preço do armazenamento está incluído no preço da instância. Os custos são baseados na us-east-1 região. A taxa de transferência e o IOPS são suposições. Os cálculos são para instâncias r6i.2xlarge e r6g.2xlarge.

Recomendações de otimização de custos

As migrações de banco de dados heterogêneas geralmente exigem a conversão do esquema do banco de dados do mecanismo de banco de dados de origem para o de destino e a migração de dados do banco de dados de origem para o de destino. A primeira etapa para a migração é avaliar e converter objetos de código e esquema do SQL Server no mecanismo de banco de dados de destino.

Você pode usar o [AWS Schema Conversion Tool \(AWS SCT\)](#) para avaliar e avaliar a compatibilidade do banco de dados com várias opções de banco de dados de código aberto de destino, como Amazon RDS para MySQL ou Amazon RDS para PostgreSQL, Aurora MySQL e PostgreSQL. Você

também pode usar a ferramenta Babelfish Compass para avaliar a compatibilidade com o Babelfish para Aurora PostgreSQL. Isso torna o Compass AWS SCT e o Compass ferramentas poderosas para entender o trabalho inicial envolvido antes de decidir sobre uma estratégia de migração. Se você decidir continuar, AWS SCT automatiza as alterações necessárias no esquema. A filosofia central por trás do Babelfish Compass é permitir que o banco de dados SQL se mova para o Aurora sem ou com poucas modificações. O Compass avaliará o banco de dados SQL existente para determinar se isso pode ser feito. Dessa forma, o resultado é conhecido antes que qualquer esforço seja gasto na migração de dados do SQL Server para o Aurora.

AWS SCT automatiza a conversão e a migração do esquema e do código do banco de dados para o mecanismo do banco de dados de destino. Você pode usar o Babelfish for Aurora PostgreSQL para migrar seu banco de dados e aplicativo do SQL Server para o Aurora PostgreSQL sem nenhuma ou mínima alteração no esquema. Isso pode acelerar suas migrações.

Depois que o esquema for migrado, você poderá usá-lo AWS DMS para migrar os dados. AWS DMS pode realizar o carregamento total dos dados e replicar as alterações para realizar a migração com o mínimo de tempo de inatividade.

Esta seção explora as seguintes ferramentas com mais detalhes:

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

AWS Schema Conversion Tool

Você pode usar AWS SCT para avaliar seus bancos de dados SQL Server existentes e avaliar a compatibilidade com o Amazon RDS ou o Aurora. Para simplificar o processo de migração, você também pode usar AWS SCT para converter o esquema de um mecanismo de banco de dados para outro em uma migração de banco de dados heterogênea. Você pode usar AWS SCT para avaliar seu aplicativo e converter o código do aplicativo incorporado para aplicativos escritos em C#, C++, Java e outras linguagens. Para obter mais informações, consulte [Convertendo o SQL do aplicativo usando AWS SCT](#) na AWS SCT documentação.

AWS SCT é uma AWS ferramenta gratuita que oferece suporte a várias [fontes](#) de banco de dados. Para usá-lo AWS SCT, você o aponta para o banco de dados de origem e, em seguida, executa uma avaliação. Em seguida, [AWS SCT](#) avalia o esquema e gera o relatório de avaliação. Os relatórios

de avaliação incluem um resumo executivo, complexidade e esforço de migração, mecanismos de banco de dados de destino adequados e recomendações para conversão. Para fazer o download AWS SCT, [consulte Instalação, verificação e atualização AWS SCT](#) na AWS SCT documentação.

A tabela a seguir mostra um exemplo de resumo executivo gerado por AWS SCT para mostrar a complexidade envolvida na alteração do banco de dados para diferentes plataformas de destino.

Plataforma de destino	Alterações automáticas ou mínimas		Ações complexas				
	Objetos de armazenamento	Objetos de código	Ações de conversão	Objetos de armazenamento		Objetos de código	
Amazon RDS para MySQL	60 (98%)	8 (35%)	42	1 (2%)	1	15 (65%)	56
Amazon Aurora Edição Compatível com MySQL	60 (98%)	8 (35%)	42	1 (2%)	1	15 (65%)	56
Amazon RDS para PostgreSQL	60 (98%)	12 (52%)	54	1 (2%)	1	11 (48%)	26
Amazon Aurora Edição Compatível	60 (98%)	12 (52%)	54	1 (2%)	1	11 (48%)	26

I com
PostgreSQL
L

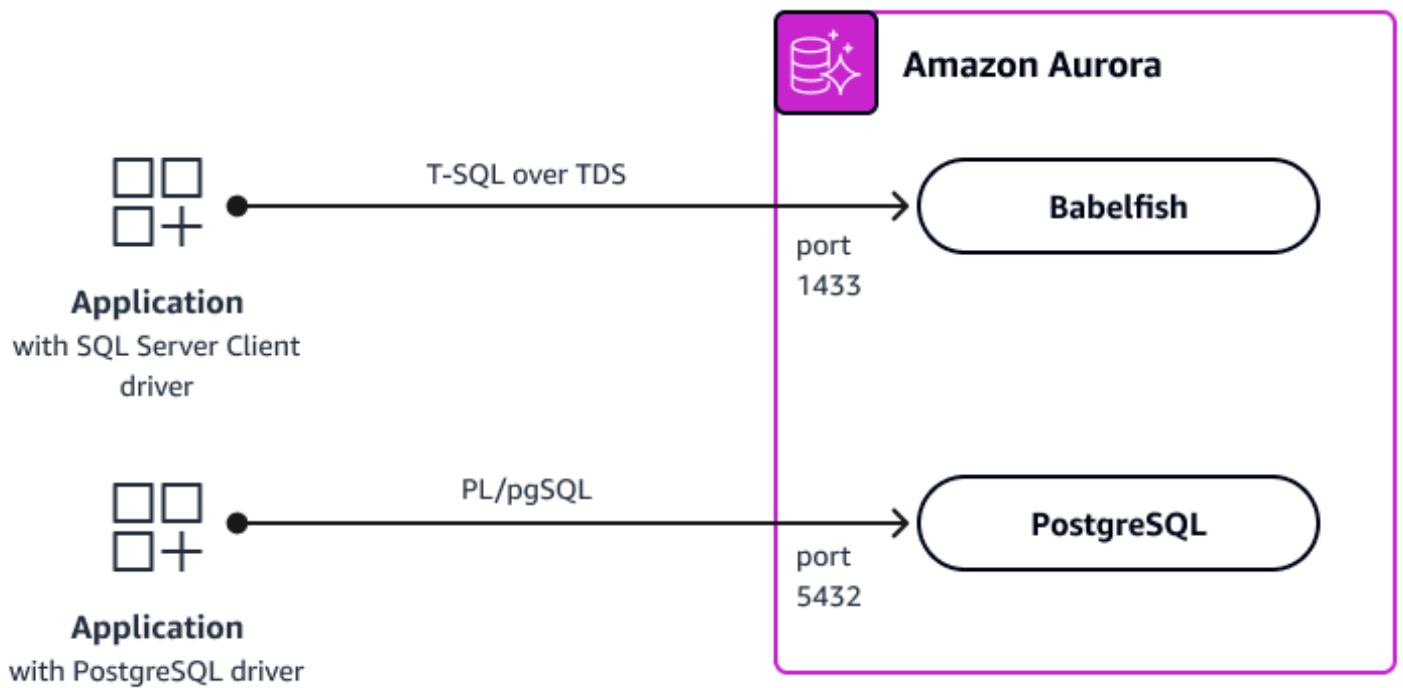
Amazon RDS para MariaDB	60 (98%)	7 (30%)	42	1 (2%)	1	16 (70%)	58
Amazon Redshift	61 (100%)	9 (39%)	124	0 (0%)	0	14 (61%)	25
AWS Glue	0 (0%)	17 (100%)	0	0 (0%)	0	0 (0%)	0
Babelfish	59 (97%)	10 (45%)	20	2 (3%)	2	12 (55%)	30

Um AWS SCT relatório também fornece detalhes sobre os elementos do esquema que não podem ser convertidos automaticamente. Você pode fechar as lacunas AWS SCT de conversão e otimizar os esquemas de destino consultando os manuais de [AWS migração](#). Há muitos manuais de migração de banco de dados para auxiliar em migrações heterogêneas.

Babelfish for Aurora PostgreSQL

O Babelfish for Aurora PostgreSQL estende o Aurora PostgreSQL com a capacidade de aceitar conexões de banco de dados de clientes do SQL Server. O Babelfish permite que aplicativos originalmente criados para o SQL Server funcionem diretamente com o Aurora PostgreSQL, com poucas alterações no código e sem alterar os drivers do banco de dados. O Babelfish torna o Aurora PostgreSQL bilíngue para que o Aurora PostgreSQL possa funcionar com as linguagens T-SQL e PL/pgSQL. O Babelfish minimiza os esforços de migração do SQL Server para o Aurora PostgreSQL. Isso acelera as migrações, minimiza os riscos e reduz significativamente os custos de migração. Você pode continuar usando as pós-migrações do T-SQL, mas também há a [opção de usar ferramentas nativas do PostgreSQL para desenvolvimento](#).

O diagrama a seguir ilustra como um aplicativo usando T-SQL se conecta à porta padrão 1433 no SQL Server e usa o tradutor Babelfish para se comunicar com o banco de dados Aurora PostgreSQL, enquanto um aplicativo usando PL/pgSQL pode se conectar direta e simultaneamente ao banco de dados Aurora PostgreSQL usando a porta padrão 5432 no Aurora PostgreSQL.



O Babelfish não oferece suporte a determinados recursos T-SQL do SQL Server. Por esse motivo, a Amazon fornece ferramentas de avaliação para fazer uma line-by-line análise de suas instruções SQL e determinar se alguma delas não é suportada pelo Babelfish.

Há duas opções para avaliações do Babelfish. AWS SCT pode avaliar a compatibilidade do seu banco de dados SQL Server com o Babelfish. Outra opção é a ferramenta Babelfish Compass, que é uma solução recomendada porque a ferramenta Compass é atualizada de acordo com as novas versões do Babelfish para Aurora PostgreSQL.

Bússola Babelfish

O [Babelfish Compass](#) é uma ferramenta gratuita para download que se alinha à versão mais recente do Babelfish para Aurora PostgreSQL. Em contraste, AWS SCT suportará versões mais recentes do Babelfish depois de algum tempo. O [Babelfish Compass](#) é executado no esquema do banco de dados SQL Server. Você também pode extrair o esquema do banco de dados SQL Server de origem usando ferramentas como o SQL Server Management Studio (SSMS). Em seguida, você pode executar o esquema por meio do Babelfish Compass. Isso gera o relatório detalhando a compatibilidade do esquema do SQL Server com o Babelfish e se alguma alteração é necessária antes da migração. A ferramenta Babelfish Compass também pode automatizar muitas dessas mudanças e, finalmente, acelerar suas migrações.

Depois que a avaliação e as alterações forem concluídas, você poderá migrar o esquema para o Aurora PostgreSQL usando ferramentas nativas do SQL Server, como SSMS ou sqlcmd. Para obter instruções, consulte a postagem [Migrar do SQL Server para o Amazon Aurora usando o Babelfish](#) no blog do banco de dados. AWS

AWS Database Migration Service

Depois que o esquema for migrado, você poderá usar AWS Database Migration Service (AWS DMS) para migrar os dados AWS com o mínimo de tempo de inatividade. AWS DMS não apenas carrega os dados por completo, mas também replica as alterações da origem para o destino enquanto o sistema de origem está funcionando. Depois que os bancos de dados de origem e de destino estiverem sincronizados, a atividade de transição pode ocorrer onde o aplicativo é direcionado para o banco de dados de destino, concluindo a migração. AWS DMS atualmente só executa o carregamento total de dados com o Babelfish para um destino do Aurora PostgreSQL e não replica as alterações. Para obter mais informações, consulte [Usando o Babelfish como alvo AWS Database Migration Service](#) na AWS DMS documentação.

AWS DMS pode fazer migrações homogêneas (no mesmo mecanismo de banco de dados) e heterogêneas (em diferentes mecanismos de banco de dados). AWS DMS oferece suporte a vários mecanismos de banco de dados de origem e destino. Para obter mais informações, consulte a publicação Como [migrar seu banco de dados do SQL Server para o Amazon RDS for SQL Server AWS DMS](#) usando AWS o blog do banco de dados.

Recursos adicionais

- [Adeus Microsoft SQL Server, Olá Babelfish \(Blog de notícias\)](#)AWS
- [Converta esquemas de banco de dados e SQL de aplicativos usando a AWS Schema Conversion Tool CLI](#)AWS (Database Blog)
- [Migre o SQL Server para o Amazon Aurora PostgreSQL usando as melhores práticas e lições aprendidas](#) em campo (blog do banco de dados)AWS
- [Valide objetos de banco de dados após a migração do Microsoft SQL Server para o Amazon RDS for PostgreSQL e o Amazon Aurora PostgreSQL](#) (blog do banco de dados)AWS

Otimize o armazenamento para SQL Server

Visão geral

Esta seção se concentra nas otimizações de custo do armazenamento SSD Amazon Elastic Block Store (Amazon EBS) para SQL Server em cargas de trabalho do EC2.

Você tem uma grande variedade de opções de armazenamento para implantar e executar cargas de trabalho do SQL Server em AWS. A seleção do armazenamento certo deve ser baseada na finalidade, na arquitetura, na durabilidade, no desempenho, na capacidade e no custo. AWS os clientes que executam cargas de trabalho do SQL Server geralmente utilizam uma combinação de armazenamento Amazon EBS, NVMe, Amazon FSx e Amazon Simple Storage Service (Amazon S3).

O Amazon EBS é um armazenamento conectado à rede conectado a instâncias computacionais do EC2 e utilizado para armazenar e processar arquivos gerais do sistema operacional, do aplicativo, do banco de dados e de backup. O armazenamento em unidade de estado sólido (SSD) do Amazon EBS inclui SSD de uso geral (gp2 e gp3) e SSD de IOPS provisionada (io1, io2 e IO2bx). Considere o seguinte:

- Algumas instâncias do EC2, como r5d, têm SSDs NVMe locais fisicamente conectados à instância host. Esses volumes fornecem armazenamento em nível de bloco que é comumente usado para a extensão tempdb ou de buffer pool do SQL Server.
- O Amazon FSx for Windows File Server é um serviço de armazenamento de arquivos totalmente gerenciado, enquanto o Amazon FSx NetApp for ONTAP é um armazenamento compartilhado totalmente gerenciado NetApp baseado no popular sistema de arquivos ONTAP. O Amazon FSx é frequentemente usado para executar cargas de trabalho do SQL Server em uma configuração de instância clusterizada de failover (FCI) do SQL Server de alta disponibilidade. Essa solução hospeda arquivos de dados e log do SQL Server, o que reduz os requisitos de desempenho do EBS nas instâncias do EC2.
- O Amazon S3 é um serviço de armazenamento de objetos que oferece escalabilidade, disponibilidade de dados, segurança e desempenho líderes do setor. Você pode armazenar arquivos de backup nativos do SQL Server, AMIs, snapshots do EBS, logs de aplicativos e muito mais no Amazon S3.

Tipos, desempenho e custo de armazenamento SSD para o Amazon EBS

Os custos de armazenamento SSD do Amazon EBS geralmente aumentam à medida que a durabilidade e o desempenho aumentam. Atualmente, o armazenamento vem em cinco tipos de volume, cada um com suas [próprias métricas de desempenho exclusivas](#). Para obter um resumo dos casos de uso e das características dos volumes baseados em SSD, consulte a tabela na seção Volumes de unidades de [estado sólido \(SSD\) da documentação](#) do Amazon EBS.

Você pode usar CloudWatch a Amazon para monitorar o desempenho do SSD, capturar dados de tendências e definir alarmes quando determinados limites forem atingidos. Se você estiver executando cargas de trabalho do SQL Server em AWS, considere habilitar o [monitoramento detalhado](#) e implantar [métricas CloudWatch personalizadas para capturar métricas](#) detalhadas de desempenho de volume, como latência de disco, IOPS, taxa de transferência, tamanho da fila de disco, capacidade usada versus capacidade livre e muito mais. Você pode usar essas métricas de CloudWatch desempenho para identificar armazenamento subprovisionado e superprovisionado e fornecer pontos de dados históricos para definir com precisão os requisitos de armazenamento.

Os custos de armazenamento SSD para o Amazon EBS também variam com base na capacidade alocada. A tabela abaixo mostra uma comparação dos diferentes tipos de volume. Todos os tipos de volume têm 1 TB de capacidade e configurações de desempenho semelhantes.

Tipo de volume	IOPS máximo (16 KiB I/O)	Taxa de transferência máxima (128 KiB I/O)	Preço por 1 TB	Porcentagem de economia de custos
gp2	3.000	250	\$102,40	
gp3	3.000	250	\$86,92	15%
io1	16.000	500	\$1.168	
io2	16.000	500	\$1.168	
gp3	16.000	500	\$146,92	87%
io2bx	16.000	4.000	\$1.168	
gp3	16.000	1.000	\$181,92	84%

Note

As métricas de desempenho e custo na tabela anterior são por volume, com base em uma [estimativa](#) da AWS Pricing Calculator. Conta da AWS É necessário um para acessar a estimativa no AWS Pricing Calculator.

Os volumes SSD gp3 do Amazon EBS oferecem excelente desempenho a um baixo custo. Você pode economizar até 87% se escolher um volume gp3 em vez de volumes io1 ou io2 para cargas de trabalho que exigem menos de 16.000 IOPS e 500 taxas de transferência. MiBps

Os volumes io2 Block Express (IO2Bx) oferecem maior desempenho em relação aos volumes io2 regulares. Com 16.000 IOPS, os volumes io1 ou io2 são capazes de atingir apenas 500 MiBps taxas de transferência, enquanto os volumes IO2Bx podem ser configurados com até 4.000 taxas de transferência. MiBps Em comparação com os volumes io1 e io2, os volumes IO2Bx fornecem mais de quatro vezes a taxa de transferência entre 16.000 e 64.000 IOPS, exatamente pelo mesmo preço. Volumes io2 regulares podem ser convertidos em volumes IO2Bx anexando-os a instâncias EC2 compatíveis com IO2BX. Para obter uma lista de instâncias do EC2 compatíveis com IO2BX, consulte Volumes [SSD de IOPS provisionados](#) na documentação do Amazon EBS. Antes de implantar um novo armazenamento, você pode usar o [AWS Pricing Calculator](#) para estimar seu custo mensal e entender o impacto no custo com base nas compensações entre durabilidade, desempenho e capacidade.

Otimização geral de custos de SSDs para Amazon EBS

Recomendamos que você avalie o que está armazenando e garanta que está usando o tipo e a classe de armazenamento corretos. Por exemplo, o Amazon S3 oferece um ótimo preço, políticas de ciclo de vida integradas e opções de replicação ideais para backups do SQL Server. O SQL Server 2022 tem a capacidade de fazer backup diretamente no Amazon S3, enquanto as versões anteriores do SQL Server dependem de backups locais nativos. Se você estiver executando versões mais antigas do SQL Server, considere fazer backup nos volumes HDD do Amazon EBS e depois copiar o backup para o Amazon S3. Essa solução pode economizar 53% em vez de usar volumes gp3 para backups.

A tabela a seguir mostra a diferença de preço de 1 TB de armazenamento no Amazon EBS gp3, Amazon EBS HDD st1 e Amazon S3.

Tipo de armazenamento	Capacity	Preço pm
EBS gp3 500 MiBps	1 TB	\$96,92
EBS st1 burst 500 MiBps		\$46,08
S3 Standard		\$23,55
S3 Standard (acesso infrequente)		\$12,80
S3 Glacier Deep Archive		\$1,03

Note

As métricas de custo na tabela anterior são baseadas em uma [estimativa](#) na AWS Pricing Calculator. Conta da AWS É necessário um para acessar a estimativa no AWS Pricing Calculator.

Recomendamos que você considere o seguinte:

- Habilite o monitoramento detalhado e implante métricas CloudWatch personalizadas para capturar com precisão seus requisitos de desempenho de armazenamento.
- Atualize o armazenamento Amazon EBS de gp2 para gp3 para reduzir custos, aumentar a flexibilidade e melhorar o desempenho.
- Atualize o armazenamento do Amazon EBS de io1 para io2 para maior durabilidade e flexibilidade de desempenho.
- Use IO2bx em vez de io1 ou io2 quando possível para aumentar a durabilidade e o desempenho.
- Considere uma mix-and-match abordagem ao escolher o armazenamento para ajudar a reduzir os requisitos de capacidade e o custo de volumes de alto desempenho. Por exemplo, você pode usar volumes gp3 de baixo custo para seu volume raiz (sistema operacional), instalação do SQL Server, bancos de dados do sistema (excluindo tempdb) e bancos de dados de usuários de baixo desempenho. Isso pode ajudar a reduzir a capacidade e o custo de um volume io2, que pode ser dedicado a bancos de dados de usuários de alto desempenho.

- Se você estiver hospedando bancos de dados do SQL Server em AWS, recomendamos que você use vários arquivos de dados do SQL Server por banco de dados. Isso permite a oportunidade de distribuir cargas de trabalho de leitura/gravação em vários volumes, reduzindo os requisitos de desempenho e capacidade por volume e, conseqüentemente, reduzindo os custos.
- Mesmo que as cargas de trabalho de produção exijam armazenamento de alto desempenho, como io1 ou IO2/io2Bx, considere os volumes gp3 para cargas de trabalho que não sejam de produção para ajudar a reduzir custos.
- Acompanhe e defina a tendência da utilização do armazenamento ao longo do tempo para identificar facilmente picos de uso e custos inesperados.
- Use [AWS Compute Optimizer](#) para obter recomendações sobre como aumentar ou diminuir os volumes do EBS com base na utilização real.
- Use a elasticidade de AWS para ajustar as necessidades de desempenho e capacidade de seus volumes SSD para o Amazon EBS. Ao contrário dos ambientes locais, você não precisa provisionar em excesso o desempenho e a capacidade de armazenamento para futuras cargas de trabalho. Você pode migrar suas cargas de trabalho existentes do SQL Server AWS e ajustar o desempenho ou a capacidade conforme necessário, mantendo seus bancos de dados on-line.

Recursos adicionais

- [Tipos de volume do Amazon EBS](#) (documentação do Amazon EBS)
- [Amazon Elastic Block Store \(Amazon EBS\) \(documentação do Amazon EBS\)](#)
- [Volumes SSD de IOPS provisionados \(documentação do Amazon EBS\)](#)
- [Volumes de armazenamento de instâncias SSD](#) (documentação do Amazon EC2)
- [CloudWatch Métricas da Amazon para Amazon EBS](#) (documentação do Amazon EBS)
- [Especificações para instâncias otimizadas para armazenamento do Amazon EC2 \(documentação do Amazon EC2\)](#)
- [Proteja suas cargas de trabalho do SQL Server usando o NetApp SnapCenter Amazon FSx NetApp for ONTAP AWS](#) (Storage Blog)
- [Perguntas frequentes sobre o Amazon EC2 \(página AWS do produto\)](#)

Otimize o licenciamento do SQL Server usando o Compute Optimizer

Orientação sobre como otimizar licenças para o SQL Server usando o AWS Compute Optimizer

Visão geral

[AWS Compute Optimizer](#) pode recomendar oportunidades de otimização de licenciamento para cargas de trabalho do Microsoft SQL Server no Amazon Elastic Compute Cloud (Amazon EC2). O Compute Optimizer pode fornecer recomendações automatizadas para reduzir os custos de licenciamento. As recomendações do Compute Optimizer estão listadas ao lado de cada uma de suas instâncias do EC2 com licenças do Microsoft SQL Server. As informações fornecidas incluem oportunidades de economia recomendadas, preços sob demanda de instâncias EC2 e preços por hora para trazer sua própria licença (BYOL). Essas informações podem ajudá-lo a decidir se você deve fazer o downgrade de sua edição de licença.

O Compute Optimizer descobre automaticamente suas instâncias do SQL Server no Amazon EC2 por tipo de carga de trabalho inferido. Para visualizar as recomendações de licenciamento, você pode selecionar as instâncias do SQL Server no Compute Optimizer e, em seguida, autenticar-se com o [CloudWatch Amazon Application Insights](#) usando suas credenciais de banco de dados somente para leitura. O Compute Optimizer analisa se você está usando algum recurso do SQL Server Enterprise Edition. Se nenhum recurso da edição Enterprise estiver sendo usado, o Compute Optimizer recomenda que você faça o downgrade para a edição Standard para reduzir os custos de licenciamento.

Você também pode usar o Compute Optimizer para fazer recomendações de dimensionamento para suas instâncias do Amazon EC2 que executam cargas de trabalho do SQL Server. Para obter mais informações, consulte [Otimizar o dimensionamento do SQL Server usando o Compute Optimizer neste](#) guia.

Recomendações de otimização de custos

As recomendações de licença no Compute Optimizer podem ajudá-lo a avaliar os recursos que você está usando no Microsoft SQL Server e escolher a edição mais econômica para suas cargas de trabalho. A edição SQL Server Enterprise é significativamente mais cara do que a edição Standard. Para obter mais informações, consulte [Compare as edições do SQL Server](#) neste guia e consulte os [preços do SQL Server 2022](#) no site da Microsoft. Investir tempo para configurar o Compute Optimizer

para avaliar sua frota de SQL Server e fornecer recomendações pode reduzir drasticamente seus custos de licenciamento.

A página de detalhes da licença fornece as seguintes informações:

- Use a tabela para comparar suas configurações de licença atuais (como edição, modelo e número de núcleos de instância) com as recomendações do Compute Optimizer.
- Use os gráficos de utilização para analisar o número de recursos da edição Enterprise que foram usados durante o período de análise.

Para obter mais informações, consulte [Visualização dos detalhes de uma recomendação de licença de software comercial](#) na documentação do Compute Optimizer.

Configurar o Compute Optimizer

O Compute Optimizer analisa as licenças de software comercial usando a métrica.

`mssql_enterprise_features_used` Para obter mais informações sobre essa métrica, consulte [Métricas para licenças de software comercial](#).

1. Verifique se você tem as permissões apropriadas para optar pelo Compute Optimizer. Para obter mais informações, consulte as informações a seguir.
 - [Política de adesão ao Compute Optimizer](#)
 - [Políticas para conceder acesso ao Compute Optimizer para uso autônomo Contas da AWS](#)
 - [Políticas para conceder acesso ao Compute Optimizer para uma conta de gerenciamento de uma organização](#)
2. Anexe as funções e a política de instância necessárias para o CloudWatch Application Insights. Para obter instruções, consulte [Políticas para habilitar recomendações de licenças de software comercial](#).
3. Habilite o CloudWatch Application Insights usando suas credenciais do banco de dados Microsoft SQL Server. Para obter instruções, consulte [Comece a usar o Amazon CloudWatch Application Insights](#).

Note

Para gerar recomendações para licenças de software comercial, são necessárias pelo menos 30 horas consecutivas de dados CloudWatch métricos. Para obter mais informações, consulte os [requisitos CloudWatch métricos](#).

4. Use a consulta SQL a seguir para configurar o acesso com privilégios mínimos para o Application Insights. CloudWatch

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

Isso habilita um novo serviço, o PrometheusSqlExporter SQL.

5. Na conta de gerenciamento de destino Conta da AWS ou da organização, opte pelo Compute Optimizer. Para obter instruções, consulte [Como ativar sua conta](#).

Note

Depois que você aceita, as descobertas e recomendações de otimização podem levar até 24 horas para serem geradas.

6. No console do [Compute Optimizer](#), escolha Licenças no painel de navegação.
7. Na coluna Descobertas, pesquise todas as instâncias que tenham a descoberta de métricas insuficientes. O Compute Optimizer retorna essa descoberta se detectar CloudWatch que o Application Insights não está ativado ou tem permissões insuficientes. Para obter mais informações, consulte [Encontrando motivos](#). Faça o seguinte para resolver essas descobertas:
 - a. Escolha a instância.
 - b. Adicione um segredo.
 - c. Confirme se a função e a política da instância estão anexadas.
 - d. Escolha Ativar recomendações de licença.
8. Na coluna Descobertas, pesquise todas as instâncias que tenham a descoberta Não otimizada. O Compute Optimizer retorna essa descoberta se detectar que sua infraestrutura do Amazon EC2 não está usando nenhum dos recursos de licença do Microsoft SQL Server pelos quais você está pagando. Para obter mais informações, consulte [Encontrando motivos](#). Faça o seguinte para resolver essas descobertas:

- a. Escolha a instância.
- b. Compare a edição da licença atual com a edição recomendada.
- c. Analise o gráfico atual de utilização da licença.
- d. Se você quiser fazer o downgrade da licença, escolha Implementar recomendação.
- e. Analise os requisitos e siga as instruções para fazer o downgrade da licença. Se você quiser automatizar o processo, consulte Fazer o [downgrade da edição Enterprise do SQL Server usando o AWS Systems Manager Documento para reduzir custos](#) (AWS Blog).

Recursos adicionais

- [Reduza os custos de licenciamento do Microsoft SQL Server com AWS Compute Optimizer](#) (AWS Blog)
- [O que AWS Compute Optimizer é](#) (AWS documentação)
- [Visualizando recomendações de licenças de software comercial](#) (AWS documentação)
- Faça o [downgrade de sua edição do Microsoft SQL Server](#) (AWS documentação)
- [Microsoft SQL Server ativado AWS](#) (AWS)
- [Licenciamento da Microsoft ativado AWS](#) (AWS)
- [Preços do Microsoft SQL Server 2019](#) (Microsoft)
- [Preços do Microsoft SQL Server 2022](#) (Microsoft)

Otimize o dimensionamento do SQL Server usando o Compute Optimizer

Visão geral

[AWS Compute Optimizer](#) ajuda administradores de banco de dados (DBAs) a descobrir as cargas de trabalho do Microsoft SQL Server na Amazon Elastic Compute Cloud (Amazon EC2) e a ajustar as instâncias EC2 do tamanho certo para reduzir os custos de licença em até 25%. O recurso de [tipo de carga de trabalho inferido](#) no Compute Optimizer usa aprendizado de máquina (ML) e detecta automaticamente os aplicativos que podem estar sendo executados em seus recursos. AWS O Compute Optimizer inclui suporte para o SQL Server como um tipo de carga de trabalho inferida. Ao usar o recurso de tipo de carga de trabalho inferido, você pode identificar oportunidades de redução

de custos com base na carga de trabalho específica em execução em suas instâncias do Amazon EC2.

Com esse recurso, você pode categorizar oportunidades de economia de custos de acordo com os tipos de carga de trabalho inferidos compatíveis, como o SQL Server. O Compute Optimizer pode descobrir automaticamente instâncias do SQL Server EC2 que estão superprovisionadas. Você pode mudar para o console do EC2 para reduzir o tamanho da instância, o que ajuda a reduzir os custos de licenciamento e infraestrutura.

Você também pode usar o Compute Optimizer para fazer recomendações de licenciamento do SQL Server. Para obter mais informações, consulte [Otimizar o licenciamento do SQL Server usando o Compute Optimizer neste](#) guia.

Configurar o Compute Optimizer

Para obter instruções sobre como usar o Compute Optimizer com cargas de trabalho inferidas do SQL Server, [consulte Otimizando o desempenho e reduzindo os custos de licenciamento: aproveitando AWS Compute Optimizer as instâncias do SQL Server do Amazon EC2](#) (Blog).AWS Você pode optar por contas autônomas, contas que são membros de uma organização e contas de gerenciamento de uma organização. Para contas independentes e de membros, a ativação ativa o Compute Optimizer somente para essa conta. Para uma conta de gerenciamento da organização, você pode escolher se deseja ativar o Compute Optimizer somente nessa conta ou em todas as contas membros da organização.

O processo de aceitação do Compute Optimizer cria automaticamente AWS Identity and Access Management uma função vinculada ao serviço (IAM). Para obter mais informações, consulte [Usando funções vinculadas a serviço para AWS Compute Optimizer](#).

O Compute Optimizer analisa seus recursos com base nas métricas da CloudWatch Amazon, como CPU, E/S, rede e uso do Amazon Elastic Block Store (Amazon EBS). Para gerar recomendações, são necessárias pelo menos 30 horas consecutivas de dados CloudWatch métricos nos últimos 14 dias. Se você ativar o recurso aprimorado de métricas de infraestrutura, ele estenderá as métricas de utilização para 93 dias. Para obter mais informações, consulte [os requisitos CloudWatch métricos e as métricas de infraestrutura aprimorada](#) na documentação do Compute Optimizer.

O Compute Optimizer fornece opções e as economias associadas a cada opção, com base na vCPU, memória, armazenamento, rede, risco e esforço de migração. Você pode usar o painel de CloudWatch métricas para analisar os dados que estão sendo usados para fazer a recomendação. Com esses dados, você pode dimensionar corretamente suas instâncias do EC2 que estão

executando cargas de trabalho do SQL Server. Para obter mais informações sobre como alterar seu tipo de instância, consulte [Alterar o tipo de instância](#) na documentação do Amazon EC2.

Recursos adicionais

- [AWS Compute Optimizer identifica e filtra cargas de trabalho do Microsoft SQL Server](#) (AWS)
- [Otimizando o desempenho e reduzindo os custos de licenciamento: aproveitando AWS Compute Optimizer as instâncias do SQL Server do Amazon EC2 \(Blog\)](#) AWS
- [O que AWS Compute Optimizer é](#) (AWS documentação)
- [Visualizando recomendações de instâncias do EC2](#) (AWS documentação)

Análise Trusted Advisor as recomendações para cargas de trabalho do SQL Server

Visão geral

[AWS Trusted Advisor](#) fornece recomendações que ajudam você a seguir as AWS melhores práticas. Ao analisar seu uso, configuração e gastos, Trusted Advisor fornece recomendações práticas para reduzir seus custos, melhorar a disponibilidade e o desempenho do sistema ou ajudar a fechar lacunas de segurança. Esta seção se concentra nas Trusted Advisor verificações que podem ajudá-lo a reduzir os custos de operação das cargas de trabalho do SQL Server no Nuvem AWS.

Recomendações de otimização de custos

Trusted Advisor fornece recomendações que ajudam você a otimizar suas cargas de trabalho do SQL Server no Amazon Elastic Compute Cloud (Amazon EC2). As verificações inspecionam suas cargas de trabalho do SQL Server e listam automaticamente as instâncias que precisam de otimização. A operacionalização Trusted Advisor das recomendações pode reduzir custos e melhorar a postura de segurança da sua organização.

A seguir estão as Trusted Advisor verificações que se concentram no Microsoft SQL Server:

- [Instâncias do Amazon EC2 provisionadas em excesso para o Microsoft SQL Server](#) — Essa verificação analisa suas instâncias do Amazon EC2 que estão executando o SQL Server e alerta você se uma instância exceder o limite de vCPU do software SQL Server. Por exemplo, uma instância com a edição SQL Server Standard pode usar até 48 vCPUs. Uma instância com o SQL Server Web pode usar até 32 vCPUs.

Edição	vCPU mínima.	vCPU máx.
Web	4	32
Padrão	4	48
Enterprise	4	Limites do sistema operacional

- [Consolidação de instâncias do Amazon EC2 para Microsoft SQL Server](#) — Essa verificação analisa suas instâncias do Amazon EC2 e alerta se sua instância tiver menos do que o número mínimo de licenças do SQL Server. Você pode consolidar instâncias menores do SQL Server para ajudar a reduzir custos. Se você tiver muitas instâncias pequenas do SQL Server, incluindo licenças, considere a consolidação. De acordo com o [guia de licenciamento do Microsoft SQL Server 2019](#), o SQL Server exige um mínimo de 4 licenças de vCPU por instância. Se você consolidar esses bancos de dados, poderá economizar nos custos de licenciamento. Você pode tomar sua decisão com base no número de bancos de dados na instância, no tamanho máximo do banco de dados e no tamanho total dos bancos de dados. A consolidação é compatível com as edições Web, Standard e Enterprise do SQL Server. Para obter mais informações, consulte [Consolidação de bancos de dados do SQL Server](#) (postagem no blog da Microsoft).

AWS não recomenda colocar grandes bancos de dados de produção em apenas um servidor. No entanto, você pode consolidar os menores usados em ambientes que não sejam de produção, como desenvolvimento, teste e preparação. Isso depende do uso atual do SQL Server; se você tiver bancos de dados de baixo uso, poderá consolidar em um servidor.

Configurar o Trusted Advisor

Faça o seguinte para avaliar as verificações focadas no SQL Server Trusted Advisor.

1. Faça login no AWS Management Console.
2. Abra o [console de AWS Trusted Advisor](#).
3. No painel de navegação, em Recomendações, escolha Otimização de custos.
4. Na lista de verificações de otimização de custos, analise o status da consolidação de instâncias do Amazon EC2 para o Microsoft SQL Server e das instâncias do Amazon EC2 superprovisionadas para verificações do Microsoft SQL Server.

- Símbolos de verificação verdes indicam que suas instâncias do Amazon EC2 estão configuradas de forma ideal.
 - Os símbolos de alerta laranja indicam que há oportunidades de melhoria.
5. Escolha um cheque para ver seus detalhes e recomendações.
 6. Siga as instruções fornecidas pela verificação para otimizar suas instâncias do Amazon EC2 que estão executando cargas de trabalho do SQL Server.
 7. Monitore suas instâncias regularmente e atualize as verificações periodicamente.

Recursos adicionais

- [Trusted Advisor referência de verificação](#) (AWS documentação)
- [Microsoft SQL Server ativado AWS](#) (AWS)
- [Licenciamento da Microsoft ativado AWS](#) (AWS)
- [Preços do SQL Server 2019](#) (Microsoft)
- [AWS Launch Wizard para SQL Server](#) (AWS documentação)

Contêineres

A modernização é uma jornada transformacional que oferece muitas opções, incluindo a decomposição de monólitos em microsserviços, a rearquitetura de aplicativos para serem orientados por eventos usando funções sem servidor (AWS Lambda) e a reutilização de bancos de dados do SQL Server para o Amazon Aurora ou bancos de dados gerenciados específicos. Os caminhos de modernização para reformular aplicativos .NET Framework para contêineres Linux e Windows exigem menos esforço do que outras opções de modernização. Os contêineres oferecem os seguintes benefícios:

- **Acelere a inovação** — migrar para contêineres facilita a automatização de estágios do ciclo de vida de desenvolvimento, que incluem a criação, o teste e a implantação de aplicativos. Ao automatizar esses processos, as equipes de desenvolvimento e operações têm mais tempo para se concentrar na inovação.
- **Reduzir o custo total de propriedade (TCO)** — migrar para contêineres também pode reduzir sua dependência das ferramentas de gerenciamento de licenças e proteção de terminais. Como os contêineres são unidades de computação efêmeras, você pode automatizar e simplificar as tarefas de gerenciamento, como aplicação de patches, escalabilidade, backup e restauração. Isso reduz o TCO da administração e operação de cargas de trabalho baseadas em contêineres. Por fim, os contêineres são mais eficientes em comparação às máquinas virtuais porque você pode usar contêineres para maximizar o posicionamento de seus aplicativos, fornecendo melhor isolamento. Isso aumenta a utilização dos recursos de infraestrutura do seu aplicativo.
- **Melhore a utilização dos recursos** — Os contêineres são mais eficientes em comparação com as máquinas virtuais porque você pode usar contêineres para maximizar o posicionamento de seus aplicativos. Isso aumenta a utilização dos recursos de infraestrutura do seu aplicativo, fornecendo melhor isolamento.
- **Elimine a lacuna de habilidades** — AWS oferece dias de imersão para aprimorar suas equipes de desenvolvimento em tecnologia e DevOps práticas de contêineres.

Esta seção abrange os seguintes tópicos:

- [Mova aplicativos do Windows para contêineres](#)
- [Otimize os custos AWS Fargate das tarefas no Amazon ECS](#)
- [Obtenha visibilidade dos seus custos do Amazon EKS](#)
- [Reorganize os aplicativos do Windows com o App2Container](#)

[Para obter informações sobre licenciamento, consulte a seção Licenciamento da Amazon Web Services e da Microsoft: Perguntas frequentes ou envie suas perguntas por e-mail para \[microsoft@amazon.com\]\(mailto:microsoft@amazon.com\).](#)

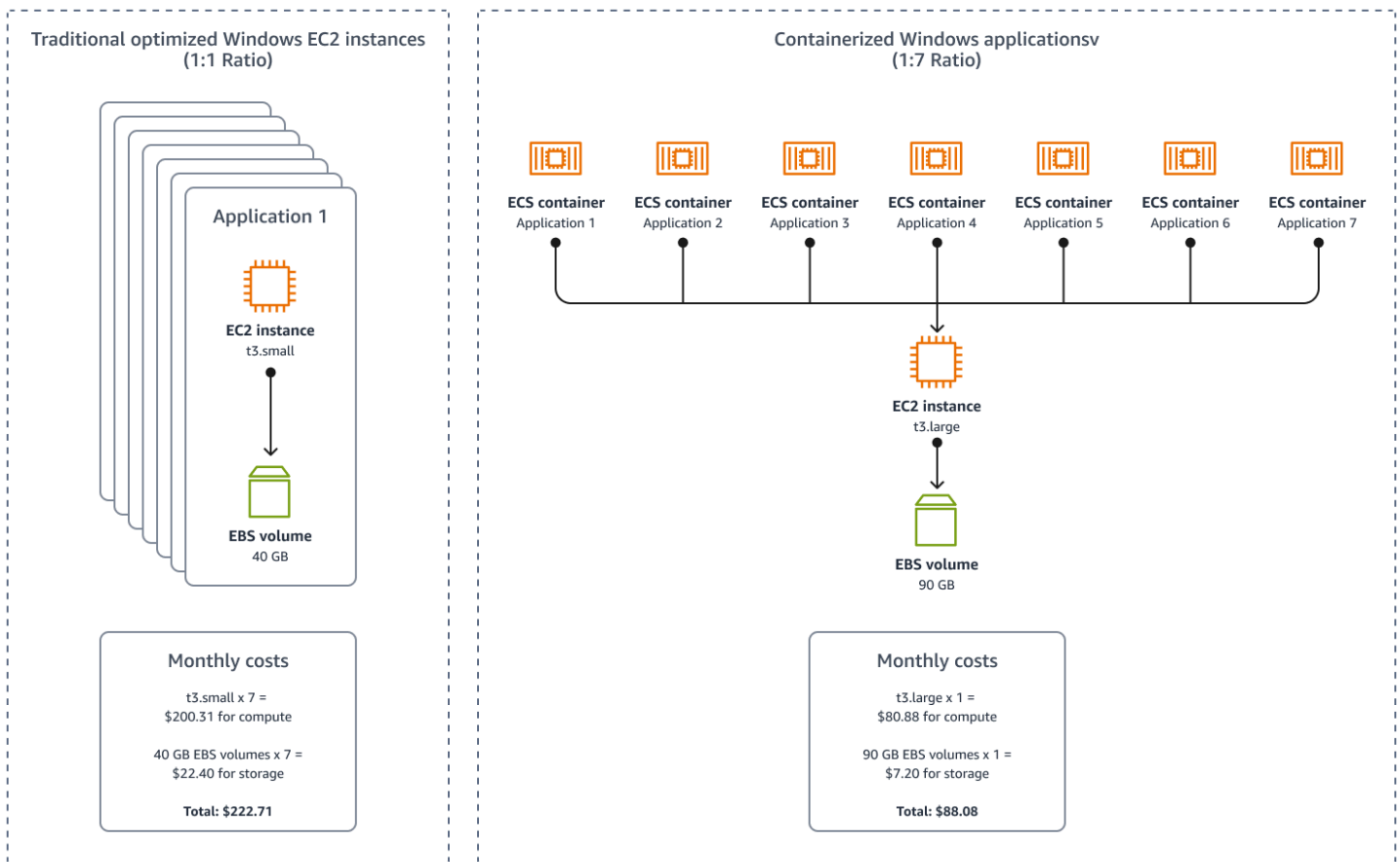
Mova aplicativos do Windows para contêineres

Visão geral

De acordo com a [Pesquisa Anual da CNCF de 2021](#), 96% das organizações estão usando ou avaliando contêineres para modernizar sua infraestrutura. Isso ocorre porque os contêineres podem ajudar sua organização a reduzir riscos, aumentar a eficiência e a velocidade operacionais e possibilitar a agilidade. Você também pode usar contêineres para reduzir o custo de execução de seus aplicativos. Esta seção oferece recomendações para a execução econômica de contêineres em vários serviços de AWS contêineres, incluindo [Amazon Elastic Container Service \(Amazon ECS\)](#), [Amazon Elastic Kubernetes Service \(Amazon EKS\)](#) e [AWS Fargate](#)

Benefícios de custo

[O infográfico a seguir mostra a economia de custos que uma empresa pode obter consolidando seus aplicativos ASP.NET Framework em instâncias do Amazon Elastic Compute Cloud \(Amazon EC2\) com base AWS em uma recomendação de Avaliação de Otimização e Licenciamento \(OLA\).AWS](#) O infográfico a seguir mostra quais economias adicionais podem ser obtidas movendo aplicativos para um contêiner do Windows.



O AWS OLA recomendou que a empresa fizesse uma mudança para instâncias t3.small individuais. A empresa poderia obter essas economias executando sete aplicativos ASP.NET em servidores locais, conforme mostra a análise de utilização de desempenho a seguir.

Server name	Storage	Operating system	On-premises CPU AVG utilization	On-premises CPU peak utilization	On-premises RAM (GB)	On-premises RAM AVG utilization (GB)	On-premises RAM peak utilization (GB)	Instance size	vCPU	RAM (GB)
1 AppServer01	60	Windows Server 2012	7.00%	17.00%	8	13.50%	17.10%	t3.small	2	2
2 AppServer02	39	Windows Server 2012	20.07%	22.00%	16	7.50%	12.40%	t3.small	2	2
3 AppServer03	39	Windows Server 2012	24.00%	25.50%	16	8.80%	11.90%	t3.small	2	2
4 AppServer04	4	Windows Server 2012	21.40%	24.00%	16	7.80%	10.70%	t3.small	2	2
5 AppServer05	40	Windows Server 2012	21.30%	23.00%	16	8.20%	12.00%	t3.small	2	2
6 AppServer06	39	Windows Server 2012	21.50%	23.50%	16	7.90%	10.90%	t3.small	2	2
7 AppServer07	39	Windows Server 2012	21.60%	22.90%	16	8.40%	11.50%	t3.small	2	2

Uma análise mais aprofundada revelou que a empresa poderia economizar ainda mais em custos executando suas cargas de trabalho em contêineres. Os contêineres reduzem a sobrecarga do sistema operacional sobre os recursos do sistema, como CPU, RAM e uso de disco (explicado na próxima seção). Nesse cenário, a empresa poderia consolidar todos os sete aplicativos em uma instância t3.large e ainda ter 3 GB de RAM de sobra. A migração para contêineres pode

ajudar a empresa a obter uma economia média de 64% em computação e armazenamento usando contêineres em vez do Amazon EC2.

Recomendações de otimização de custos

A seção a seguir oferece recomendações para otimizar custos por meio da consolidação de aplicativos e do uso de contêineres.

Reduza sua presença no Windows no Amazon EC2

Os contêineres do Windows podem reduzir sua presença do Windows no Amazon EC2, permitindo que você consolide mais aplicativos em menos instâncias do EC2. Por exemplo, suponha que você tenha 500 aplicativos ASP.NET. Se você estiver executando um núcleo por aplicativo para Windows no Amazon EC2, isso equivale a 500 instâncias do Windows (t3.small). Se você presumir uma proporção de 1:7 (que pode aumentar significativamente dependendo do tipo/tamanho da instância EC2) para usar contêineres do Windows (com t3.large), precisará apenas de aproximadamente 71 instâncias do Windows. Isso representa uma redução de 85,8% em sua presença no Windows no Amazon EC2.

Reduza os custos de licenciamento do Windows

Se você licenciar uma instância do Windows, não precisará licenciar contêineres em execução nessa instância. Como resultado, consolidar seus aplicativos ASP.NET usando contêineres do Windows pode reduzir significativamente seus custos de licenciamento do Windows.

Reduza seu espaço de armazenamento

Toda vez que você executa uma nova instância do EC2, você cria e paga por um novo volume do Amazon Elastic Block Store (Amazon EBS) para abrigar o sistema operacional. À medida que isso aumenta, o custo aumenta com ele. Se você usa contêineres, pode reduzir os custos de armazenamento porque todos os contêineres compartilham o mesmo sistema operacional básico. Além disso, os contêineres usam o conceito de camadas para reutilizar partes imutáveis de uma imagem de contêiner para todos os contêineres em execução com base nessa imagem. No cenário de exemplo anterior, todos os contêineres estão executando o .NET Framework e, portanto, todos compartilham a camada intermediária e imutável da estrutura ASP.NET.

Migre end-of-support servidores para contêineres

O suporte para Windows Server 2012 e Windows Server 2012 R2 terminou em 10 de outubro de 2023. Você pode migrar seus aplicativos em execução no Windows Server 2012 ou em versões

anteriores, colocando-os em contêineres para serem executados em novos sistemas operacionais. Dessa forma, você evita executar seus aplicativos em sistemas operacionais não compatíveis e, ao mesmo tempo, aproveita a economia, o risco reduzido, a eficiência operacional, a velocidade e a agilidade que os contêineres oferecem.

Uma ressalva a ser considerada com essa abordagem é se seu aplicativo requer APIs específicas relacionadas à versão do sistema operacional atualmente em uso (com Interop, por exemplo). Nesse caso, você deve testar a migração do aplicativo para uma versão mais recente do Windows. Os contêineres do Windows alinham sua imagem de contêiner base (por exemplo, Windows Server 2019) com o sistema operacional do host do contêiner (por exemplo, Windows Server 2019). Testar e migrar para contêineres pode facilitar as atualizações do sistema operacional no futuro, alterando a imagem base em seu Dockerfile e implantando em um novo conjunto de hosts executando a versão mais recente do Windows.

Remova ferramentas e licenças de gerenciamento de terceiros

O gerenciamento de sua frota de servidores requer o uso de várias ferramentas de operação de sistema de terceiros para gerenciamento de patches e configuração. Isso pode tornar o gerenciamento da infraestrutura complexo e, muitas vezes, você incorre em custos de licenciamento de terceiros. Se você usa contêineres AWS, não precisa gerenciar nada no lado do sistema operacional. O tempo de execução do contêiner gerencia os contêineres. Isso significa que o hospedeiro subjacente é efêmero e pode ser facilmente substituído. Você pode executar seus contêineres sem a necessidade de gerenciar diretamente o host do contêiner. Além disso, você pode usar ferramentas gratuitas, como AWS Systems Manager Session Manager acessar facilmente o host e solucionar problemas.

Melhore o controle e a portabilidade

Os contêineres oferecem um controle mais granular sobre os recursos do servidor, como CPU e RAM, do que sobre as instâncias do EC2. Para instâncias do EC2, você pode controlar a CPU e a RAM selecionando uma família de instâncias, tipo de instância e [opções de CPU](#). No entanto, com contêineres, você pode definir exatamente quanta CPU ou RAM você deseja alocar para um contêiner nas suas definições de tarefas do ECS ou para [pods no Amazon EKS](#). Na verdade, recomendamos [especificar a CPU e a memória em nível de contêiner](#) para contêineres do Windows. Esse nível de granularidade traz benefícios de custo. Considere o seguinte exemplo de código:

```
json
{
```

```
"taskDefinitionArn": "arn:aws:ecs:us-east-1:123456789012:task-definition/demo-
service:1",
  "containerDefinitions": [
    {
      "name": "demo-service",
      "image": "mcr.microsoft.com/dotnet/framework/samples:aspnetapp-
windowsservercore-ltsc2019",
      "cpu": 512,
      "memory": 512,
      "links": [],
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 0,
          "protocol": "tcp"
        }
      ]
    }
  ],
```

Acelere a inovação

A mudança para contêineres facilita a automatização dos estágios do ciclo de vida de desenvolvimento, que incluem a criação, o teste e a implantação de aplicativos. Se você automatizar esses processos, dará às suas equipes de desenvolvimento e operações mais tempo para se concentrarem na inovação.

Reduza o TCO

A mudança para contêineres geralmente reduz a dependência das ferramentas de gerenciamento de licenças e proteção de terminais. Como os contêineres são unidades de computação efêmeras, você pode automatizar e simplificar as tarefas de gerenciamento, como aplicação de patches, escalabilidade, backup e restauração. Isso pode reduzir o TCO da administração e operação de cargas de trabalho baseadas em contêineres. Os contêineres são mais eficientes em comparação com as máquinas virtuais porque permitem maximizar o posicionamento dos aplicativos para que você possa aumentar a utilização dos recursos de infraestrutura do aplicativo.

Feche a lacuna de habilidades

AWS oferece programas e dias de imersão para aprimorar as equipes de desenvolvimento de clientes em contêineres e DevOps tecnologia. Isso inclui consultoria e capacitação práticas.

Refatore para o.NET 5+ e use contêineres Linux

Embora você possa reduzir custos transferindo seus aplicativos.NET Framework para contêineres, você pode obter ainda mais economia de custos ao refatorar aplicativos.NET legados para alternativas nativas da nuvem. AWS

Remova os custos de licenciamento

Refatorar seu aplicativo do.NET Framework no Windows para o.NET Core no Linux resulta em uma economia de aproximadamente 45%.

Acesse os aprimoramentos mais recentes

Refatorar seus aplicativos do.NET Framework no Windows para.NET Core no Linux dá acesso aos aprimoramentos mais recentes, como o Graviton2. O Graviton2 oferece um preço de desempenho 40% melhor do que instâncias comparáveis.

Melhore a segurança e o desempenho

Refatorar seu aplicativo do.NET Framework no Windows para o.NET Core em contêineres Linux traz melhorias na segurança e no desempenho. Isso ocorre porque você obtém os patches de segurança mais recentes, se beneficia do isolamento de contêineres e tem acesso a novos recursos.

Use contêineres do Windows em vez de executar vários aplicativos em uma instância do IIS

Considere as seguintes vantagens de usar contêineres do Windows em vez de executar vários aplicativos em uma instância EC2 do Windows com o Internet Information Services (IIS):

- **Segurança** — Os contêineres fornecem um nível de segurança pronto para uso que não é alcançado por meio do isolamento no nível do IIS. Se um site ou aplicativo do IIS for comprometido, todos os outros sites hospedados ficarão expostos e vulneráveis. O escape de contêineres é raro e uma vulnerabilidade mais difícil de explorar do que obter o controle de um servidor por meio de uma vulnerabilidade na web.
- **Flexibilidade** — A capacidade de executar contêineres isoladamente no processo e ter sua própria instância permite opções de rede mais granulares. Os contêineres também oferecem métodos complexos de distribuição em muitas instâncias do EC2. Você não obtém esses benefícios ao consolidar aplicativos em uma única instância do IIS.

- Sobrecarga de gerenciamento — A Indicação de Nome de Servidor (SNI) cria uma sobrecarga que requer gerenciamento e automação. Além disso, você precisa lidar com operações típicas de gerenciamento do sistema operacional, como aplicação de patches, solução de problemas de BSOD (se o dimensionamento automático não estiver em vigor), proteção de terminais e assim por diante. Configurar sites do IIS de acordo com [as melhores práticas de segurança](#) é uma atividade demorada e contínua. Talvez você até precise configurar [níveis de confiança](#), o que também aumenta a sobrecarga de gerenciamento. Os contêineres são projetados para serem apátridas e imutáveis. Em última análise, suas implantações são mais rápidas, mais seguras e reproduzíveis se você usar contêineres do Windows em vez disso.

Próximas etapas

Investir em uma infraestrutura moderna para executar suas cargas de trabalho legadas traz imensos benefícios para sua organização. AWS os serviços de contêiner facilitam o gerenciamento de sua infraestrutura subjacente, seja no local ou na nuvem, para que você possa se concentrar na inovação e nas necessidades de sua empresa. Quase 80% de todos os contêineres na nuvem funcionam AWS atualmente. AWS fornece um rico conjunto de serviços de contêiner para praticamente todos os casos de uso. Para começar, consulte [Contêineres em AWS](#).

Recursos adicionais

- [Otimize o custo de cargas de trabalho de contêineres com provedores de capacidade ECS e instâncias spot do EC2 \(blog\)AWS](#)
- [Lista de verificação de otimização de custos para Amazon ECS e AWS Fargate \(AWS Blog\)](#)
- [Amazon EKS no AWS Graviton2 disponível ao público em geral: considerações sobre aplicativos de várias arquiteturas \(Blog\)AWS](#)
- [Otimização de custos para Kubernetes em AWS \(Blog\)AWS](#)
- [Otimizando seus custos de computação do Kubernetes com a consolidação do Karpenter \(Blog\)AWS](#)

Otimize os custos AWS Fargate das tarefas no Amazon ECS

Visão geral

AWS Fargate As tarefas de dimensionamento correto são uma etapa importante para a otimização de custos. Muitas vezes, os aplicativos são criados com tamanhos arbitrários para tarefas do

Fargate e nunca são revisitados. Isso pode causar superprovisionamento das tarefas do Fargate e gastos desnecessários. Esta seção mostra como usar para fornecer recomendações práticas [AWS Compute Optimizer](#) para que você possa otimizar a CPU e a memória de tarefas para seus serviços Amazon Elastic Container Service (Amazon ECS) executados no Fargate. O Compute Optimizer também quantifica o impacto nos custos da adoção dessas recomendações. Isso permite que você priorize seus esforços de otimização com base no tamanho da oportunidade de economia. As recomendações do Compute Optimizer fornecem configurações de CPU e memória em nível de contêiner para tarefas de redução de tamanho.

Benefícios de custo

O dimensionamento correto das tarefas do Amazon ECS no Fargate pode reduzir os custos em 30 a 70 por cento para tarefas de longa execução. Sem revisar as métricas de desempenho do aplicativo para dimensionar corretamente o tamanho da tarefa, você pode aplicar a mesma mentalidade usada nas instâncias de computação do EC2 ao dimensionamento de contêineres. Isso resulta em tarefas superdimensionadas do Fargate que aumentam os custos de recursos ociosos. Você pode usar o Compute Optimizer para descobrir as oportunidades de dimensionamento certas de forma reativa. O ideal é que o proprietário do aplicativo revise as métricas específicas de desempenho do aplicativo e remova a sobrecarga do sistema operacional para garantir que o tamanho adequado da tarefa seja especificado. Para obter mais informações, consulte a seção [Mover aplicativos do Windows para contêineres](#) deste guia.

Recomendações de otimização de custos

Esta seção oferece recomendações para usar o Compute Optimizer para dimensionar corretamente seu Amazon ECS nas tarefas do Fargate.

Como parte do processo de otimização de custos, recomendamos que você faça o seguinte:

- Ativar o Compute Optimizer
- Consuma os resultados do Compute Optimizer
- Marque as tarefas para que tenham o tamanho certo
- Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento
- Implemente recomendações de dimensionamento corretas
- Analise os custos antes e depois no Cost Explorer

Ativar o Compute Optimizer

Você pode habilitar [AWS Compute Optimizer](#) no nível da organização ou da conta única em AWS Organizations. A configuração de toda a organização fornece relatórios contínuos para instâncias novas e existentes em toda a sua frota para todas as contas dos membros. Isso permite que o dimensionamento correto seja uma atividade recorrente em vez de uma point-in-time atividade.

Nível da organização

Para a maioria das organizações, a forma mais eficiente de usar o Compute Optimizer é no nível organizacional. Isso fornece visibilidade de várias contas e várias regiões em sua organização e centraliza os dados em uma fonte para análise. Para habilitar isso no nível da organização, faça o seguinte:

1. Entre na sua [conta AWS Organizations de gerenciamento](#) com uma função que tenha as [permissões necessárias](#) e opte por se cadastrar em todas as contas dessa organização. A organização deve ter [todos os recursos habilitados](#).
2. Depois de ativar a conta de gerenciamento, você pode entrar na conta, ver todas as outras contas de membros e consultar suas recomendações.

Note

É uma prática recomendada configurar uma [conta de administrador delegado](#) para o Compute Optimizer. Isso permite que você exerça o princípio do menor privilégio, minimizando o acesso à conta de AWS Organizations gerenciamento e, ao mesmo tempo, fornecendo acesso ao serviço de toda a organização.

Nível de conta única

Se você tem como alvo uma conta com altos custos, mas não tem acesso a ela AWS Organizations, você ainda pode ativar o Compute Optimizer para essa conta e região. Para saber mais sobre o processo de aceitação, consulte [Introdução ao AWS Compute Optimizer](#).

Note

As recomendações são atualizadas diariamente e podem levar até 12 horas para serem geradas. Lembre-se de que o Compute Optimizer exige 24 horas de métricas nos últimos 14 dias para gerar recomendações para o Amazon ECS no Fargate. Para obter mais

informações, consulte [Requisitos para os serviços do Amazon ECS no Fargate na documentação](#) do Compute Optimizer.

O Compute Optimizer analisa automaticamente as seguintes métricas de utilização do Amazon e do CloudWatch Amazon ECS para seus serviços do Amazon ECS no Fargate:

- `CPUUtilization`— A porcentagem da capacidade da CPU usada no serviço.
- `MemoryUtilization`— A porcentagem de memória usada no serviço.

Consuma os resultados do Compute Optimizer

Considere um exemplo que se concentra em fazer as alterações de tamanho corretas em uma única conta e região. Neste exemplo, o Compute Optimizer está ativado no nível da organização em todas as contas. Lembre-se de que o dimensionamento correto é um processo disruptivo que, na maioria dos casos, é executado com precisão pelos proprietários do aplicativo durante uma janela de manutenção programada de várias semanas.

Se você navegar até o Compute Optimizer de dentro da conta de gerenciamento de uma organização (conforme mostrado nas etapas a seguir), poderá escolher a conta que deseja investigar. Neste exemplo, uma tarefa está sendo executada em uma única conta superprovisionada. `us-east-1` O foco é redimensionar para o tamanho recomendado para o serviço Amazon ECS.

1. Abra o console do [Compute Optimizer](#).
2. Na página Dashboard, filtre por `Findings=overprovisioned` para ver todos os serviços do Amazon ECS no Fargate.
3. Para revisar as recomendações detalhadas para serviços ECS provisionados em excesso no Fargate, role para baixo e escolha Exibir recomendações.
4. Escolha Exportar e salve o arquivo para uso futuro.

Note

Para salvar recomendações para análise futura, você deve ter um bucket S3 disponível para gravação no Compute Optimizer em cada região. Para obter mais informações, consulte a [política de bucket do Amazon S3 AWS Compute Optimizer](#) na documentação do Compute Optimizer.

Para ver as recomendações do Compute Optimizer, faça o seguinte:

1. No console do [Compute Optimizer](#), acesse a página de recomendações de exportação.
2. Para o destino do bucket S3, escolha seu bucket S3.
3. Na seção Filtros de exportação, em Tipo de recurso, escolha Serviços ECS no Fargate.
4. Na página Recomendações para serviços ECS no Fargate, analise um dos serviços ECS no Fargate e veja as recomendações de CPU e memória do Compute Optimizer. Por exemplo, revise as recomendações nas seções Comparar as configurações atuais com o tamanho recomendado da tarefa e Comparar as configurações atuais com o tamanho recomendado do contêiner.

Para obter a lista de serviços do ECS para Fargate que você precisa do tamanho certo, faça o seguinte:

1. Abra o [console Amazon S3](#).
2. No painel de navegação, escolha Buckets e, em seguida, escolha o bucket para o qual você exportou seus resultados.
3. Na guia Objetos, selecione seu objeto e escolha Baixar.
4. Nos resultados baixados, filtre a coluna de descoberta para mostrar somente serviços Amazon ECS OVER_PROVISIONED no Fargate. Isso mostra os serviços do Amazon ECS que você planeja direcionar para o tamanho certo.
5. Armazene as definições da tarefa em um editor de texto para uso posterior.

Tarefas de etiquetas de dimensionamento correto

Marcar suas cargas de trabalho é uma ferramenta poderosa para organizar seus recursos em AWS. Você pode usar etiquetas para obter uma visibilidade detalhada dos custos e permitir o estorno. Há muitos métodos e estratégias para adicionar tags aos AWS recursos para lidar com o estorno e a automação. Para obter mais informações, consulte o AWS whitepaper [Best Practices for AWS tagging](#) Resources. O exemplo a seguir é usado [AWS CloudShell](#) para marcar todas as tarefas que fazem parte de qualquer serviço do Amazon ECS na conta de destino e. Região da AWS

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
```

```
ClustersArns=$( w secs list-clusters -query 'clusterArns' -output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$( w secs list-services -cluster $ClustersArn -query 'serviceArns' -output
text)
  for ServiceArn in $ServiceArns; do
    TasksArns=$( w secs list-tasks -cluster $ClustersArn -service-name $ServiceArn -query
'taskArns' -output text)
    for TasksArn in $TasksArns; do
      w secs tag-resource -resource-arn $TasksArn -tags key=$TAG_KEY,value=$TAG_VALUE
    done
  done
done
```

O exemplo de código a seguir mostra como habilitar a [propagação de tags](#) para todos os serviços do Amazon ECS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$(aws ecs list-clusters --query 'clusterArns' --output text)
for ClustersArn in $ClustersArns; do
  ServiceArns=$(aws ecs list-services --cluster $ClustersArn --query 'serviceArns' --
output text)
  for ServiceArn in $ServiceArns; do
    aws ecs update-service --cluster $ClustersArn --service $ServiceArn --propagate-tags
SERVICE &>/dev/null
    aws ecs tag-resource --resource-arn $ServiceArn --tags key=$TAG_KEY,value=$TAG_VALUE
  done
done
```

Ative a etiqueta de alocação de custos para funcionar com ferramentas de AWS faturamento

Recomendamos ativar a tag de alocação de custos definida pelo usuário. Isso permite que a tag Rightsizing seja reconhecida e filtrável nas ferramentas de AWS cobrança (por exemplo, e). AWS Cost Explorer AWS Cost and Usage Report Se você não habilitar isso, a opção de filtragem de tags e os dados não estarão disponíveis. Para obter informações sobre o uso de tags de alocação de custos, consulte [Ativação de tags de alocação de custos definidas pelo usuário](#) na documentação. AWS Billing and Cost Management

Depois de esperar 24 horas, você pode ver a tag no Cost Explorer antes de implementar as recomendações de dimensionamento correto na próxima seção. Para fazer isso, pesquise a tag Rightsizing no Cost Explorer.

Implemente recomendações de dimensionamento corretas

O Compute Optimizer fornecerá recomendações de tamanho de tarefas ou contêineres. Para implementar as recomendações de dimensionamento correto, faça o seguinte.

1. Abra o [console do Amazon ECS](#).
2. Na barra de navegação, selecione a região que contém a definição de tarefa.
3. No painel de navegação, escolha Task definitions (Definições de tarefa).
4. Na página Task definitions (Definições de tarefa), escolha a tarefa e, em seguida, escolha Create new revision (Criar nova revisão).
5. Na página Create new task definition revision (Criar nova revisão da definição de tarefa), faça as alterações. Para atualizar a recomendação de tamanho do contêiner, `cpu` atualize e use `memory` o bloco ContainerDefinitions em [sua](#) definição de tarefa do ECS. Por exemplo: .

```
"containerDefinitions": [  
  {  
    "name": "your-container-name",  
    "image": "your-image",  
    "cpu": 1024,  
    "memory": 2048,  
  }  
],
```

6. Verifique as informações e escolha Create (Criar).

Para atualizar o serviço Amazon ECS, faça o seguinte:

1. Abra o [console do Amazon ECS](#).
2. Na página Clusters, selecione o cluster.
3. Na página Cluster overview (Visão geral do cluster), selecione o serviço e escolha Update (Atualizar).
4. Em Task definition (Definição de tarefa), escolha a família de definição de tarefa e a revisão que serão usadas.

Para operadores avançados, você pode usar CloudShell para atualizar o serviço Amazon ECS. Por exemplo: .

```
bash
#!/bin/bash
# Set variables
ClustersName="workshop-cluster"
ServiceName="lab7-fargate-service"
TaskDefinition="lab7-fargate-demo:3"
# update the service
aws ecs update-service --cluster $ClustersName --service $ServiceName --task-definition
$TaskDefinition
```

Analise os custos antes e depois

Depois de dimensionar corretamente seus recursos, você pode usar o Cost Explorer para mostrar os custos antes e depois usando a tag Rightsizing. Lembre-se de que você pode usar [tags de recursos](#) para controlar os custos. Ao usar várias camadas de tags, você pode obter visibilidade granular de seus custos. No exemplo abordado neste guia, a tag Rightsizing é usada para aplicar uma tag genérica a todas as instâncias de destino. Em seguida, uma tag de equipe é usada para organizar ainda mais os recursos. A próxima etapa é introduzir tags de aplicativo para mostrar ainda mais o impacto nos custos de operar um aplicativo específico.

Considere um exemplo da redução de custos que pode ser alcançada usando a tag Rightsizing para um único nível de conta. Neste exemplo, os custos operacionais vão de 30,26 USD por dia para 7,56 USD por dia. Supondo 744 horas por mês, o custo anual antes do dimensionamento correto é de \$11.044,9. Após o dimensionamento correto, o custo anual cai para \$2.759,4. Isso se traduz em uma redução de 75% nos custos de computação dessa conta. Imagine o impacto disso em uma grande organização.

Antes de embarcar na viagem do tamanho certo, considere o seguinte:

- AWS oferece muitas opções para redução de custos. Isso inclui o [AWS OLA](#), onde AWS analisa suas instâncias locais antes de migrar para o. AWS O AWS OLA também fornece recomendações de dimensionamento e orientação de licenciamento corretos.
- Complete todo o tamanho correto antes de comprar [Savings Plans](#). Isso pode ajudá-lo a evitar compras excessivas em seu compromisso com Savings Plans.

Próximas etapas

Recomendamos as seguintes etapas:

1. Analise seu cenário atual e considere a conversão de volumes gp2 do Amazon EBS em volumes gp3.
2. Analise os [Savings Plans](#).

Recursos adicionais

- [Introdução ao Compute AWS Optimizer](#) (documentação)
- [Melhores práticas para a marcação de AWS recursos](#) (AWS whitepapers)
- [Contêineres do Windows ativados AWS](#) (AWS Workshop Studio)

Obtenha visibilidade dos seus custos do Amazon EKS

Visão geral

Uma visão holística é necessária para monitorar com eficácia o custo de uma implantação do Kubernetes. O único custo fixo e conhecido é para o plano de controle do Amazon Elastic Kubernetes Service (Amazon EKS). Isso inclui todos os outros componentes que compõem a implantação, da computação e armazenamento à rede, sendo uma quantidade variável com base nas necessidades do seu aplicativo.

Você pode usar o [Kubecost](#) para analisar o custo da sua infraestrutura Kubernetes, desde os [namespaces](#) e [serviços](#) até os [pods](#) individuais, e depois exibir os dados em um painel. O Kubecost apresenta custos em cluster, como computação e armazenamento, e custos out-of-cluster como buckets do Amazon [Simple Storage Service \(Amazon S3\)](#) e [instâncias do Amazon Relational Database Service](#) (Amazon RDS). O Kubecost fará recomendações de dimensionamento correto com base nesses dados e exibirá alertas críticos que podem afetar o sistema. O Kubecost pode [se integrar AWS Cost and Usage Report](#) para mostrar economias em [Compute Savings Plans, Reserved Instances e outros programas de descontos](#).

Benefícios de custo

O Kubecost fornece relatórios e painéis que visualizam o custo de suas implantações do Amazon EKS. Ele permite que você faça uma análise detalhada do cluster em cada um dos vários

componentes, como controladores, serviços, nós, pods e volumes. Isso lhe dá uma visão holística de seus aplicativos em execução em um ambiente Amazon EKS. Ao habilitar essa visibilidade, você pode agir de acordo com as recomendações do KubeCost ou visualizar os custos de cada aplicativo em um nível granular. O dimensionamento correto de um grupo de nós do Amazon EKS oferece a mesma economia potencial que as instâncias EC2 padrão. Se você conseguir dimensionar corretamente seus contêineres e nós, poderá remover o inchaço computacional do tamanho da instância necessária para executar o contêiner e do número de instâncias do EC2 necessárias no grupo de auto scaling.

Recomendações de otimização de custos

Para aproveitar as vantagens do KubeCost, recomendamos que você faça o seguinte:

1. Implemente o KubeCost em seu ambiente
2. Obtenha um detalhamento granular dos custos dos aplicativos do Windows
3. Nós de cluster do tamanho certo
4. Solicitações de contêiner do tamanho certo
5. Gerencie nós subutilizados
6. Corrija cargas de trabalho abandonadas
7. Agir de acordo com as recomendações
8. Atualizar nós autogerenciados

Implemente o KubeCost em seu ambiente

O [Amazon EKS Finhack Workshop](#) ensina como implantar um ambiente Amazon EKS configurado para usar o KubeCost em uma conta própria. AWS Isso permite que você tenha experiência prática com a tecnologia. Se você estiver interessado em realizar esse workshop em sua organização, entre em contato com a equipe da sua conta.

Para implantar o KubeCost em seu cluster Amazon EKS usando o [Helm](#), consulte a publicação [AWS e a KubeCost colabore para fornecer monitoramento de custos para clientes do EKS no blog](#). AWS Como alternativa, você pode consultar a [documentação oficial do KubeCost](#) para obter instruções sobre como instalar e configurar o KubeCost. Para obter informações sobre o suporte do KubeCost para nós do Windows, consulte [Windows Node Support na documentação](#) do KubeCost.

Obtenha um detalhamento granular dos custos dos aplicativos do Windows

Embora você possa obter economias de custo significativas usando [instâncias spot do Amazon EC2](#), você também pode se beneficiar do fato de que as cargas de trabalho do Windows tendem a ser monitoradas. O uso de Instâncias Spot depende do aplicativo, e recomendamos que você verifique se elas serão aplicáveis ao seu caso de uso.

Para obter uma análise granular dos custos de seus aplicativos do Windows, [faça login no Kubecost](#). Na página de navegação, escolha Poupança.

Nós de cluster do tamanho certo

No [Kubecost](#), escolha Savings na barra de navegação e escolha Rightsize your cluster node.

Considere um exemplo em que o Kubecost relata que o cluster está superprovisionado em termos de vCPU e RAM. A tabela a seguir mostra os detalhes e as recomendações do Kubecost.

	Atual	Recomendação: Simples	Recomendação: Complexo
Contagem total	US \$3462,57 por mês	US \$137,24 por mês	US \$303,68 por mês
Contagem de nós	4	5	4
CPU	74 vCPUs	10 vCPUs	8 vCPUs
RAM	152 GB	20 GB	18 GB
Detalhamento da instância	2 c5.xlarge + 2 mais	5 t3a.médio	2 c5n.large + 1 mais

Conforme descrito na postagem do blog Kubecost [Encontre um conjunto ideal de nós para um cluster Kubernetes](#), a opção simples utiliza um único grupo de nós, enquanto a complexa utiliza uma abordagem de grupo de vários nós. O botão Saiba como adotar pode realizar o redimensionamento do cluster com um clique. Isso requer a instalação do [Kubecost Cluster](#) Controller.

Se você estiver usando [nós autogerenciados do Windows](#) que não foram criados pelo [eksctl](#), consulte [Atualização de um grupo de nós autogerenciados existente](#). Essas instruções mostram como alterar o tipo de instância no modelo de execução do Amazon EC2 usado pelo grupo [Auto Scaling](#).

Solicitações de contêiner do tamanho certo

No [Kubecost](#), escolha Savings na barra de navegação e acesse a página Solicitar recomendações de dimensionamento correto. Esta página mostra a [eficiência](#) dos pods, as recomendações de tamanho correto e a economia estimada de custos. Você pode usar o botão Personalizar para filtrar por cluster, nó, namespace\ Controller e muito mais.

Como exemplo, considere que o Kubecost calculou que alguns de seus pods estão superprovisionados em termos de CPU e RAM (memória). Em seguida, a Kubecost recomenda que você se ajuste aos novos valores de CPU e RAM para obter sua economia mensal estimada. Para alterar os valores de CPU e RAM, você deve atualizar seu arquivo de [manifesto de implantação](#).

Gerencie nós subutilizados

No [Kubecost](#), escolha Economia na barra de navegação e, em seguida, escolha Gerenciar nós subutilizados.

Considere um exemplo em que a página mostra que um nó no cluster é subutilizado em termos de CPU e RAM (memória) e, portanto, pode ser drenado e encerrado ou redimensionado. Escolher os nós que não passam nas verificações de nós e pods fornecerá mais informações sobre por que eles não podem ser drenados.

Corrija cargas de trabalho abandonadas

No [Kubecost](#), escolha Savings na barra de navegação e, em seguida, escolha a página Cargas de trabalho abandonadas. Neste exemplo, você filtra por namespace chamado windows. Esta página mostra os pods que não atingiram o limite de tráfego e são considerados abandonados. Os pods precisam enviar ou receber uma certa quantidade de tráfego de rede durante o período definido.

Depois de considerar cuidadosamente que um ou mais pods foram abandonados, você pode economizar custos reduzindo o número de réplicas, excluindo a implantação, redimensionando-a para consumir menos recursos ou notificando o proprietário do aplicativo de que acredita que a implantação foi abandonada.

Agir de acordo com as recomendações

Na seção Dimensione corretamente seus nós do cluster, o Kubecost analisa o uso dos nós de trabalho no cluster e faz recomendações sobre o dimensionamento correto dos nós para reduzir os custos. Há dois tipos de grupos de nós que podem ser usados com o Amazon EKS: [autogerenciados e gerenciados](#).

Atualizar nós autogerenciados

Para obter informações sobre a atualização de nós autogerenciados, consulte [Atualizações de nós autogerenciados na documentação do Amazon EKS](#). Ele afirma que os grupos de nós criados com `eksctl` podem ser atualizados e devem ser migrados para um novo grupo de nós com a nova configuração.

Como exemplo, suponha que você tenha um grupo de nós do Windows chamado `ng-windows-m5-2xlarge` (que usa uma instância EC2 `m5.2xlarge`) e queira migrar os pods para um [novo grupo de nós](#) chamado `ng-windows-t3-large` (que é apoiado por uma instância EC2 `t3.large` para economizar custos).

Para migrar para um novo grupo de nós ao usar grupos de nós implantados pelo `eksctl`, faça o seguinte:

1. Para encontrar o nó em que o pod está atualmente, execute o `kubectl describe pod <pod_name> -n <namespace>` comando.
2. Execute o comando `kubectl describe node <node_name>`. A saída mostra que o nó está sendo executado em uma instância `m5.2xlarge`. Também corresponde ao nome do grupo de nós (`ng-windows-m5-2xlarge`).
3. Para alterar a implantação para usar o grupo de nós `ng-windows-t3-large`, exclua o grupo de nós `ng-windows-m5-2xlarge` e execute `kubectl describe svc,deploy,pod -n windows`. A implantação começa imediatamente a ser reimplantada agora que seu grupo de nós foi excluído.

Note

Haverá um tempo de inatividade do serviço quando você excluir o grupo de nós.

4. Execute o `kubectl describe svc,deploy,pod -n windows` comando novamente após alguns minutos. A saída mostra que os pods estão todos em estado de execução novamente.
5. Para mostrar que os pods agora estão sendo executados no grupo de nós `ng-windows-t3-large`, execute os `kubectl describe node <node_name>` comandos `kubectl describe pod <pod_name> -n <namespace>` and novamente.

Métodos alternativos de redimensionamento

Esse método se aplica a qualquer combinação de grupos de nós autogerenciados ou gerenciados. A postagem do blog [Migrar perfeitamente as cargas de trabalho do grupo de nós autogerenciados do EKS para os grupos de nós gerenciados pelo EKS](#) fornece orientação sobre como migrar suas cargas de trabalho de um grupo de nós com o tipo de instância superdimensionado para o grupo de nós que foi dimensionado corretamente sem nenhum tempo de inatividade.

Próximas etapas

O Kubecost facilita a visualização do custo de seus ambientes Amazon EKS. A profunda integração do Kubecost com o Kubernetes e as AWS APIs pode ajudar você a encontrar possíveis economias de custo. Você pode vê-las como recomendações no painel de economia do Kubecost. O Kubecost também pode implementar algumas dessas recomendações para você por meio do recurso de [controlador de cluster](#).

Recomendamos que você analise a step-by-step implantação no blog [AWS Containers e que a Kubecost colabore para fornecer monitoramento de custos para clientes do AWS EKS](#).

Recursos adicionais

- [Workshop do Amazon EKS \(Workshop do Amazon EKS\)](#)
- [AWS e a Kubecost colaboram para fornecer monitoramento de custos para clientes da EKS \(Blog\)AWS](#)
- Workshop [Amazon EKS Finhack \(estúdio de AWS oficina\)](#)
- [Contêineres do Windows ativados AWS](#) (AWS Workshop Studio)

Reorganize os aplicativos do Windows com o App2Container

Visão geral

[AWS App2Container](#) é uma ferramenta de linha de comando para migrar e modernizar aplicativos web Java e .NET em contêineres. O App2Container analisa e cria um inventário de todos os aplicativos executados em máquinas virtuais, máquinas virtuais, instâncias do Amazon Elastic Compute Cloud (Amazon EC2) ou em outros provedores de nuvem. Você seleciona o aplicativo que deseja colocar em contêineres. O App2Container empacota os artefatos e dependências do aplicativo em imagens de contêiner, configura as portas de rede e gera os artefatos de implantação

necessários do Amazon Elastic Container Service (Amazon ECS) e do Amazon Elastic Kubernetes Service (Amazon EKS), que são modelos de infraestrutura como código (IaC). O App2Container provisiona a infraestrutura de nuvem e os pipelines de CI\ CD necessários para implantar o aplicativo em contêineres em um ambiente de produção. Para obter mais informações, consulte [Como o App2Container funciona](#) na documentação do App2Container.

Com o App2Container, você pode migrar AWS e modernizar seus aplicativos como contêineres, além de padronizar a implantação e as operações de seus aplicativos. Você pode usar o App2Container para ajudar a criar rapidamente uma prova de conceito (PoC) ou acelerar a implantação de cargas de trabalho de produção em contêineres.

Há várias coisas que você deve ter em mente ao trabalhar com aplicativos do Windows. O App2Container oferece suporte à containerização de aplicativos ASP.NET implantados no Microsoft Internet Information Services (IIS), incluindo aplicativos Windows Communication Foundation (WCF) hospedados no IIS que são executados no Windows Server 2016, Windows Server 2019 ou Windows Server Core 2004. Para obter mais informações, consulte [Aplicativos compatíveis para Windows](#) na documentação do App2Container. O App2Container usa o Windows Server Core como imagem base para seus artefatos de contêiner, combinando a versão do contêiner do Windows Server Core com a versão do sistema operacional (OS) do servidor em que você executa os comandos de containerização. Essa abordagem separa o aplicativo do sistema operacional subjacente para que você possa atualizar o sistema operacional sem realizar uma migração tradicional.

Se você usa uma máquina de trabalho para colocar seu aplicativo em contêineres, a imagem base do contêiner, como o canal de manutenção de longo prazo (LTSC) do Windows Server 2019, corresponde ao sistema operacional da máquina de trabalho, como o Windows Server 2019. Se você estiver executando a containerização diretamente nos servidores de aplicativos, a versão corresponde ao sistema operacional do seu servidor de aplicativos. Se seus aplicativos estiverem sendo executados no Windows Server 2008 ou 2012 R2, você ainda poderá usar o App2Container configurando uma máquina de trabalho para as etapas de containerização e implantação. O App2Container não oferece suporte a aplicativos executados em sistemas operacionais clientes Windows, como o Windows 7 ou o Windows 10. O App2Container oferece suporte às estruturas Tomcat, TomEE e JBoss (modo autônomo) para processos Java. Para obter mais informações, consulte [Compatibilidade com App2Container](#).

Benefícios de custo

A containerização e a consolidação de seus aplicativos podem gerar até [60% de economia de computação](#) em comparação com um padrão de design de implantação one-application-to-one de servidor. O App2Container ajuda a agilizar o processo de containerização de aplicativos. A seguir estão alguns dos benefícios de usar o App2Container para suas necessidades de modernização:

- O App2Container é oferecido sem custo adicional.
- O App2Container oferece suporte a vários aplicativos em uma imagem de contêiner.
- Solucione os sistemas operacionais que estão chegando ao fim do suporte usando o App2Container para mover seus aplicativos.NET legados para contêineres. Você pode migrar para um sistema operacional mais novo, evitar pagar por suporte estendido e reduzir os riscos de segurança.
- Os contêineres são um método eficiente e econômico de empacotar seus aplicativos.NET. Analise os benefícios dos contêineres na [Recomendação MACO - Movendo para contêineres](#).
- A consolidação e a containerização de aplicativos ajudam a reduzir seu espaço de computação, armazenamento e licenciamento usando seus recursos computacionais com mais eficiência.
- A mudança para contêineres pode reduzir a sobrecarga operacional e os custos de infraestrutura, além de aumentar a portabilidade do desenvolvimento e a agilidade de implantação.

Recomendações de otimização de custos

Para obter instruções sobre como usar o App2Container, consulte [Introdução ao. AWS App2Container](#) Para obter informações sobre os comandos do App2Container, consulte a referência do comando do [App2Container](#).

Próximas etapas

O App2Container pode acelerar o processo de containerização de aplicativos e implantação no Amazon EKS ou no Amazon ECS. A implantação de aplicativos em contêineres reduz os custos de computação, rede e armazenamento e reduz a sobrecarga operacional dos operadores de aplicativos.

[Para uma experiência prática com o App2Container, consulte o Modernize with Workshop. AWS App2Container](#) Se você quiser ter uma experiência de aprendizado aprofundada, peça à sua equipe de AWS contas que configure um dia de imersão no App2Container.

Recursos adicionais

- [Colocando em contêineres aplicativos Windows complexos de várias camadas usando AWS App2Container](#) (AWS postagem do blog)
- [Containerizando aplicativos ASP.NET legados usando AWS App2Container](#) (postagem do blog)AWS
- [Aplicativos compatíveis com o App2Container](#) (documentação)AWS
- [Modernize com o AWS App2Container Workshop](#) (AWS Workshop Studio)
- [AWS App2Container Perguntas frequentes](#) (AWS site)

Armazenamento

Escolher o armazenamento certo para suas cargas de trabalho da Microsoft é uma decisão arquitetônica crítica. Como parte do processo de tomada de decisão, recomendamos que você desenvolva um plano de armazenamento e determine os requisitos funcionais para seus aplicativos e serviços. Este capítulo fornece uma visão geral das seguintes opções de armazenamento que podem influenciar seu planejamento.

Seções:

- [Amazon EBS](#)
- [Amazon FSx](#)
- [AWS Storage Gateway](#)

Amazon EBS

O Amazon Elastic Block Store (AmazonEBS) é um serviço de armazenamento em blocos totalmente gerenciado que permite armazenar volumes persistentes de armazenamento em nível de bloco que você pode usar com instâncias do Amazon Elastic Compute Cloud (AmazonEC2). Você pode aproveitar vários recursos da Amazon EBS para gerenciar e otimizar com eficácia seus recursos de armazenamento para cargas de trabalho do Windows na nuvem. Por exemplo, você pode usar EBS a Amazon para provisionar a quantidade exata IOPS e a taxa de transferência necessárias para sua carga de trabalho, selecionar entre uma variedade de tipos de volume para atender às suas necessidades de carga de trabalho e usar ferramentas para identificar e eliminar o desperdício de recursos de armazenamento. Esse controle granular sobre o desempenho e o uso do armazenamento ajuda você a otimizar seus recursos de armazenamento, evitando custos desnecessários.

Esta seção abrange os seguintes tópicos:

- [Migre EBS volumes da Amazon de gp2 para gp3](#)
- [Modificar EBS instantâneos da Amazon](#)
- [Excluir volumes Amazon EBS não anexados](#)

Migre EBS volumes da Amazon de gp2 para gp3

Visão geral

Uma unidade de estado sólido (SSD) é a opção de armazenamento padrão para cargas de trabalho de produção e de alto desempenho. A Amazon EBS oferece um [SSD volume de uso geral](#) para cargas de trabalho de médio a alto desempenho. O padrão em muitos Serviços da AWS (incluindo a AmazonEC2) é o [gp2](#), a segunda geração desses volumes de uso SSD geral. A terceira geração de uso geral SSDs, chamada [gp3](#), foi lançada em dezembro de 2020.

A oferta gp3 fez melhorias significativas nos aspectos de personalização de desempenho em relação à geração anterior. Para volumes Amazon EBS gp2, o desempenho está estreitamente associado ao tamanho do volume. Para cada 1 GB de capacidade, os volumes gp2 obtêm 3 IOPS de desempenho. Ou seja, um volume gp2 de 2.000 GB é capaz de 6.000 IOPS. Para volumes gp3, o desempenho pode ser personalizado independentemente da capacidade de armazenamento. Isso permite que até mesmo pequenos volumes de capacidade atinjam capacidades de desempenho de até 16.000 IOPS e 1.000 MB/s de taxa de transferência.

Outra grande mudança nos volumes gp3 é o desempenho básico IOPS. Os volumes do gp3 começam em 3.000 IOPS. Em comparação, os volumes gp2 devem atingir 1 TiB de tamanho antes de atingir a mesma capacidade de desempenho. Para o Windows Server, que geralmente tem uma unidade C: muito menor que 1 TiB, a atualização de gp2 para gp3 é uma melhoria significativa no desempenho.

Finalmente, o preço dos volumes gp3 é uma das maiores melhorias em comparação com os volumes gp2. Os volumes gp3 oferecem todos os recursos de desempenho aprimorados a 20% do custo dos volumes gp2.

Impacto nos custos

Com a capacidade de escalar o desempenho independentemente da capacidade, é importante entender os aspectos de preços da adição IOPS e da taxa de transferência. Para volumes de gp2, o preço é baseado na capacidade provisionada de 0,10 USD por GiB por mês. Para volumes gp3, o preço é semelhante aos volumes [provisionados de alto desempenho, que têm um custo de capacidade e um custo separado para IOPS SSD volume](#) adicional e taxa de transferência. IOPS

Conforme observado na tabela a seguir, os volumes gp3 têm um preço de capacidade de 0,08 USD por GiB por mês (20% mais barato que o gp2) e custos separados de 0,005 USD por mês provisionado acima de 3.000 USD e 0,04 USD por mês provisionado IOPS acima de 125 USD para IOPS taxa de transferência. MiBs MiBs

	gp3	gp2
Tamanho do volume	1 GiB – 16 TiB	1 GiB – 16 TiB
Linha de base IOPS	3.000	IOPS3/GiB (mínimo 100IOPS) até um máximo de 16.000 IOPS
		Volumes menores que 1 TiB podem estourar até 3.000 IOPS
Volume IOPS máximo/	16.000	16.000
Taxa de transferência de linha de base	125 MiBs	O limite de taxa de transferência está entre 128 MiBs e 250 MiBs, dependendo do tamanho do volume
Throughput máximo/volume	1.000 MiBs	250 MiBs
Preço	0,08 USD/GiB por mês 3.000 IOPS gratuitos e 0,005 USD/provisionado — mês acima de 3.000 IOPS 125 MiBs gratuitos e 0,04 MiBs USD/provisionado - mês acima de 125 MiBs	USD 0,10 USD/Gib por mês

⚠ Important

Embora os volumes gp3 tenham custos separados de capacidade e desempenho, os volumes gp3 são sempre mais baratos do que os volumes gp2 se estiverem configurados nos mesmos níveis de desempenho.

As tabelas a seguir mostram exemplos de economias de custo que podem ser obtidas com a conversão de volumes gp2 em gp3 em várias configurações de capacidade e desempenho.

Exemplo de configuração do gp2

Tamanho do volume (GiB)	Máximo IOPS	Rendimento () MiBs	Custo (USD/mês)
30	3000	128	\$3,00
100	3000	128	\$10,00
500	3000	250	\$50,00
1000	3000	250	\$100,00
2000	6000	250	\$200,00
6000	16000	250	\$600,00

Exemplo de configuração gp3 (linha de base)

Máximo IOPS	Rendimento () MiBs	Custo (USD/mês)	Redução de custos (em comparação com gp2)
3000	125	\$2,40	20%
3000	125	\$8,00	20%
3000	125	\$40,00	20%
3000	125	\$80,00	20%
3000	125	\$160,00	20%
3000	125	\$480,00	20%

Exemplo de configuração gp3 (correspondência com gp2)

Máximo IOPS	Rendimento (IOPS) MiBs	Custo (USD/mês)	Redução de custos (em comparação com gp2)
3000	128	\$2,52	16%
3000	128	\$8,12	19%
3000	250	\$45,00	10%
3000	250	\$85,00	15%
6000	250	\$180,00	10%
16000	250	\$550,00	8%

[Para uma análise de custos, consulte a seção Calculadora de economia de custos de migração de EBS gp2 para gp3 no recurso da Amazon. EBS](#) Você pode baixar a calculadora e usá-la para descobrir quanto você pode economizar migrando seus volumes gp2 para gp3.

Recomendações de otimização de custos

Para obter instruções sobre como concluir o processo de migração, consulte a publicação [Migre seus EBS volumes da Amazon de gp2 para gp3 e economize até 20% em custos no](#) blog de armazenamento. AWS

Recursos adicionais

- [Migre seus EBS volumes da Amazon de gp2 para gp3 e economize até 20% em custos \(Storage Blog\)AWS](#)
- [Crie uma regra AWS Config personalizada para otimizar os tipos de EBS volume da Amazon](#) (blog de operações e migrações AWS na nuvem)
- [Controle seus AWS custos excluindo EBS volumes não utilizados da Amazon](#) (blog de operações e migrações AWS na nuvem)
- [Utilitário de EBS migração da Amazon](#) (GitHub)
- [Como economizar com os anúncios do re:Invent de 2020 \(Cloud Financial Management\)AWS](#)
- [Workshop de otimização de custos](#) (AWS Well-Architected Labs)

- Calculadora de [economia de custos de migração de gp2 para gp3](#) (download)

Modificar EBS instantâneos da Amazon

Visão geral

Excluir EBS volumes e gerenciar a retenção e o arquivamento de instantâneos é um aspecto importante para controlar os custos desde o início. Você pode fazer backup dos dados dos seus EBS volumes no Amazon Simple Storage Service (Amazon S3) tirando snapshots. point-in-time Os instantâneos são backups incrementais, portanto, eles salvam somente os blocos nos dispositivos que foram alterados após o instantâneo mais recente. Isso minimiza o tempo necessário para criar o snapshot e economiza em custos de armazenamento ao não duplicar os dados. Cada instantâneo contém todas as informações necessárias para restaurar seus dados (de quando o instantâneo foi criado) em um novo EBS volume.

As cobranças por EBS instantâneos são calculadas em gigabytes por mês. Você é cobrado pelo tamanho do instantâneo e por quanto tempo você o mantém. Os preços variam de acordo com o nível de armazenamento. Para o [nível Padrão](#), você é cobrado somente pelos blocos alterados que são armazenados. Para o nível de arquivamento, você é cobrado por todos os blocos de instantâneos armazenados. [Você também é cobrado pela recuperação de instantâneos da camada de arquivamento](#). Veja a seguir exemplos de cenários para cada nível de armazenamento:

- Nível padrão — você tem um volume que armazena 100 GB de dados. Você é cobrado pelos 100 GB completos de dados do primeiro snapshot (snap A). No momento do próximo instantâneo (snap B), você tem 105 GB de dados. Em seguida, você será cobrado apenas pelos 5 GB adicionais de armazenamento para o snap incremental B.
- Nível de arquivamento — Você arquiva o snap B. O snapshot é então movido para o nível Archive e você é cobrado pelo bloco completo de snapshot de 105 GB.

Você pode usar o [Amazon Data Lifecycle Manager](#) para ajudá-lo a configurar um ciclo de vida para reter e gerenciar seus snapshots dentro do cronograma.

Impacto nos custos

As cobranças por EBS volumes e instantâneos são gerenciadas separadamente. EBSos snapshots são cobrados a uma taxa menor do que os volumes ativos EBS. Quando uma instância é encerrada, o valor do [DeleteOnTermination atributo](#) para cada EBS volume anexado determina se o volume deve ser preservado ou excluído. Por padrão, o DeleteOnTermination atributo é definido como

`True` para o volume raiz. Está configurado `False` para todos os outros tipos de volume. Isso cria situações em que o operador pretende excluir uma EC2 instância, mas deixa para trás volumes que foram adicionados à instância além do volume raiz. Para obter instruções sobre como verificar volumes (e seus snapshots associados) que você não precisa mais, consulte [Exibir informações sobre um EBS volume da Amazon](#) na EBS documentação da Amazon.

Por padrão, quando você cria um snapshot, ele é armazenado no nível Amazon EBS Snapshot Standard (nível padrão). Os snapshots armazenados na camada padrão são incrementais. Isso significa que apenas os blocos de volumes que foram alterados após o snapshot mais recente são salvos. O [Amazon EBS Snapshots Archive](#) é um novo nível de armazenamento que você pode usar para armazenamento de baixo custo e de longo prazo de seus snapshots raramente acessados que não exigem recuperação frequente ou rápida. A diferença de preço do padrão para o arquivamento é significativa e deve ser uma consideração importante ao configurar sua estratégia de captura instantânea. O Amazon EBS Snapshots Archive oferece custos de armazenamento de snapshots até 75% menores para snapshots que você planeja armazenar por 90 dias ou mais e que raramente precisa acessar.

Armazenamento de EBS snapshots da Amazon	Custo
Padrão	0,05 USD/GB por mês
Arquivo	0,0125 USD/GB por mês

Em ambientes menores, a economia de custos pode não ser significativa. A economia é mais significativa em grande escala, quando há várias contas e milhares de EC2 instâncias sem TBs EBS instantâneos, mesmo quando os EBS volumes são excluídos.

A tabela a seguir compara os níveis padrão e de arquivamento por mês com apenas 50 TB de uso. Mesmo nessa escala mais baixa, ainda são milhares de dólares em economias anuais.

Armazenamento de EBS snapshots da Amazon	Custo por mês	Custo por ano
Padrão 50 TB	\$312,50	\$3.750
Arquivar 50 TB	\$78,13	\$937,60
	Poupança anual	\$2.812,40

Recomendações de otimização de custos

A exclusão de um snapshot pode não reduzir os custos de armazenamento de dados de sua organização. Outros snapshots podem fazer referência aos dados desse snapshot e os dados referenciados serão sempre preservados. Por exemplo, quando você executa o primeiro snapshot de um volume com 10 GiB de dados, o tamanho do snapshot também é de 10 GiB. Como os snapshots são incrementais, o segundo snapshot que você tira do mesmo volume contém apenas os blocos de dados alterados desde que o primeiro snapshot foi tirado. O segundo snapshot também faz referência aos dados no primeiro snapshot. Se você alterar 4 GiB de dados e tirar um segundo instantâneo, o tamanho do segundo instantâneo será de 4 GiB. Além disso, o segundo snapshot faz referência aos 6 GiB inalterados no primeiro snapshot. Para obter mais informações, consulte [Por que meus custos de armazenamento não diminuíram depois que eu excluí um instantâneo do meu EBS volume e depois excluí o volume em si?](#) no Centro de AWS Conhecimento.

Considere o seguinte:

- Você não é cobrado por instantâneos que outra pessoa Conta da AWS possui e compartilha com sua conta. Você é cobrado somente quando copia o snapshot compartilhado para sua conta. Você também é cobrado pelos EBS volumes criados a partir do snapshot compartilhado.
- Se um instantâneo (snap A) for referenciado por outro snapshot (snap B), a exclusão do snap B pode não reduzir os custos de armazenamento. Quando você exclui um instantâneo, somente os dados exclusivos desse instantâneo são removidos. Os dados referenciados por outros instantâneos permanecem, e você é cobrado por esses dados referenciados. Para excluir um instantâneo incremental, consulte Exclusão [incremental de instantâneos na documentação da Amazon](#). EBS

A limpeza de instantâneos é uma prática operacional padrão ao executar suas cargas de trabalho em AWS. Com o tempo, os instantâneos podem resultar em cobranças caras por dados que você não precisa.

Recursos adicionais

- [Controle seus AWS custos excluindo EBS volumes não utilizados da Amazon](#) (blog de operações e migrações AWS na nuvem)
- [Excluir um EBS snapshot da Amazon](#) (EBSdocumentação da Amazon)
- [Workshop de otimização de custos](#) (AWS Well-Architected Labs)

- [Arquive automaticamente os Amazon EBS Snapshots com o Amazon Data Lifecycle Manager](#) (Storage Blog)AWS

Excluir volumes Amazon EBS não anexados

Visão geral

EBS Volumes não conectados (órfãos) podem gerar custos de armazenamento desnecessários em seu ambiente. AWS É essencial incorporar a revisão e a exclusão regulares de EBS volumes não utilizados e não utilizados como parte da higiene do ambiente. AWS É uma prática recomendada ter um processo para analisar continuamente o uso dos EBS volumes. Você pode usar o [AWS Compute Optimizer](#) para analisar instâncias subutilizadas. Esta seção ajuda você a identificar, gerenciar e excluir EBS volumes que estão desanexados ou subutilizados.

Amazon EBS

O [Amazon Elastic Block Store \(AmazonEBS\)](#) é um dispositivo em nível de bloco que oferece volumes de armazenamento para instâncias do Amazon Elastic Compute Cloud (AmazonEC2). EBS fornece armazenamento persistente, com a flexibilidade de se conectar e desconectar das EC2 instâncias. Isso significa que o ciclo de vida dos EBS volumes persiste mesmo se uma EC2 instância for encerrada. O [DeleteOnTermination](#) atributo é um recurso que controla se os EBS volumes anexados devem ser preservados ou excluídos após o encerramento da instância. Por padrão, o atributo é definido como `True` para o volume raiz, resultando na exclusão. Está configurado `False` para outros volumes, resultando em preservação.

Impacto nos custos

EBS Os volumes não anexados, também chamados de volumes não utilizados ou órfãos, incorrem nas mesmas cobranças que os volumes anexados, com base no tamanho e no tipo de armazenamento provisionado. Embora o custo médio das EBS cobranças da Amazon possa parecer mínimo de 0,10 USD por GB por mês, é crucial reconhecer que o acúmulo de EBS volumes não utilizados pode resultar em custos significativos ao longo do tempo.

Por exemplo, considere as consequências de reter 50 EBS volumes não utilizados, cada um provisionado com um tamanho de armazenamento de 100 GB, conforme mostra a tabela a seguir.

Número de volumes de armazenamento	Tipo de volume	Tamanho	Custo mensal total
50 volumes	gp2 (\$0,10) USD	100 GB	100 GB 50,00 EBS volumes por mês 0,10 USD = 500,00 USD USD USD

O cenário da tabela anterior gera uma redução de custo de aproximadamente \$500 por mês ou \$6.000 por ano. Esse é um passo efetivo em direção à redução de custos. Certifique-se de incorporar a exclusão de EBS volumes não acoplados como uma prática regular na higiene do AWS ambiente.

Recomendações de otimização de custos

Você pode usar AWS para automatizar facilmente a exclusão de volumes não anexados EBS. Por exemplo, você pode usar AWS Lambda, AWS Config CloudWatch, Amazon e AWS Systems Manager para definir critérios para excluir volumes não anexados com base na idade, tags e outras especificações. Você também pode usá-los Serviços da AWS para automatizar o processo de limpeza em grande escala.

Para evitar consequências não intencionais, recomendamos que você realize a devida diligência antes de excluir volumes não anexados. EBS

Gerencie volumes não anexados EBS

Recomendamos que você considere as seguintes práticas recomendadas:

- Atenda aos requisitos de conformidade — verifique se a exclusão de EBS volumes não anexados está em conformidade com os requisitos de governança e conformidade da sua organização.
- Defina políticas de backup e retenção de dados — [Antes de excluir um EBS volume independente, faça backup de todos os dados importantes em outro repositório de armazenamento \(por exemplo, Amazon S3\)](#). Para retenção de dados, os [EBS snapshots da Amazon](#) são uma forma mais econômica de reter dados do que EBS volumes, e eles podem restaurar o volume, se necessário, no futuro. Para obter mais informações sobre o gerenciamento eficaz de snapshots, consulte a seção [Modificar EBS snapshots da Amazon](#) deste guia.
- Verifique se há dependências — Verifique se há dependências entre EBS volumes não anexados e outros recursos. AWS Você pode usar o [AWS Management Console ou an API](#) para coletar

informações descritivas sobre seus EBS volumes, como tamanho, status e recursos associados. Essa é uma etapa importante para evitar a exclusão de recursos temporariamente desvinculados.

- Crie uma política de retenção — estabeleça um período de retenção para EBS volumes não anexados. Isso pode ajudá-lo a identificar o momento apropriado para excluir volumes não conectados, garantindo que seu AWS ambiente permaneça otimizado. Por exemplo, você pode criar uma EventBridge regra da [Amazon](#) para iniciar uma função Lambda de forma programada. A função Lambda pode usar o AWS SDK para identificar ativamente quaisquer EBS volumes não anexados, aplicar um mecanismo de marcação para facilitar o rastreamento e enviar notificações quando um EBS volume não anexado atinge ou excede um limite definido.
- Marcar EBS volumes não anexados — [Marcar](#) EBS volumes é uma prática útil que pode ajudar na organização e identificação de volumes com base em atributos como ambiente, aplicativo ou proprietário. Isso pode ser particularmente útil ao decidir quais volumes não anexados excluir, pois permite identificar rapidamente os volumes que não são mais necessários com base em suas tags.
- Garanta a exclusão segura — Analisar quando um EBS volume foi anexado pela última vez pode ajudá-lo a determinar se é seguro excluir o volume. Para obter mais informações, consulte [Como eu uso AWS CLI comandos para listar os anexos ou o histórico de destacamentos de um volume específico](#) da Amazon? EBS no Centro de AWS Conhecimento.
- Identifique EBS volumes subutilizados — Identificar e remover EBS volumes subutilizados é uma prática altamente recomendada para reduzir os custos de armazenamento e manter um ambiente otimizado. AWS AWS Trusted Advisor e [AWS Compute Optimizer](#) pode ajudá-lo a identificar EBS volumes subutilizados e fornecer recomendações para reduzir custos e melhorar a eficiência. Por exemplo, consulte [Configurar a automação para otimizar EBS volumes com AWS Trusted Advisor](#) (GitHub), [Estabelecer uma Trusted Advisor organização \(TAO\) no painel \(AWS Workshop Studio\)](#) e [Otimizar os custos dos volumes da EBS Amazon AWS Compute Optimizer](#) usando AWS (Storage Blog).

Automatize a limpeza de volumes não acoplados EBS

Recomendamos que você considere as seguintes ferramentas para ajudá-lo a automatizar a limpeza de volumes não acoplados EBS:

- [AWS APIs\(DescribeVolumes\)](#) — Você pode filtrar e encontrar EBS volumes não anexados usando AWS SDKs ou o AWS Command Line Interface (AWS CLI). Você pode economizar tempo e esforço automatizando esse processo com um script ou uma [função Lambda](#) que é executada em um cronograma. Um [exemplo de script](#) GitHub demonstra como isso funciona. O script usa o Lambda para analisar AWS CloudTrail registros e identificar volumes não anexados EBS.

- [AWS Systems Manager Automação](#) — Isso permite que você automatize as tarefas rotineiras de manutenção e remediação em sua infraestrutura. Para começar, [crie um runbook de automação](#), que define uma série de etapas a serem executadas em uma ordem específica. Por exemplo, você pode criar um runbook que primeiro cria um instantâneo do volume não anexado e depois exclui o EBS volume em si. Isso pode ajudá-lo a automatizar tarefas que, de outra forma, seriam demoradas e propensas a erros se fossem feitas manualmente.
- [AWS Config](#)— Isso permite que você avalie, audite e acompanhe as mudanças em seus AWS recursos ao longo do tempo. Ao capturar as alterações de configuração, você pode usar AWS Config para avaliar a conformidade, a governança e a utilização de recursos em seu ambiente. Por exemplo, AWS Config pode identificar [EBSvolumes não utilizados](#). Além disso, você pode associar a AWS Systems Manager automação AWS Config para corrigir automaticamente a exclusão de volumes não EBS utilizados.

Recursos adicionais

- [Exclua volumes não utilizados do Amazon Elastic Block Store \(AmazonEBS\) usando AWS Config e AWS Systems Manager](#) (orientação AWS prescritiva)
- [Controle seus AWS custos excluindo EBS volumes não utilizados da Amazon](#) (blog de operações e migrações AWS na nuvem)
- [AWSConfigRemediation- DeleteUnused EBSVolume](#) (Referência do runbook de AWS Systems Manager automação)

Amazon FSx

O Amazon FSx para Windows File Server é um serviço de armazenamento de arquivos totalmente gerenciado que é otimizado para cargas de trabalho do Windows. Ele fornece uma solução simples e escalável para executar seus aplicativos e cargas de trabalho baseados em Windows, sem a necessidade de gerenciamento complexo da infraestrutura de armazenamento. Você pode usar o FSx Windows File Server para provisionar e acessar facilmente o armazenamento compartilhado de arquivos que ofereça suporte nativo aos aplicativos do Windows, incluindo Microsoft SQL Server SharePoint, Microsoft e personalizados. NETaplicações. Além disso, o File Server FSx for Windows ajuda você a gerenciar custos fornecendo opções flexíveis de preços, como cotas de armazenamento pay-as-you-go e deduplicação automática de dados para reduzir o espaço ocupado pelo armazenamento e otimizar o desempenho e o custo.

Esta seção abrange os seguintes tópicos:

- [Escolha o armazenamento de SMB arquivos certo](#)
- [Habilite a deduplicação de dados na Amazon FSx](#)
- [Entenda a fragmentação de dados no FSx Windows File Server](#)
- [Entenda HDD o uso do volume na Amazon FSx](#)
- [Use uma única zona de disponibilidade](#)

Escolha o armazenamento de SMB arquivos certo

Visão geral

AWS oferece uma variedade de serviços de armazenamento totalmente gerenciados que oferecem os recursos avançados dos serviços de arquivos líderes do setor, combinando as mais recentes inovações de AWS infraestrutura e segurança. Você pode incorporar AWS serviços aos fluxos de trabalho de infraestrutura como código (IaC) e integrá-los aos serviços de AWS computação, monitoramento e proteção de dados. Para cargas de trabalho do Windows, você pode escolher entre dois serviços de arquivos totalmente gerenciados que podem ser usados para atender às necessidades do seu aplicativo: FSx para Windows File Server e Amazon FSx for NetApp ONTAP.

FSx para Windows File Server

O Amazon FSx para Windows File Server fornece armazenamento compartilhado totalmente gerenciado baseado no Windows Server e oferece uma ampla variedade de recursos administrativos, de acesso a dados e de gerenciamento de dados. FSx para Windows O File Server se integra facilmente aos ambientes Windows porque é um serviço nativo do Windows. Recomendamos usar o FSx Windows File Server para compartilhamentos de usuários e grupos, instâncias de cluster de failover Always On para SQL servidor, aplicativos Windows e infraestrutura de desktop virtual (VDI). FSx para Windows, o File Server também se integra bem ao Amazon FSx File Gateway, ao Amazon Kendra, aos logs de auditoria do Amazon S3 e ao Amazon Data Firehose.

FSx para ONTAP

FSx for ONTAP é baseado no sistema NetApp de ONTAP arquivos proprietário da. Isso exige algum nível de aprimoramento e é recomendado principalmente para usuários locais NetApp existentes. Os casos de uso típicos incluem compartilhamentos de usuários e grupos, instâncias de cluster de failover Always On para SQL servidor e aplicativos Windows. FSx for ONTAP compatível com vários protocolos, sistemas de arquivos maiores que 64 TB (escala PB sem um servidor de DFS namespace), clonagem, replicação, instantâneos, compactação (eficiência de armazenamento) e hierarquização inteligente de dados.

Impacto nos custos

FSx para Windows File Server

FSx for Windows File Server foi a primeira solução de armazenamento compartilhado AWS para implantar instâncias de cluster de failover para SQL servidor. Com FSx o Windows File Server, você pode iniciar instâncias de cluster de failover usando o licenciamento da edição SQL Standard. No entanto, isso impede que você confie nos grupos de disponibilidade Always On, que exigem licenças da edição SQL Server Enterprise. Ao mudar da edição SQL Server Enterprise Standard para a edição SQL Server Standard, você pode economizar de 65 a 75 por cento no licenciamento do [SQL Server](#).

Você pode usar o FSx Windows File Server para instâncias de cluster de failover para descarregar a E/S de armazenamento do armazenamento típico. EBS Ao transferir a I/O FSx para o Windows File Server, você pode reduzir as EC2 instâncias, que dependem da alta taxa de transferência da AmazonIOPS, sem afetar a taxa de EBS transferência do armazenamento.

FSx para ONTAP

Você pode usar o FSx for ONTAP para executar seu cluster de failover da Microsoft no protocolo de bloco iSCSI e se beneficiar da inicialização instantânea de arquivos do SQL servidor, do uso da replicação entre regiões, do suporte a antivírus e da SnapMirror clonagem. Se você criar várias cópias de bancos de dados para teste, a clonagem pode fazer uma diferença significativa no consumo de espaço e na rapidez com que essas cópias do banco de dados podem ser criadas. Além disso, você pode usar NetApp SnapCenter para gerenciar a funcionalidade de backup, restauração e clonagem com suas EC2 instâncias para SQL Server usando FSx for ONTAP. FSx for ONTAP também fornece hierarquização automática de SSD até um armazenamento em pool de capacidade de baixo custo para uma combinação de desempenho e economia.

FSx for ONTAP support NetApp file system (ONTAP), diferentemente FSx do Windows File Server, que oferece suporte a um sistema de NTFS arquivos nativo do Windows. O tamanho mínimo FSx para for ONTAP é de 1024 GB, enquanto FSx para o Windows File Server pode começar a partir de 32 GB.

Integração com o Microsoft Distributed File System

FSx para Windows File Server e FSx para ONTAP integração com o [Sistema de Arquivos Distribuído da Microsoft \(DFS\)](#) para integração perfeita em implantações existentes. Lembre-se do seguinte ao planejar sua arquitetura:

- FSx para Windows File Server e FSx para ONTAP suporta [DFSNamespaces \(DFSN\)](#) em ambos os tipos de implantação (várias zonas de disponibilidade e zonas de disponibilidade únicas).
- Somente FSx para Windows, o File Server oferece suporte [DFS à replicação \(DFSR\)](#) e somente ao usar zonas de disponibilidade únicas.

Recomendações de otimização de custos

O desempenho do Windows File Server e do FSx ONTAP for depende muito da configuração, assim como seus preços. FSx para Windows File Server, o preço depende principalmente da capacidade e do tipo de armazenamento, da capacidade de taxa de transferência, do backup e dos dados transferidos. Com FSx for ONTAP, você paga pelo SSD armazenamento SSD IOPS, pelo uso do pool de capacidade, pela capacidade de transferência e pelo backup.

Serviço de arquivos	Custo do armazenamento de 5 TB	Configuração	Região
FSx para Windows File Server	\$982,78	Zona de disponibilidade única SSD(15.000IOPS) 32 MBps Backup de 5 TB (sem economia de deduplicação)	Leste dos EUA (Norte da Virgínia)
FSx para ONTAP	\$979,28	Zona de disponibilidade única 100% SSD Nível de capacidade de leitura e gravação de 15.000 15.000 SSD IOPS 128 MBps	Leste dos EUA (Norte da Virgínia)

Serviço de arquivos	Custo do armazenamento de 5 TB	Configuração	Região
		Backup de 5 TB (sem economia de deduplicação)	

Lembre-se do seguinte:

- A deduplicação e a compactação permitem que você armazene mais dados em dispositivos físicos reduzindo o tamanho dos dados, mas você paga pelo armazenamento provisionado da unidade de estado sólido (SSD) ou da unidade de disco rígido (HDD).
- Você pode usar FSx for ONTAP para hierarquizar seus dados. É extremamente raro que 100% dos seus dados sejam acessados regularmente e exijam SSD armazenamento. Você pode mover dados frios e acessados com pouca frequência para um nível de capacidade para reduzir custos.
- Os preços mencionados aqui são calculados com dados de 100% no SSD nível e 15.000 IOPS no SSD nível.

Backup

Por padrão, tanto FSx para ONTAP quanto FSx para o Windows File Server armazenam seu backup totalmente gerenciado no Amazon S3. No entanto, com FSx for ONTAP há uma opção adicional de uso de backup SnapVault, que pode configurar os backups para residirem no nível de capacidade. O backup com SnapVault é um mecanismo autogerenciado que é mais econômico do que a opção padrão de backup totalmente gerenciado. A opção de backup totalmente gerenciado custa 0,05 USD por GB por mês. O SnapVault backup ativado FSx para ONTAP (10:1 SSD até a capacidade de armazenamento do pool) é de 0,03221 USD (0,9x0,0219+0,1x0,125).

Lembre-se do seguinte:

- AWS os backups gerenciados oferecem granularidade de uma hora. [SnapVault](#) permite que você vá até cinco minutos.
- Você pode usar as ferramentas NetApp da (como a CLI e API) para configurar os SnapVault relacionamentos e a replicação de instantâneos.
- Ative a política all de classificação em camadas em um SnapVault volume para usar a camada de capacidade como armazenamento para os dados de backup.

- SnapVault os destinos podem estar na mesma região Região da AWS, em várias regiões ou no local. Isso geralmente é para um único destino de backup do sistema de arquivos da Zona de Disponibilidade ou de várias Zonas de Disponibilidade. Em comparação, AWS Backup é apoiado pela resiliência regional do Amazon S3.

Tamanho certo

Você também pode economizar nos custos e tirar o máximo proveito do seu sistema de arquivos, dimensionando corretamente e evitando o provisionamento excessivo.

Para obter o tamanho certo, faça o seguinte:

1. Identifique suas necessidades atuais com base nos dados. Para cargas de trabalho típicas do Windows, você pode usar ferramentas integradas do sistema operacional, como o [Monitor de Desempenho](#).
2. No Monitor de Desempenho, use os contadores a seguir para avaliar suas necessidades atuais de desempenho. O intervalo de captura é definido para um segundo, com um tamanho máximo de registro de 1.000 MB e a substituição ativada.

```
Logman.exe create counter PerfLog-Short -o "c:\perflogs\PerfLog-Long.blg" -f bincirc -v mmdhhmm -max 1024 -c "\LogicalDisk(*)\*" "\Memory\*" "\.NET CLR Memory(*)\*" "\Cache\*" "\Network Interface(*)\*" "\Paging File(*)\*" "\PhysicalDisk(*)\*" "\Processor(*)\*" "\Processor Information(*)\*" "\Process(*)\*" "\Thread(*)\*" "\Redirector\*" "\Server\*" "\System\*" "\Server Work Queues(*)\*" "\Terminal Services\*" -si 00:00:01
```

3. Para iniciar a captura do log, execute o `logman start PerfLog-Short` comando. Para interromper a captura do log, execute o `logman stop PerfLog-Short` comando.

Note

Você pode encontrar arquivos de log de desempenho em `c:\perflogs` no servidor que executa a captura. Para obter mais informações, consulte [Visão geral do Monitor de Desempenho do Windows](#) na documentação da Microsoft.

4. Depois de identificar a configuração correta, teste se sua estimativa está correta no sistema de FSx arquivos da Amazon usando ferramentas de estresse de disco, como a Microsoft [DISKSPD](#).
5. Se você estiver satisfeito com o desempenho, vá para o compartilhamento de arquivos.

Recomendamos uma abordagem conservadora da capacidade de armazenamento, pois ela só pode ser ampliada. A capacidade de produção pode ser ampliada e reduzida conforme necessário.

Recursos adicionais

- [Amazon FSx para NetApp ONTAP FAQs](#) (AWS site)
- [Otimizando o desempenho do Amazon FSx para Windows File Server com novas métricas](#) (AWS Storage Blog)

Habilite a deduplicação de dados na Amazon FSx

Visão geral

A deduplicação de dados é um recurso que permite armazenar seus dados com mais eficiência e com menos requisitos de capacidade. Ela envolve encontrar e remover a duplicação nos dados sem comprometer sua fidelidade ou integridade. A deduplicação de dados usa fragmentação e compactação de tamanho variável de subarquivos, que oferecem taxas de otimização de 2:1 para servidores de arquivos gerais e de até 20:1 para dados de virtualização. A deduplicação de dados é muito mais eficaz do que NTFS a compactação. A resiliência durante falhas de hardware é inerente à arquitetura de deduplicação, com validação completa da soma de verificação dos dados e metadados, incluindo redundância para metadados e os fragmentos de dados mais acessados.

FSx para Windows File Server oferece suporte total à deduplicação de dados. Seu uso pode resultar em uma economia média de 50 a 60% em compartilhamentos de arquivos de uso geral. Em ações, as economias variam de 30 a 50% para documentos de usuários e de 70 a 80% para conjuntos de dados de desenvolvimento de software. É importante entender que a economia de armazenamento que você pode obter com a deduplicação de dados depende da natureza do seu conjunto de dados, incluindo a quantidade de duplicação existente entre os arquivos. A deduplicação não é uma boa opção se os dados armazenados forem de natureza dinâmica.

Impacto nos custos

Para lidar com o crescimento do armazenamento de dados na empresa, os administradores consolidam os servidores e fazem do dimensionamento da capacidade e da otimização de dados os principais objetivos. As configurações padrão da deduplicação de dados podem proporcionar economia imediata, ou os administradores podem ajustar as configurações para obter ganhos adicionais. Por exemplo, você pode configurar a eliminação de duplicação para ser executada somente em determinados tipos de arquivo ou criar uma programação de trabalho personalizada.

Em um alto nível, a deduplicação tem três tipos de trabalhos: otimização, coleta de lixo e depuração. Esteja ciente de que o espaço não será liberado até que você execute um trabalho de coleta de lixo após a otimização. Você pode programar o trabalho ou executá-lo manualmente. Todas as configurações disponíveis quando você agenda um trabalho de deduplicação de dados também estão disponíveis quando você inicia um trabalho manualmente (exceto aquelas que são específicas do agendamento).

Mesmo com uma economia efetiva de apenas 25% com a deduplicação, há uma economia significativa FSx para o Windows File Server. Essas economias projetadas são baseadas em uma [estimativa](#) no AWS Pricing Calculator.

Recomendações de otimização de custos

A deduplicação ativada FSx para sistemas de arquivos do Windows File Server não está habilitada por padrão. Para habilitar a deduplicação usando o [gerenciamento remoto ativado PowerShell](#), você deve executar o `Enable-FSxDedup` comando e, em seguida, usar o `Set-FSxDedupConfiguration` comando para definir a configuração. Para obter mais informações, consulte [Administração de sistemas de arquivos](#) na documentação do FSx Windows File Server.

Para ativar a deduplicação, execute o seguinte comando:

```
PS C:\Users\Admin> Invoke-Command -ComputerName amznfsxzzzzzzzzz.corp.example.com -  
ConfigurationName FSxRemoteAdmin -ScriptBlock {Enable-FsxDedup }
```

Para verificar sua configuração de deduplicação, execute o seguinte comando:

```
Invoke-Command -ComputerName amznfsxzzzzzzzzz.corp.example.com -ConfigurationName  
FSxRemoteAdmin -ScriptBlock {  
Set-FSxDedupSchedule -Name "CustomOptimization" -Type Optimization -Days  
Mon,Tues,Wed,Sat -Start 09:00 -DurationHours 7  
}
```

Ao executar o PowerShell `Measure-DedupFileMetadata` cmdlet, você pode determinar quanto espaço em disco potencial pode ser recuperado em um volume se você excluir um grupo de pastas, uma única pasta ou um único arquivo e, em seguida, executar um trabalho de coleta de lixo. Especificamente, o `DedupDistinctSize` valor informa quanto espaço você recupera se excluir esses arquivos. Os arquivos geralmente têm partes que são compartilhadas em outras pastas, então o mecanismo de deduplicação calcula quais partes são exclusivas e seriam excluídas após o trabalho de coleta de lixo.

As [programações de trabalho padrão de deduplicação de dados](#) foram projetadas para funcionar bem nas cargas de trabalho recomendadas e serem o mais não intrusivas possível (excluindo a tarefa de otimização prioritária habilitada para o tipo de uso de backup). Se as cargas de trabalho tiverem grandes requisitos de recursos, recomendamos que você agende trabalhos executados somente durante as horas de inatividade ou para reduzir ou aumentar a quantidade de recursos do sistema que um trabalho de deduplicação de dados pode consumir.

Por padrão, a deduplicação de dados usa 25% da memória disponível. No entanto, isso pode ser aumentado usando `-memory switch`. Para trabalhos de otimização, recomendamos que você defina um intervalo de 15 a 50. Para trabalhos agendados, você pode usar um maior consumo de memória. Por exemplo, com trabalhos de coleta e depuração de lixo (que normalmente são programados para serem executados fora do horário de expediente), você pode definir um maior consumo de memória (como 50).

Para obter informações adicionais sobre as configurações de deduplicação de dados, consulte [Reduzindo os custos de armazenamento com a deduplicação de dados na documentação](#) do Windows FSx File Server.

Recursos adicionais

- [Compreendendo a deduplicação de dados \(documentação\)](#) da Microsoft)
- [Reduzindo os custos de armazenamento com a deduplicação de dados](#) (FSx para documentação do Windows File Server)

Entenda a fragmentação de dados no FSx Windows File Server

Visão geral

FSx para o Windows File Server, o desempenho depende da configuração. Ele se baseia principalmente no tipo de armazenamento, na capacidade de armazenamento e na configuração da taxa de transferência. A capacidade de taxa de transferência que você seleciona determina os recursos de desempenho disponíveis para o servidor de arquivos, incluindo os limites de E/S de rede, a memória CPU e os limites de E/S de disco impostos pelo servidor de arquivos. A capacidade de armazenamento e o tipo de armazenamento selecionados determinam os recursos de desempenho disponíveis para os volumes de armazenamento — os limites de E/S de disco impostos pelos discos de armazenamento. Além do desempenho, as opções de configuração também influenciam o custo. FSx para Windows File Server, o preço depende principalmente da

capacidade e do tipo de armazenamento, da capacidade de taxa de transferência, do backup e dos dados transferidos.

Se você tiver requisitos relativamente grandes de armazenamento e desempenho de arquivos, poderá se beneficiar da fragmentação de dados. A fragmentação de dados envolve [dividir os dados do arquivo em conjuntos de dados](#) menores (fragmentos) e armazená-los em diferentes sistemas de arquivos. As aplicações que acessam seus dados de várias instâncias podem alcançar altos níveis de performance lendo e gravando nesses fragmentos em paralelo. Ao mesmo tempo, você ainda pode apresentar às suas aplicações uma visão unificada sob um namespace comum. Além disso, também pode ajudar a escalar o armazenamento de dados de arquivos além do que cada sistema de arquivos suporta (64 TB) para grandes conjuntos de dados de arquivos — até centenas de petabytes.

Impacto nos custos

Para grandes conjuntos de dados, geralmente é mais eficaz implantar vários sistemas de arquivos pequenos FSx para Windows File Server, em vez de um grande SSD compartilhamento para alcançar o mesmo nível de desempenho. O uso de uma combinação dos tipos FSx de servidor de arquivos HDD e SSD armazenamento do Windows permite uma melhor economia de custos e permite que você combine a carga de trabalho com o melhor subsistema de disco subjacente. Nas tabelas a seguir, você pode ver a diferença entre um único sistema de arquivos de 17 TB e compará-lo com vários sistemas de arquivos menores que aumentam a mesma capacidade.

Sistema de SSD arquivos grande com várias cargas de trabalho

Nome do servidor	Custo	Configuração	Região
Servidor FSx de arquivos Amazon para Windows	\$5.716 USD	17 TB SSD 30% de deduplicação 256 Mbps Backup de 17 TB	Leste dos EUA (Norte da Virgínia)

Carga de trabalho particionada usando DFSN

Nome do servidor	Custo	Configuração	Região	Compartilhar
Servidor FSx de arquivos Amazon para Windows	\$1.024 USD	2 TB SSD 20% de deduplicação 128 Mbps Backup de 2 TB Multi-AZ	Leste dos EUA (Norte da Virgínia)	Compartilhe 1
Servidor FSx de arquivos Amazon para Windows	\$2.132 USD	5 TB SSD 30% de deduplicação 256 Mbps Backup de 5 TB Multi-AZ	Leste dos EUA (Norte da Virgínia)	Compartilhe 2
Servidor FSx de arquivos Amazon para Windows	\$1.036 USD	10 TB HDD 40% de deduplicação 128 Mbps Backup de 10 TB Multi-AZ	Leste dos EUA (Norte da Virgínia)	Compartilhe 3
DFSNEC2Instâncias do Windows	\$27 USD	t3a.medium 2 vCPUs	Leste dos EUA (Norte da Virgínia)	DFSNIInstâncias

Nome do servidor	Custo	Configuração	Região	Compartilhar
		Memória de 4 GiB		

O custo anual de um sistema de SSD arquivos grande é de \$68.592. O custo anual de uma carga de trabalho particionada é de \$50.640. Neste exemplo, é possível obter uma economia de 26% ao combinar a carga de trabalho com o armazenamento de back-end apropriado. Para obter mais informações sobre a estimativa de preços, consulte a [AWS Pricing Calculator](#) estimativa.

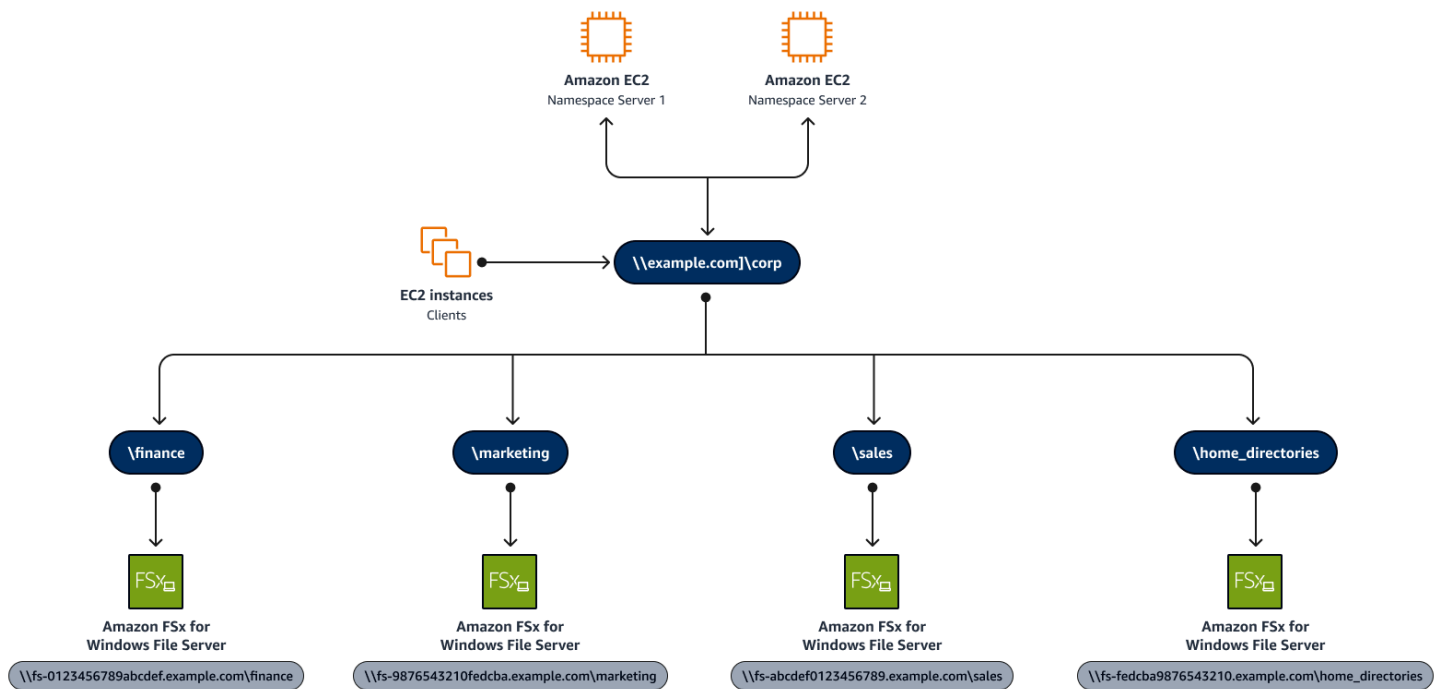
Recomendações de otimização de custos

Para implantar uma solução de deduplicação de dados, você deve configurar um [Microsoft DFS Namespace](#) com base no tipo de dados, tamanho de E/S e padrão de acesso de E/S. Cada namespace suporta até 50.000 compartilhamentos de arquivos e centenas de petabytes de capacidade de armazenamento em conjunto.

É mais eficiente escolher uma convenção de fragmentação que distribua a E/S uniformemente em todos os sistemas de arquivos que você planeja usar. O monitoramento de sua carga de trabalho ajudará na otimização adicional ou na redução de custos. Se precisar de ajuda para avaliar as informações de desempenho do sistema de FSx arquivos da Amazon, consulte FSx o [desempenho do Windows File Server](#) na documentação do FSx Windows File Server.

Depois de escolher uma estratégia de fragmentação, você pode agrupar os sistemas de arquivos para facilitar o acesso aos seus compartilhamentos usando DFS namespaces. Isso permite que os usuários vejam um sistema de arquivos homogêneo quando, na realidade, estão acessando uma variedade de sistemas de arquivos diferentes com casos de uso específicos. É importante criar os compartilhamentos com uma convenção de nomenclatura adequada para que seus usuários finais possam decifrar facilmente para qual carga de trabalho os compartilhamentos foram projetados. Também é importante rotular compartilhamentos de produção e de não produção, para que os usuários finais não coloquem arquivos no sistema de arquivos errado por engano.

O diagrama a seguir mostra como um único DFS namespace pode ser usado como ponto de acesso para vários sistemas de FSx arquivos da Amazon.



Lembre-se do seguinte:

- Você pode adicionar compartilhamentos existentes FSx do Windows File Server a uma DFS árvore.
- A Amazon não FSx pode ser adicionada à raiz do caminho de DFS compartilhamento. Você tem apenas uma subpasta.
- Você deve implantar uma EC2 instância para atender à configuração do DFS namespace.

Para obter mais informações sobre a configuração DFS -N, consulte [Visão geral dos DFS namespaces na documentação](#) da Microsoft. Para obter mais informações sobre o uso de DFS namespaces, consulte o vídeo Como [usar DFS namespaces com o FSx Amazon para Windows File Server](#) sobre. YouTube

Recursos adicionais

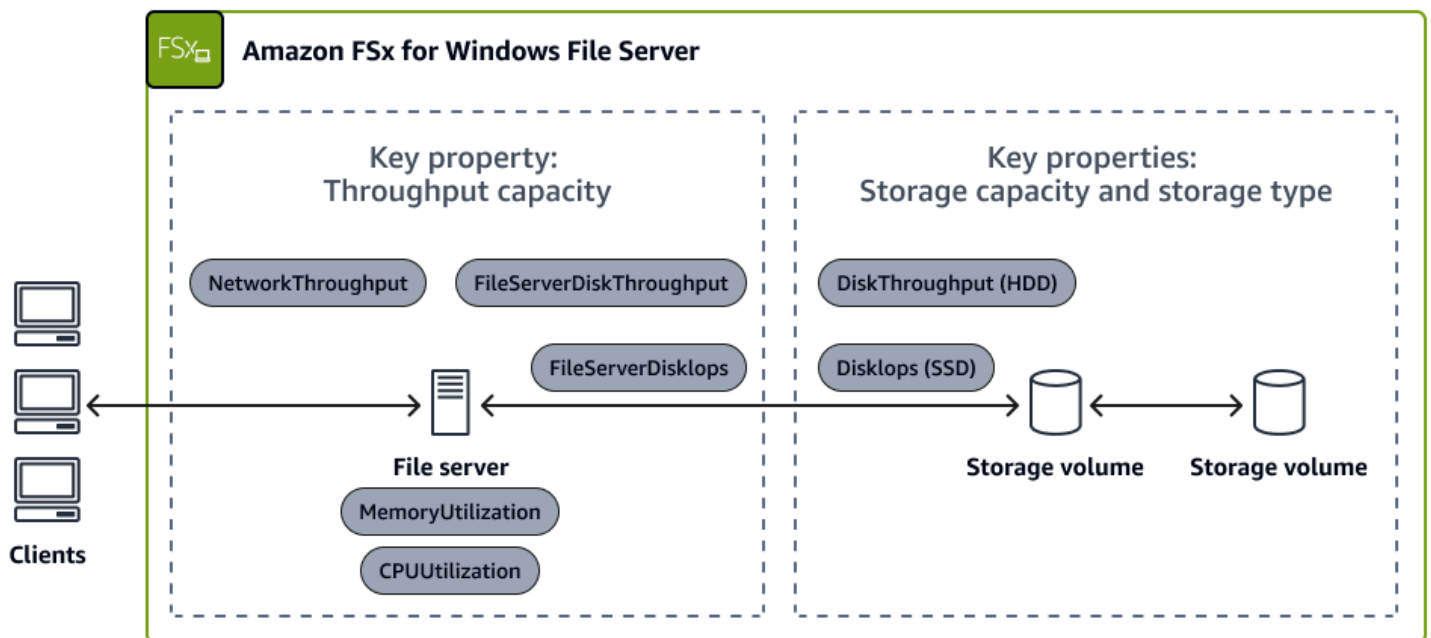
- [Agrupamento de vários sistemas de arquivos com DFS namespaces \(documentação da Amazon\) FSx](#)
- [Passo a passo 6: Escalando o desempenho com fragmentos](#) (documentação da Amazon) FSx
- [Usando DFS namespaces com o Amazon FSx para Windows File Server](#) (AWS Labs)

Entenda HDD o uso do volume na Amazon FSx

Visão geral

O Amazon FSx para Windows File Server oferece a flexibilidade de escolher a taxa de transferência independentemente da capacidade do sistema de arquivos. Duas configurações de capacidade estão disponíveis: HDD (unidade de disco rígido) e SSD (unidade de estado sólido). EBSAs [unidades st1](#) são usadas para o armazenamento do sistema de arquivos em HDD. EBSOs [drives io1](#) são usados para SSD.

O diagrama a seguir mostra a relação entre as configurações de taxa de transferência e armazenamento.



Com o armazenamento HDD baseado, você recebe uma IOPS linha de base de 12 com 80 discos intermitentes IOPS (por IOPs TiB de armazenamento) e uma taxa de transferência de 12 megabytes/segundo, linha de base com 80 megabytes/segundo (por TiB de armazenamento). Por exemplo, se seu compartilhamento tiver 50 TB, você obterá $50 * 12 = 600$ como linha de base tanto para a taxa de transferência quanto para IOPS.

O Amazon FSx para Windows File Server fornece 80 burst. IOPS Os créditos Burst são recarregados automaticamente quando sua utilização está abaixo da taxa básica e são consumidos automaticamente quando sua utilização está acima da taxa básica. Por exemplo, se sua carga de trabalho estiver utilizando apenas IOPS 10/TB por uma hora (IOPS2/TB abaixo da taxa básica), você

poderá utilizar IOPS 14/TB (IOPS2/TB acima da linha de base) na hora seguinte antes de ficar sem créditos intermitentes novamente.

Para operações de arquivos, o Amazon FSx for Windows File Server fornece latências consistentes de menos de um milissegundo com SSD armazenamento e latências de um dígito em milissegundos com armazenamento HDD. Para todos os sistemas de arquivos, incluindo aqueles com HDD armazenamento, o Amazon FSx for Windows File Server fornece um cache rápido (na memória) no servidor de arquivos, para que você possa obter alto desempenho e latências inferiores a um milissegundo para dados acessados ativamente, independentemente do tipo de armazenamento.

Quando apropriado, o uso do HDD armazenamento pode ajudar a reduzir o custo de sua capacidade geral de armazenamento e fornecer uma plataforma de armazenamento confiável para suas necessidades.

Impacto nos custos

O desempenho do Amazon FSx para Windows File Server depende de três fatores: capacidade de armazenamento, tipo de armazenamento e taxa de transferência. O desempenho de E/S de rede e o tamanho do cache na memória são determinados exclusivamente pela capacidade de taxa de transferência, enquanto o desempenho de E/S de disco é determinado por uma combinação de capacidade de taxa de transferência, tipo de armazenamento e capacidade de armazenamento.

Embora SSD seja recomendado para cargas de trabalho intensivas de E/S, há uma variedade de cargas de trabalho cujas necessidades podem ser atendidas com HDD especificações de desempenho. HDDo armazenamento foi projetado para um amplo espectro de cargas de trabalho, incluindo diretórios iniciais, compartilhamentos de usuários e departamentos e sistemas de gerenciamento de conteúdo. Por exemplo, se seus usuários precisam apenas de acesso de baixa latência aos dados que dão suporte aos projetos atuais, a maioria dos dados que você está armazenando é acessada com pouca frequência.

Você pode usar o [AWS Pricing Calculator](#) para fornecer uma comparação de 20 TB SSD com um sistema de HDD arquivos em us-east-1. Como mostra a tabela a seguir, mesmo sem economia na deduplicação, a diferença de custo é significativa ao comparar sistemas de HDD arquivos com sistemas de SSD arquivos.

Configuração do sistema FSx de arquivos da Amazon	Custos mensais
Multi-AZ de 20 TB SSD () us-east-1	\$4.699,30

Configuração do sistema FSx de arquivos da Amazon	Custos mensais
Multi-AZ de 20 TB HDD () us-east-1	\$542,88
Economia mensal estimada	\$4.156,42

Note

Para obter mais FSx economias sobre o Windows File Server, consulte a FSx seção [Habilitar a deduplicação de dados na Amazon](#) deste guia.

Ao identificar corretamente suas necessidades de desempenho, você pode selecionar o armazenamento certo para sua carga de trabalho e reduzir seus custos.

Recomendações de otimização de custos

Se você decidir usar o HDD armazenamento, teste seu sistema de arquivos para garantir que ele atenda aos seus requisitos de desempenho. HDDo armazenamento tem um custo menor em relação ao SSD armazenamento, mas com níveis mais baixos de taxa de transferência de disco e disco IOPS por unidade de armazenamento. Ele pode ser adequado para compartilhamentos de usuários de uso geral e diretórios pessoais com baixos requisitos de E/S, grandes sistemas de gerenciamento de conteúdo em que os dados são recuperados com pouca frequência ou conjuntos de dados com pequenos números de arquivos grandes.

O tipo de armazenamento de um sistema de arquivos existente não pode ser alterado. Para converter o tipo de armazenamento de um sistema de arquivos Amazon FSx para Windows File Server, você deve fazer backup do seu sistema de arquivos existente e restaurá-lo em um novo sistema de arquivos com o tipo de armazenamento desejado. Se você deseja converter um sistema de SSD arquivos existente em um sistema de HDD arquivos, saiba que ele HDD tem uma capacidade mínima muito maior de 2 TB.

Para restaurar um backup com um tipo de armazenamento diferente, faça o seguinte:

1. [Faça backup do seu sistema de arquivos existente.](#)
2. [Crie um novo sistema de FSx arquivos da Amazon](#) com o tipo HDD de armazenamento.
3. Restaure o backup no novo sistema de arquivos com o tipo de armazenamento desejado.

4. Verifique se o novo sistema de arquivos tem o tipo de armazenamento correto e se seus dados estão intactos.

Antes de transferir suas alterações para a produção, recomendamos que você analise o desempenho do seu sistema de FSx arquivos da Amazon e verifique se a alteração é aceitável. Para obter mais orientações, consulte a publicação [Otimizando o desempenho do Amazon FSx para Windows File Server com novas métricas](#) no blog AWS de armazenamento.

Recursos adicionais

- [Otimizando custos com a Amazon FSx](#) (FSxdocumentação da Amazon)

Use uma única zona de disponibilidade

Visão geral

Esta seção explica quando é mais vantajoso usar uma única implementação de zona de disponibilidade do [Amazon FSx para Windows File Server](#). Ele abrange cenários em que a mudança para uma única zona de disponibilidade reduz custos e ainda permite que você use o Amazon FSx for Windows File Server como seu serviço gerenciado de armazenamento de arquivos. Recomendamos que você implemente uma única zona de disponibilidade para a Amazon FSx para cargas de trabalho de produção. Isso pode ajudar a garantir que você tenha a redundância de várias zonas de disponibilidade.

Impacto nos custos

Um único sistema de arquivos de zona de disponibilidade oferece uma redução de custo de aproximadamente 40% em comparação com a implementação de várias zonas de disponibilidade. Com um sistema de arquivos com várias zonas de disponibilidade, você paga 0,230 USD por GB por mês SSD e 0,025 USD por GB por mês, em HDD comparação com 0,130 USD por GB por mês SSD e 0,013 USD por GB por mês em um único sistema de arquivos de zona de disponibilidade. HDD Você pode ver uma comparação de custos e criar suas próprias estimativas usando [AWS Pricing Calculatoro](#).

Para um sistema de arquivos de 10 TB, isso pode ser a diferença de pagar aproximadamente 1.200 USD por mês por várias zonas de disponibilidade ou 680 USD por mês por uma única zona de disponibilidade. Este [exemplo](#) usa um sistema de arquivos de 10 TB FSx para Windows File Server com SSD. A economia estimada com a desduplicação é de 50%. No geral, uma única zona

de disponibilidade tem um custo de entrada menor, mas vem com algumas ressalvas que serão abordadas na próxima seção.

Recomendações de otimização de custos

Implantações de zona de disponibilidade única

Para garantir que uma única zona de disponibilidade seja a opção certa, leve em consideração sua própria zona interna SLAs para os dados que estão sendo armazenados no FSx Windows File Server. Isso implica entender se você precisa fornecer SLAs aos seus clientes (internos e externos) e se os três nove de disponibilidade de uma FSx única zona de disponibilidade da Amazon ainda permitirão que você os atenda. SLAs FSx para Windows File Server com uma única zona de disponibilidade ainda tem um tempo de atividade de 99,9%. O valor da Amazon SLA FSx para várias zonas de disponibilidade é superior a 99,99 por cento. Para cargas de trabalho essenciais, recomendamos que você use várias zonas de disponibilidade em uma única zona de disponibilidade, mesmo com custo adicional.

As implantações de zona de disponibilidade única são ideais para cargas de trabalho, como backups para bancos de dados SQL do servidor. Eles podem fornecer armazenamento de baixo custo em um HDD nível, ao mesmo tempo em que oferecem um tempo de atividade consistente. Se você precisar de um nível mais alto de disponibilidade para uma carga de trabalho de produção, como SQL servidores de alta disponibilidade ou acesso a aplicativos de produção, uma única zona de disponibilidade não é a opção certa para suas cargas de trabalho. Para backups, testes de não produção e ambientes de desenvolvimento, uma implementação de zona de disponibilidade FSx única da Amazon pode reduzir seus custos operacionais.

Um caso de uso em que um sistema de arquivos de zona de disponibilidade FSx única da Amazon funciona bem é em uma situação de produção em que vários sistemas de arquivos de zona de disponibilidade FSx única da Amazon estão em uso, como armazenamento por servidor em um cluster de SQL servidores altamente disponível usando grupos de disponibilidade Always On. Para obter mais informações, consulte a postagem [Otimizando o custo de suas implantações de SQL servidores de alta disponibilidade no AWS](#) blog AWS de armazenamento.

Replicação em várias regiões

Uma opção potencial para reduzir custos com um único sistema de arquivos de zona de disponibilidade (em que apenas um único sistema de arquivos de zona de disponibilidade funciona) é se você quiser aproveitar as vantagens de uma replicação multirregional com a Amazon. FSx Você pode implantar [sistemas de arquivos Single-AZ](#) que suportem o uso com o Microsoft DFS -R nativo. DFS-R tem a capacidade de replicar dados automaticamente entre regiões e vários locais. Para

obter mais informações sobre como configurar DFS -R usando a AmazonFSx, consulte [Usando a replicação do sistema de arquivos distribuído da Microsoft na documentação](#) da Amazon. FSx

Outra alternativa para economia de custos em várias regiões é usar AWS Storage Gateway. Isso permite que você implemente um [Amazon FSx File Gateway](#) em outra região para acesso multirregional da AmazonFSx. Para obter mais informações, consulte a seção [AWS Storage Gateway](#) desse guia.

Se você trabalha em várias regiões, deve considerar o custo de transferência de dados para o tráfego de dados entre regiões. O tráfego que se movimenta entre regiões incorre em uma cobrança de 0,02 USD por GB. Portanto, se você tiver alterações consistentes de dados em grandes volumes, isso aumentará seu custo geral. Por [exemplo](#), 1 TB de transferência de dados equivale a aproximadamente 20,48 USD.

Janela de manutenção

A janela de manutenção é uma consideração importante se você estiver usando uma única zona de disponibilidade com a AmazonFSx. Durante a janela de manutenção, o sistema de FSx arquivos da Amazon fica indisponível por aproximadamente 20 minutos, devido a correções de software de rotina para o Windows Server subjacente. Se você estiver usando o sistema de arquivos para backups noturnos, ajuste a janela de FSx manutenção da Amazon adequadamente para evitar interrupções durante o backup. Você pode ajustar a [janela de manutenção](#) depois de criar seu sistema de FSx arquivos da Amazon.

Recursos adicionais

- [Disponibilidade e durabilidade: sistemas de arquivos Single-AZ e Multi-AZ \(documentação\)](#) da AmazonFSx)
- [Preços do Amazon FSx para Windows File Server](#) (AWS site)

AWS Storage Gateway

AWS Storage Gateway é um serviço de armazenamento em nuvem híbrida que conecta ambientes locais ao armazenamento AWS em nuvem. Ele permite que você integre perfeitamente sua infraestrutura local existente AWS, permitindo que você armazene e recupere dados da nuvem e execute aplicativos em um ambiente híbrido. Para cargas de trabalho do Windows, você pode usar o Storage Gateway para armazenar e acessar dados usando protocolos nativos do Windows, como SMB e NFS. Você pode usar o Storage Gateway para reduzir os custos associados à execução de

cargas de trabalho do Windows AWS usando hardware e software locais como uma ponte para a nuvem. Isso permite que você aproveite a escalabilidade e a economia AWS sem precisar fazer alterações significativas em sua infraestrutura existente.

Sob a égide do Storage Gateway, você obtém o Amazon S3 File Gateway, o Amazon File Gateway, o Tape Gateway e o Volume Gateway. FSx O S3 File Gateway e FSx o File Gateway são mais comumente usados com cargas de trabalho da Microsoft.

Gateway de arquivos Amazon S3

[O Amazon S3 File Gateway](#) permite que você armazene seus arquivos no Amazon S3 enquanto fornece acesso aos seus usuários usando compartilhamentos tradicionais. SMB Isso fornece uma interface de usuário familiar e ajuda a reduzir custos armazenando seus dados no Amazon S3 e aproveitando os vários níveis de armazenamento do Amazon S3. Você pode implementar o Storage Gateway com o S3 Intelligent Tiering para ajudá-lo a mover automaticamente os arquivos do ciclo de vida para os níveis de armazenamento de menor custo, a fim de reduzir ainda mais seus custos. Recomendamos o S3 File Gateway para acesso escalável e somente leitura, leituras rápidas repetidas (do cache) e despejos de banco de dados. Geralmente, não é recomendado para gravações, edição de arquivos ou compartilhamentos departamentais de alto desempenho ou alta disponibilidade.

Amazon FSx File Gateway

[O Amazon FSx File Gateway](#) também pode oferecer economia de custos ao trabalhar com sistemas de arquivos Amazon FSx Windows. Você pode criar um gateway de FSx arquivos para fornecer acesso localizado a um sistema de FSx arquivos da Amazon em outra região para evitar os custos de ter dois sistemas de arquivos independentes. Isso também pode ser útil se você tiver vários servidores de arquivos locais e quiser consolidá-los para evitar pagar por vários dispositivos de hardware.

Impacto nos custos

Gateway de arquivos Amazon S3

Configurar o S3 File Gateway é fácil porque você pode usar o assistente de inicialização do Storage Gateway. Você pode implantar o gateway em questão de minutos usando uma EC2 instância em seu AWS ambiente. Depois que o gateway for configurado, você poderá configurar os compartilhamentos do Storage Gateway para serem acessíveis por meio dos NFS protocolos SMB e. Para cargas de trabalho típicas do Windows, você também pode usar essa configuração para aproveitar um

ambiente do Active Directory e definir permissões em seus compartilhamentos de arquivos. Você pode integrar efetivamente um Storage Gateway ao seu uso normal, pois ele funcionará como um compartilhamento de arquivos típico do Windows. Arquivos e pastas são armazenados como objetos e listas de controle de NTFS acesso (ACLs) como metadados.

A tabela a seguir compara os custos de 10 TB de armazenamento com três opções de armazenamento disponíveis:

- FSx para Windows File Server
- Gateway de arquivos Amazon S3
- Amazon Elastic Block Store (AmazonEBS)

O preço para armazenar 10 TB de armazenamento é consideravelmente mais barato se você usar o Amazon S3, pois você pode particionar seus dados em vários níveis de uso. Na estimativa de preços, o S3 Intelligent Tiering é usado por sua flexibilidade de preços. Isso inclui 80% no S3 Standard, 10% no Acesso Infrequente e 10% no Amazon S3 Glacier. Embora você possa usar o S3 Glacier, é importante definir as regras de ciclo de vida adequadas para garantir que os arquivos movidos para o S3 Glacier não precisem ser acessados imediatamente. O S3 Glacier é exclusivamente para uso de arquivamento, não para uso de acesso regular.

Sistemas de armazenamento	Custo de 10 TB de armazenamento	Região
FSx para Windows File Server (supondo uma economia de 50% na desduplicação)	\$683,20 USD SSD	Leste dos EUA (Norte da Virgínia)
Gateway de arquivos Amazon S3	Hierarquização inteligente de \$449,51 USD	Leste dos EUA (Norte da Virgínia)
Amazon EBS	\$1.335,69 USD GP3	Leste dos EUA (Norte da Virgínia)

Considere o seguinte:

- No S3 Glacier, você recebe erros genéricos de E/S, a menos que use o [RestoreObjectAPI](#) para restaurar o objeto de volta ao Amazon S3. Recomendamos que você use uma notificação para

esse erro de E/S usando o Amazon CloudWatch Events. Dessa forma, sua equipe de operações pode reagir ao fato de um usuário receber esse erro em um arquivo que ele talvez precise acessar. Para obter mais informações sobre esses erros, consulte [Erro: InaccessibleStorageClass na documentação](#) do Amazon S3 File Gateway.

- Além da limitação de acesso ao S3 Glacier, são [ACLs permitidos apenas 10 por objeto/pasta no Storage Gateway](#). Antes de decidir usar o Storage Gateway, verifique se você não precisa de mais de 10 ACL entradas.

Amazon FSx File Gateway

Semelhante a um gateway de arquivos do Amazon S3, um gateway de FSx arquivos fornece acesso a um sistema de arquivos que retém os dados a longo prazo. No Amazon S3 File Gateway, os dados residem no Amazon S3. Para o FSx File Gateway, seus dados residem no FSx Windows File Server. Embora as opções Multi-AZ estejam disponíveis FSx para o Windows File Server, não há uma opção multirregional. Se você tem uma empresa global ou um escritório remoto, talvez seja necessário fornecer uma plataforma de armazenamento compartilhado que esteja geograficamente mais próxima do usuário final para evitar latência. Se você implantasse outro sistema de FSx arquivos da Amazon, isso adicionaria o custo de um sistema de arquivos Amazon FSx para Windows File Server totalmente novo e o armazenamento necessário. Para evitar a criação de um sistema de arquivos totalmente novo e a duplicação de custos, você pode implantar o FSx File Gateway na região secundária. Isso fornece acesso localizado aos arquivos para os usuários, ao mesmo tempo que ajuda a reduzir seus custos gerais.

Sistemas de armazenamento	Custo de 10 TB de armazenamento	Região
Servidor FSx de arquivos Amazon para Windows	\$683,20 USD SSD	Leste dos EUA (Norte da Virgínia)
Amazon FSx File Gateway	\$503,70/Gateway único	Leste dos EUA (Norte da Virgínia)

Note

Os preços na tabela anterior são baseados nos [preços do Storage Gateway](#).

Lembre-se do seguinte:

- FSxO File Gateway pode ajudar você a economizar aproximadamente 180 USD por mês (ou 2.100 USD por ano) para cargas de trabalho multirregionais.
- As taxas de transferência de dados são muito mais baixas com o FSx File Gateway, porque ele só precisa armazenar em cache os arquivos que estão sendo acessados regularmente e não uma cópia secundária completa.
- Embora você possa ter duas implantações do FSx Windows File Server em regiões diferentes e mantê-las atualizadas com AWS Backup ou AWS DataSync, nenhuma das opções está quase em tempo real.

Recomendações de otimização de custos

Gateway de arquivos Amazon S3

O S3 File Gateway oferece uma opção de baixo custo para armazenar arquivos, mas há alguns problemas a serem considerados em relação à forma como ele implementa e usa o sistema de arquivos. Por exemplo, o S3 File Gateway exige o uso de uma máquina virtual para executar o software Storage Gateway. Em AWS, o Storage Gateway é implantado na Amazon EC2 usando uma instância m5.xlarge, por padrão. Se quiser reduzir seus custos de armazenamento local, você pode implantar o Storage Gateway como um dispositivo virtual em plataformas de virtualização como VMware o Hyper-V.

Considerações sobre alta disponibilidade

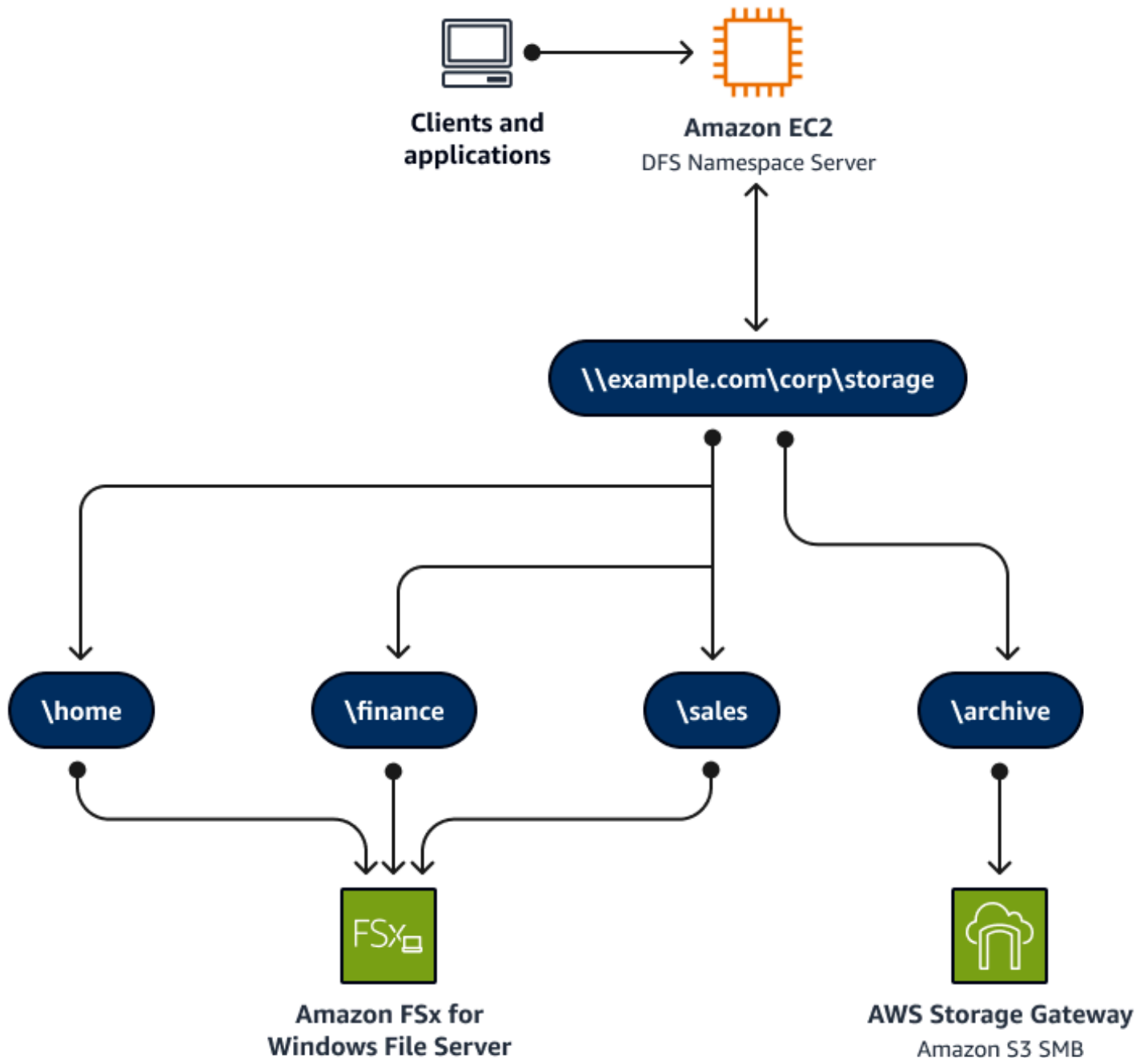
A execução do Storage Gateway é um ponto único de falha no acesso aos arquivos. Para evitar tempo de inatividade desnecessário, recomendamos que você implemente um controle de acesso rígido no qual os usuários possam fazer alterações ou interromper e iniciar a instância do Storage Gateway. Além disso, para implantações em AWS, é vantajoso usar o Amazon Data Lifecycle Manager para criar instantâneos de roteamento para recuperar rapidamente sua implementação do Storage Gateway. Se você estiver executando o Storage Gateway no local usando VMware, você pode configurá-lo para [alta disponibilidade](#).

Executando vários sistemas de arquivos

Separar suas cargas de trabalho de arquivos de uso diário de suas cargas de trabalho de arquivamento pode ajudá-lo a evitar custos de armazenamento desnecessários. O Storage Gateway tem a capacidade de ser implantado junto com um sistema de arquivos FSx para Windows File

Server. Ao usar [DFSnamespaces](#), você pode apresentar seu armazenamento principal de uso diário em execução no Windows File Server e seu armazenamento em execução no Amazon S3 (que é acessado FSx por meio do Storage Gateway).

O diagrama a seguir mostra como um único DFS namespace pode ser usado como ponto de acesso de front-end para diferentes opções de armazenamento de back-end.



Os clientes são direcionados para uma estrutura de pastas, como \\example.com\storage. Esse diretório principal contém os subdiretórios. Um sistema FSx de arquivos do Windows File Server contém os compartilhamentos de arquivos acessados normalmente. Você pode usar um compartilhamento de arquivos criado no Storage Gateway para arquivar dados. Os usuários podem arquivar itens manualmente na pasta de arquivamento ou você pode criar um processo para automatizar a movimentação de alguns arquivos dos compartilhamentos normais para a pasta de arquivamento.

Considere o seguinte:

- Analise seus requisitos de armazenamento e forneça [armazenamento adequado para o cache](#).
- Adicione seu gateway à configuração do Active Directory e use o [Windows padrão ACLs para acessar os arquivos](#).

FSxGateway de arquivos

A implantação do FSx File Gateway é semelhante à implantação do S3 File Gateway, mas é ainda mais fácil se você usar o assistente de inicialização. Para obter instruções detalhadas, consulte [Etapa 3: Criar e ativar um Amazon FSx File Gateway](#) na documentação do Amazon FSx File Gateway. Depois de implantar FSx o File Gateway em seu ambiente, você pode associá-lo aos seus sistemas de FSx arquivos existentes da Amazon e obter acesso aos seus arquivos.

O armazenamento é a principal consideração ao implantar o FSx File Gateway. O armazenamento padrão fornece 150 GB, o que é uma quantidade razoável de espaço para armazenar arquivos em cache. Criar alertas de monitoramento para pouco espaço livre pode ajudar no dimensionamento correto do armazenamento sem superalocação.

Recursos adicionais

- [AWS Storage Gateway recursos](#) (AWS documentação)

Active Directory

O Amazon Elastic Compute Cloud (Amazon EC2) executando o Windows Server é um ambiente seguro, confiável e de alto desempenho para a implantação de aplicativos e cargas de trabalho baseados no Windows. Você pode provisionar instâncias rapidamente e aumentar ou reduzir a escala conforme necessário, pagando apenas pelo que usar. Os serviços do Active Directory são usados como a principal fonte de gerenciamento de identidade em ambientes Windows Server.

Esta seção abrange os seguintes tópicos:

- [Active Directory autogerenciado no Amazon EC2](#)
- [AWS Managed Microsoft AD](#)
- [AD Connector](#)

Active Directory autogerenciado no Amazon EC2

Visão geral

Esta seção fornece recomendações para reduzir o custo de execução do Active Directory na Amazon Elastic Compute Cloud (Amazon EC2). O foco principal é garantir que você possa dimensionar adequadamente os controladores de domínio do Active Directory e usar a flexibilidade do Nuvem AWS para ajustá-los conforme necessário ao seu ambiente. AWS podem ajudá-lo a interromper facilmente uma instância e redimensioná-la para atender às suas necessidades em constante mudança, ou a reduzir o tamanho da instância se você aumentar a escala muito rápido. Escolher o tamanho e o tipo de instância corretos pode resultar em economias significativas.

Impacto nos custos

A tabela a seguir mostra a diferença entre escolher uma instância da família de instâncias com capacidade de intermitência em vez de uma instância de uso geral. Essa opção pode economizar uma quantia considerável de dinheiro a cada mês. O planejamento e o dimensionamento adequados da sua instância podem ajudar você a gerenciar os custos.

Tipo de instância	Número de instâncias	vCPU	Memória	Custo
t3a.medium	2	2	8	\$81,76/mês

Tipo de instância	Número de instâncias	vCPU	Memória	Custo
m5a.large	2	2	8	\$259,88/mês

Para obter mais informações sobre custos, consulte a AWS Pricing Calculator [estimativa](#).

Uma economia de \$178,12 por mês acaba sendo uma economia de mais de \$2.000 por ano para seus controladores de domínio. Lembre-se de que é para uma pequena área de apenas dois controladores de domínio em uma conta. Em grande escala com várias contas e controladores de domínio adicionais, essa economia pode resultar em uma redução significativa de custos.

Recomendações de otimização de custos

A Microsoft fornece [recomendações de planejamento de capacidade](#) para quando você estiver implantando seu ambiente do Active Directory. Recomendamos que você leve em consideração os seguintes componentes principais ao planejar ou escalar seu ambiente do Active Directory:

- Memória
- Rede
- Armazenamento
- Processador

Mantendo esses componentes principais em mente, você pode selecionar um tipo de instância que faça sentido para o seu ambiente do Active Directory AWS. Esta seção aborda alguns exemplos do Active Directory para cenários AWS de implantação. Esses cenários deixam claro que não é necessário replicar seu ambiente local em AWS, se você não planeja lidar com o mesmo número de usuários e computadores que você faz em seu ambiente local.

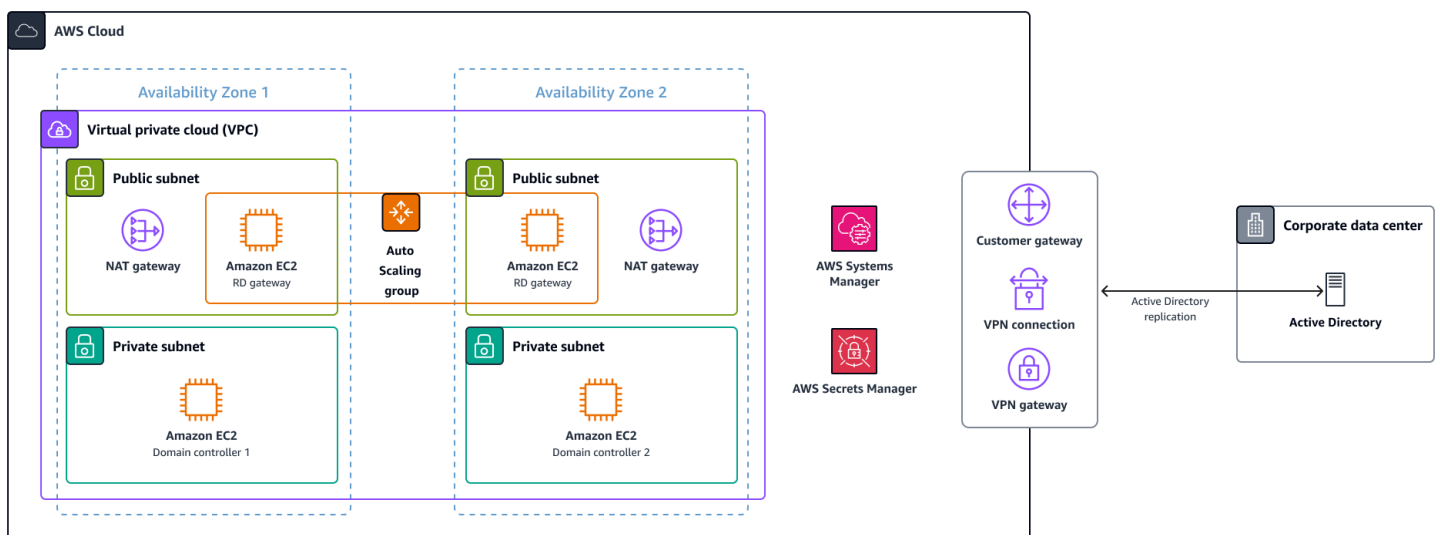
A tabela a seguir destaca componentes importantes relacionados à vCPU, memória e disco para seu AWS espaço físico.

Componente	Estimativas
Tamanho do armazenamento/banco de dados	40—60 KB para cada usuário
RAM	Tamanho do banco de dados

Componente	Estimativas
	Recomendações básicas do sistema operacional
	Aplicativos de terceiros
Rede	1 GB
CPU	1.000 usuários simultâneos para cada núcleo

Cenário de implantação híbrida

O diagrama a seguir mostra um exemplo de arquitetura para uma implantação híbrida do Active Directory.



Como mostra o diagrama, você normalmente tem uma área local e a expande para a nuvem AWS. Nas fases iniciais de uma migração, você normalmente não terá todos os seus usuários e servidores implantados na AWS. É por isso que é importante implantar inicialmente uma área menor para economizar dinheiro nos esforços de migração.

Se você quiser manter um espaço local com servidores e usuários se autenticando localmente, não precisará do mesmo espaço físico para controladores de domínio em AWS. Seguindo as melhores práticas do Active Directory, você pode implementar [sites e serviços adequados do Active Directory](#) para autenticar usuários e computadores em seu espaço local, ao mesmo tempo em que autentica seu espaço físico somente nos controladores de domínio em AWS. Isso permite que você

evite sobredimensionar sua área de cobertura do Active Directory, AWS limitando o uso apenas aos AWS recursos e não a toda a sua infraestrutura local. Para obter orientação sobre como criar uma configuração híbrida, consulte [Posicionamento adequado dos controladores de domínio e considerações sobre o site](#) na documentação da Microsoft.

Otimize para uma AWS migração por meio do dimensionamento correto

Se você estiver implantando uma nova instância do Active Directory para seus usuários ou planeja migrar totalmente AWS para sua infraestrutura do Active Directory, recomendamos que você planeje o dimensionamento de acordo com as recomendações da Microsoft para vCPU, memória e espaço em disco para a escolha de instâncias na tabela anterior.

Se esse for um novo espaço, você pode começar aos poucos e aproveitar a capacidade de [alterar facilmente os tipos de instância](#) para redimensionar seu ambiente à medida que ele cresce. AWS A seção [Windows no Amazon EC2](#) deste guia mostra como monitorar e analisar a utilização da CPU e da memória em. AWS Dessa forma, você sabe quando aumentar o tamanho da sua instância do EC2.

Se você estiver migrando totalmente seu ambiente local do Active Directory para AWS, poderá implementar os mesmos planos de dimensionamento para garantir o desempenho adequado. Antes de duplicar o que você tem no local AWS, recomendamos que você conclua uma análise completa do seu ambiente do Active Directory. Isso pode ajudar você a evitar o excesso de provisionamento. Certifique-se de usar o Monitor de Desempenho para coletar informações sobre a quantidade de tráfego e a utilização dos controladores de domínio existentes. Isso pode lhe dar uma compreensão do uso geral para que você possa dimensionar corretamente e, por fim, reduzir seus custos.

Otimize o Active Directory em AWS

Se você estiver executando o Active Directory AWS, é importante também monitorar continuamente a utilização e alterar os tamanhos das instâncias conforme necessário para reduzir seus gastos. Você pode usar AWS Compute Optimizer para obter informações sobre os recursos que você está usando AWS. Para obter informações sobre como usar o Compute Optimizer para dimensionar corretamente suas cargas de trabalho do Windows, consulte a seção Windows [no Amazon EC2 deste guia](#). Para uma análise mais abrangente, você pode usar o Monitor de Desempenho para monitorar a utilização dos controladores de domínio do Active Directory, avaliar o desempenho e redimensionar adequadamente.

Você também pode usar CloudWatch para monitorar o desempenho dos controladores de domínio. Para otimizar seus controladores de domínio (aumentando ou diminuindo a escala), você pode usar

as métricas disponíveis CloudWatch para ajudá-lo a tomar as decisões certas. Você pode usar o CloudWatch agente para configurar métricas personalizadas do Monitor de Desempenho a serem enviadas para coleta de dados. Para obter instruções, consulte [Como posso usar o CloudWatch agente para visualizar métricas do Monitor de Desempenho em um servidor Windows?](#) no Centro de AWS Conhecimento.

Depois de implantar o CloudWatch agente, você pode configurar as seguintes métricas no arquivo de configuração do agente `emmetrics_collected`:

Categoria métrica	Nome da métrica
Banco de dados para instâncias (NTDSA)	% de impacto no cache do banco de dados
Latência média de leitura do banco de dados de E/S	
Leituras do banco de dados de E/S por segundo	
Latência média de gravação de registros de E/S	
DirectoryServices (NTDS)	Tempo de ligação LDAP
Operações de replicação pendentes do DRA	
Sincronizações de replicação pendentes do DRA	
DNS	Consultas recursivas por segundo
Falha na consulta recursiva por segundo	
Consulta TCP recebida/seg	
Total de consultas recebidas por segundo	
Resposta total enviada/seg	
Consulta UDP recebida/seg	

Categoria métrica	Nome da métrica
LogicalDisk	Tamanho médio da fila de discos
% de espaço livre	
Memória	% de bytes comprometidos em uso
Vida útil (s) média de longo prazo do cache em espera	
Interface de rede	Bytes enviados/seg
Bytes recebidos/segundo	
Largura de banda atual	
NTDS	Atraso estimado na fila do ATQ
Latência da solicitação ATQ	
Leituras do diretório DS por segundo	
Pesquisas de diretório DS/seg	
Gravações de diretórios DS por segundo	
Sessões do cliente LDAP	
Pesquisas LDAP/seg	
Ligações LDAP bem-sucedidas por segundo	
Processador	% de tempo de processamento
Estatísticas de todo o sistema de segurança	Autenticações Kerberos
Autenticações NTLM	

Recursos adicionais

- [Serviços de domínio do Active Directory em AWS: Guia de implantação da solução de parceiro](#) (AWS documentação)
- [Planejamento de capacidade para os Serviços de Domínio do Active Directory](#) (documentação da Microsoft)
- [Considerações de design para executar o Active Directory em instâncias do EC2](#) (AWS whitepapers)

AWS Managed Microsoft AD

Visão geral

AWS Directory Service for Microsoft Active Directory, também conhecido como AWS Managed Microsoft AD, é alimentado por um Active Directory do Windows Server e gerenciado por AWS. Você pode usar AWS Managed Microsoft AD para migrar uma ampla variedade de aplicativos compatíveis com o Active Directory para o. Nuvem AWS AWS Managed Microsoft AD funciona com uma variedade de aplicativos e serviços nativos do Active Directory. Ele também oferece suporte a [aplicativos e serviços AWS gerenciados](#). Embora não existam muitas alavancas de otimização de custos AWS Managed Microsoft AD devido ao serviço e seus mecanismos de cobrança, existem alguns princípios de design que podem ajudar você a manter os custos no mínimo.

Impacto nos custos

Como AWS Managed Microsoft AD é um serviço gerenciado baseado nos SKUs atuais, o dimensionamento é um processo relativamente simples. Atualmente, existem dois SKUs de dimensionamento disponíveis: edições Standard e Enterprise. Outros SKUs incluem compartilhamento de diretórios, adição de controladores de domínio adicionais (incluindo regiões adicionais) e transferência de dados entre regiões.


Recomendações de otimização de custos

Há diferenças entre a AWS Managed Microsoft AD Standard Edition e a AWS Managed Microsoft AD Enterprise Edition. A Enterprise Edition suporta até 500.000 objetos do Active Directory, 125 compartilhamentos de conta (limite flexível) e tem suporte a várias regiões. A Standard Edition suporta até 30.000 objetos do Active Directory, cinco compartilhamentos de conta (limite flexível de aproximadamente 30 no máximo) e não tem suporte para várias regiões.

As questões a serem consideradas antes de selecionar seu tipo de diretório são:

- O suporte multirregional é necessário?
- O diretório será compartilhado com mais de 30 contas?
- A contagem de objetos do Active Directory será superior a 30.000?

Se a resposta for sim para qualquer uma das perguntas acima, a Enterprise Edition é necessária. Se a resposta para todas as perguntas for não, recomendamos que você comece com a Standard Edition.

 Note

Você pode atualizar um diretório da Standard Edition para a Enterprise Edition, mas um diretório não pode ser rebaixado. Implantar a Standard Edition não é passar por uma porta de mão única. Se você deseja atualizar seu diretório para a Enterprise Edition, entre em contato com AWS.

Há um custo para cada compartilhamento quando você compartilha diretórios na AWS Managed Microsoft AD Enterprise Edition. Isso é menor do que o custo de implantar um diretório em cada conta, mas lembre-se de que os custos de compartilhamento podem aumentar se não forem verificados. Recomendamos que você compartilhe somente diretórios com contas que contenham Amazon Relational Database Service (Amazon RDS) e Amazon FSx for Windows File Server, já que somente esses serviços oferecem suporte a esse recurso. Lembre-se de que você tem a opção de integrar o FSx for Windows File Server com seu Active Directory autogerenciado, incluindo um. AWS Managed Microsoft AD Se somente o Amazon FSx for necessário em outra conta, você poderá fazer uma implantação autogerenciada do Amazon FSx na sem AWS Managed Microsoft AD a necessidade de compartilhar o diretório.

Ao decidir quando implantar controladores de domínio adicionais, lembre-se de que ele AWS Managed Microsoft AD oferece suporte a apenas duas sub-redes em zonas de disponibilidade separadas na mesma VPC. Adicionar mais controladores de domínio não permite que você adicione mais sub-redes. Para determinar se você deve adicionar outros controladores de domínio devido a problemas de desempenho, revise as [métricas de desempenho do controlador de domínio em CloudWatch](#). Isso informa se um ou todos os controladores de domínio estão sobrecarregados. Se você determinar que apenas um controlador de domínio está sobrecarregado, adicionar outros controladores de domínio não aliviará a carga e você precisará se aprofundar nos aplicativos que

não fazem o balanceamento de carga nos controladores de domínio atualmente disponíveis. Se todos os controladores de domínio estiverem sendo muito usados, adicionar um controlador de domínio adicional poderá reduzir a carga dos controladores de domínio existentes. Para obter instruções sobre como automatizar a escalabilidade, consulte [Como automatizar a AWS Managed Microsoft AD escalabilidade com base nas métricas de utilização no blog de segurança](#). AWS

Se você estendeu seu diretório para várias regiões, recomendamos que você não use os compartilhamentos NETLOGON ou SYSVOL do diretório para armazenamento de arquivos. Todos os controladores de domínio replicam o conteúdo desses compartilhamentos. Não usar os compartilhamentos para armazenamento de arquivos reduz os custos de transferência de dados ao mínimo.

Você também tem a opção de se inscrever em um Contrato Enterprise com AWS. Os contratos corporativos oferecem a opção de personalizar os contratos que melhor atendam às suas necessidades. Para obter mais informações, consulte [Clientes corporativos](#).

Recursos adicionais

- [AWS Managed Microsoft AD cotas](#) (AWS Directory Service documentação)
- [AWS Directory Service Preços](#) (AWS site)
- [Serviços de domínio do Active Directory em AWS](#) (AWS whitepapers)

AD Connector

Visão geral

O [AD Connector](#) é um serviço de proxy que fornece uma maneira fácil de conectar seu Microsoft Active Directory local existente a [AWS aplicativos](#) compatíveis, como Amazon WorkSpaces QuickSight, Amazon e união perfeita de domínios para instâncias do Amazon Elastic Compute Cloud (Amazon EC2), sem armazenar em cache nenhuma informação na nuvem. Você pode usar o AD Connector para adicionar uma conta de serviço ao seu Active Directory. O AD Connector elimina a necessidade de sincronização de diretórios ou o custo e a complexidade de hospedar uma infraestrutura de federação. Embora não haja muitas alavancas de otimização de custos para o AD Connector devido à natureza do serviço e seus mecanismos de cobrança, você pode seguir as recomendações de design nesta seção para reduzir os custos ao mínimo.

Impacto nos custos

O AD Connector é um serviço gerenciado baseado em SKUs predefinidos. Isso torna o dimensionamento um processo simples. Há dois SKUs de dimensionamento disponíveis: tamanhos pequenos e grandes. Você pode usar o [AWS Pricing Calculator](#) para estimativas de custo envolvendo o AD Connector.

Recomendações de otimização de custos

Além dos recursos de computação de back-end, não há diferença entre os tamanhos de conectores pequenos e grandes.

As questões a serem consideradas antes de selecionar seu tipo de diretório são:

- Há um grande número (mais de 10.000) de usuários ativos usando AWS aplicativos integrados ao AD Connector?
- O usuário é membro de muitos grupos aninhados profundos ou circulares?

Se a resposta para ambas as perguntas for não, recomendamos que você comece com o tamanho pequeno. Se você responder sim a alguma das perguntas acima, talvez valha a pena considerar um tamanho grande. Você pode começar com um AD Connector de tamanho pequeno e, se o diretório ficar prejudicado devido ao desempenho, você pode solicitar que o diretório seja atualizado para um tamanho grande.

Note

Você pode atualizar um AD Connector de pequeno para grande, mas um AD Connector não pode ser rebaixado.

A maioria dos problemas de desempenho não está relacionada ao AD Connector, mas aos controladores de domínio do Active Directory locais que estão sobrecarregados devido ao fato de muitos usuários serem membros de muitos grupos aninhados profundos ou circulares.

Você também tem a opção de se inscrever em um Contrato Enterprise com AWS. Os contratos corporativos oferecem a opção de personalizar os contratos que melhor atendam às suas necessidades. Para obter mais informações, consulte [Clientes corporativos](#).

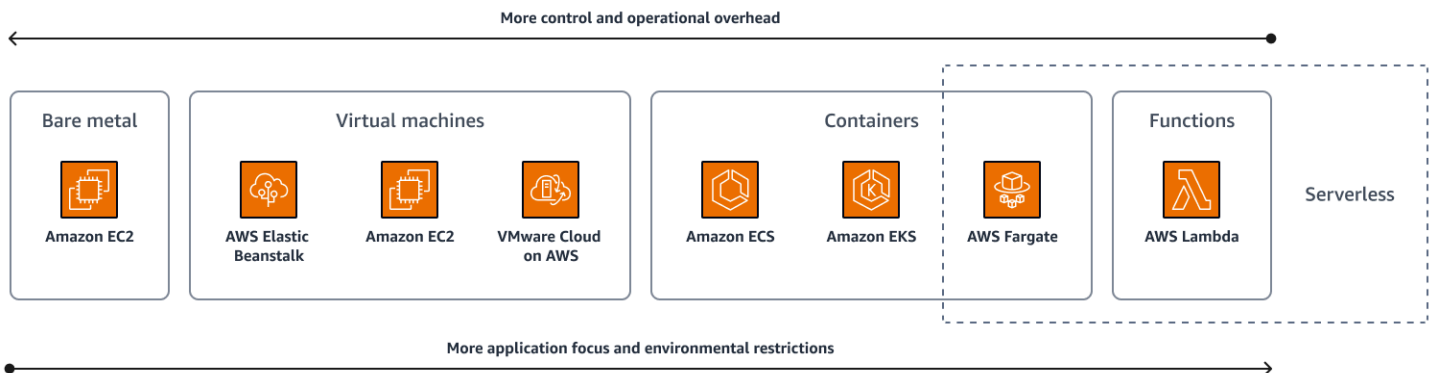
Recursos adicionais

- [Cotas do AD Connector](#) (AWS Directory Service documentação)
- [Preços de outros tipos de diretório](#) (AWS site)
- [Serviços de domínio do Active Directory em AWS](#) (AWS whitepapers)

.NET

Desenvolvimento e implantação. NETos aplicativos são uma chave importante para ajudar você a alcançar a escala e a agilidade oferecidas pela computação em nuvem. Para muitos legados. NETEm aplicativos, a opção computacional mais adequada para executar aplicativos AWS é o uso de máquinas virtuais, seja por meio do AWS Elastic Beanstalk Amazon Elastic Compute Cloud (AmazonEC2). Também é possível correr. NETaplicativos em contêineres Windows e Linux.

A introdução do. NETO núcleo permite que você crie um design moderno. NETaplicativos que aproveitam todos os benefícios da nuvem. Os aplicativos modernos podem usar o conjunto tradicional de opções de computação e também ter como alvo vários tipos de ambientes sem servidor, incluindo ou. AWS Fargate AWS Lambda. NETO 6+ agora oferece hospedagem de alto desempenho de cargas de trabalho em ARM64 EC2 instâncias como as famílias Graviton2. EC2 Isso permite o acesso à última geração de processadores disponíveis na AmazonEC2. Isso significa que seus aplicativos podem ser hospedados em computação especializada para seu tipo de carga de trabalho, como codificação de vídeo, servidores web e computação de alto desempenho (). HPC



Esta seção fornece recomendações para ajudar você a adaptar seu. NETaplicativos para aproveitar os benefícios da nuvem com foco na eficiência de custos.

Esta seção abrange os seguintes tópicos:

- [Refatore para o moderno. NETe mude para o Linux](#)
- [Contentorizar. NETaplicativos](#)
- [Use instâncias e contêineres do Graviton](#)
- [Support a escalabilidade dinâmica para estática. NETAplicativos de estrutura](#)
- [Use o armazenamento em cache para reduzir a demanda do banco de dados](#)
- [Considere a opção sem servidor. NET](#)

- [Considere bancos de dados criados especificamente](#)

Refatore para o moderno. NETe mude para o Linux

Visão geral

Modernizando o legado. NETOs aplicativos de estrutura podem ajudar você a melhorar a segurança, o desempenho e a escalabilidade. Uma forma eficaz de se modernizar. NETOs aplicativos de estrutura são migrá-los para um moderno. NETversão (6+). Aqui estão alguns dos principais benefícios de migrar esses aplicativos para o código aberto. NET:

- Para reduzir os custos de licenciamento do Windows executando-os em um sistema operacional Linux
- Aproveite a disponibilidade de idiomas modernos
- Obtenha desempenho otimizado para execução em Linux

Muitas organizações ainda estão executando versões mais antigas do. NETEstrutura. Isso pode representar riscos de segurança, já que as vulnerabilidades nas versões mais antigas não são mais tratadas pela Microsoft. A Microsoft encerrou o suporte para versões recentes do. NETEstrutura 4.5.2, 4.6 e 4.6.1. É muito importante avaliar os riscos e benefícios de continuar executando as versões mais antigas da estrutura. Para reduzir riscos e reduzir custos, pode valer a pena investir tempo e esforço na refatoração para uma versão moderna do. NET.

Impacto nos custos

Considere um tipo de EC2 instância de uso geral (m5), que ofereça um equilíbrio entre recursos de computação, memória e rede. Essas instâncias são adequadas para uma variedade de aplicativos, como servidores web, bancos de dados de médio porte e repositórios de código-fonte.

Por exemplo, uma instância m5.xlarge sob demanda com 4 vCPUs e 16 GB de memória no Windows Server (licença incluída) no Leste dos EUA (Norte da Virgínia) custa 274,48 USD mensais. Os mesmos recursos em um servidor Linux custam \$140,16 mensais. Neste exemplo, há uma redução de 49% no custo quando você migra seu aplicativo do. NETEstrutura para uma versão moderna do. NETe execute seu aplicativo em um servidor Linux. Seu custo pode variar dependendo das opções (por exemplo, tipo de instância, sistema operacional, armazenamento) que você escolher ao selecionar uma [EC2instância](#). Você pode otimizar ainda mais os custos usando [Savings Plans](#) ou [Reserved Instances](#). Para obter mais detalhes, use o [AWS Pricing Calculator](#) para executar

estimativas de custo. Para instâncias incluídas no Windows, o custo da licença é de [0,046 USD por v CPU por hora](#), independentemente do modelo de preços.

Portando esses. NETFramework de aplicativos modernos. NETexige esforço do desenvolvedor. Você deve avaliar seus aplicativos e suas dependências para ver se eles são compatíveis com a versão da plataforma de destino. [AWS Assistente de portabilidade para. NET](#) é uma ferramenta auxiliar que escaneia. NETFramework aplica e gera um. NETavaliação de compatibilidade, ajudando você a portar seus aplicativos para serem compatíveis com o Linux mais rapidamente. O assistente de portabilidade para. NETidentifica incompatibilidades com. NET, encontra substitutos conhecidos e gera uma avaliação detalhada da compatibilidade. Depois de portar sua solução, você deve fazer alterações manuais no código para que seu projeto seja compilado com êxito com dependências. Isso reduz o esforço manual envolvido na modernização de seus aplicativos para Linux. Se seu aplicativo oferece suporte a ARM processadores, migrar para o Linux libera a capacidade de usar instâncias do Graviton. Isso pode ajudá-lo a obter mais 20% em reduções adicionais de custos. Para obter mais informações, consulte [Powering. NET5 com AWS Graviton2: Benchmarks](#) in the AWS Compute Blog.

Existem outras ferramentas, como o [AWS Toolkit for. NETRefatoração e o. NETAssistente de atualização](#), que pode ajudá-lo a portar o legado. NETaplicativos de estrutura para modernos. NET.

Recomendações de otimização de custos

Para migrar. NETAplicativos de estrutura, faça o seguinte:

1. Pré-requisitos — Para usar o Porting Assistant para. NET, você deve instalar. NET5+ na máquina em que você planeja analisar o código-fonte do aplicativo. Os recursos na máquina devem ter no mínimo 1,8 velocidade de GHz processamento, 4 GB de memória e 5 Gb de espaço de armazenamento. Para obter mais informações, consulte [Pré-requisitos](#) no Assistente de portabilidade para. NETdocumentação.
2. Avaliação — Baixe o Porting Assistant para. NETcomo um [arquivo executável](#) (download). Você pode baixar e instalar a ferramenta em sua máquina para iniciar a avaliação de seus aplicativos. A página de avaliação contém projetos e pacotes portados APIs que são incompatíveis com os modernos. NET. Por esse motivo, você recebe erros de compilação na solução após a avaliação. Você pode visualizar ou baixar os resultados da avaliação em um CSV arquivo. Para obter mais informações, consulte [Portar uma solução](#) no Assistente de portabilidade para. NETdocumentação.
3. Refatoração — Depois de avaliar o aplicativo, você pode portar seus projetos para a versão da estrutura de destino. Ao portar uma solução, seus arquivos de projeto e parte do código serão

modificados pelo Assistente de portabilidade. Você pode verificar os registros para analisar as alterações em seu código-fonte. Na maioria dos casos, o código exigirá um esforço adicional para concluir a migração e o teste para deixá-lo pronto para produção. Dependendo do aplicativo, algumas das alterações podem incluir estrutura de entidades, identidade e autenticação. Para obter mais informações, consulte [Portar uma solução](#) no Assistente de portabilidade para .NET documentação.

Essa é a primeira etapa para modernizar seus aplicativos em contêineres. Pode haver vários fatores comerciais e técnicos para modernizar seu .NET aplicativos de estrutura para contêineres Linux. Um dos fatores mais importantes é reduzir o custo total de propriedade ao mudar de um sistema operacional Windows para o Linux. Isso reduz os custos de licenciamento ao migrar seu aplicativo para uma versão multiplataforma do .NET aos contêineres para otimizar a utilização dos recursos.

Depois que seu aplicativo for portado para o Linux, você poderá usá-lo [AWS App2Container](#) para containerizar seu aplicativo. O App2Container usa a ECS Amazon ou a Amazon EKS como serviços de endpoint nos quais você pode implantar diretamente. O App2Container fornece toda a infraestrutura necessária como artefatos de implantação de código (IaC) para containerizar seus aplicativos repetidamente.

Considerações e recursos adicionais

- Se você tiver aplicativos desenvolvidos em VB. NET (uma estrutura legada de 2002) e quero portá-los para o .NET6, consulte o [Port legacy VB. NET aplicativos para .NET6.0 com o Porting Assistant para .NET](#) publicação no AWS blog Microsoft Workloads on.
- Se você tiver aplicativos herdados no Windows Communication Foundation (WCF) e quiser executá-los em um ambiente moderno. NET, você pode adotar o CoreWCF. Para obter mais informações, consulte [Modernizando WCF aplicativos legados para o Core WCF usando o Porting Assistant for .NET](#) publicação no AWS blog Microsoft Workloads on.
- Você pode adicionar o assistente de portabilidade como uma extensão ao Visual Studio IDE. Isso permite que você execute todas as tarefas necessárias para converter seu código sem precisar alternar entre o seu IDE e o Assistente de portabilidade para .NET ferramenta. Para obter mais informações, consulte o [Accelerate .NET modernização de aplicativos com o Porting Assistant for .NET](#) postagem sobre a IDE extensão do Visual Studio no AWS blog Microsoft Workloads on.
- [AWS Assistente de portabilidade para .NET agora é uma ferramenta de código aberto](#) com o código-fonte e os componentes de análise de compatibilidade da avaliação. Isso pode incentivar

seus desenvolvedores a usar e compartilhar. NET transportando conhecimento e melhores práticas.

- Você pode portar. NET aplicativos de estrutura para modernos. NET no Linux usando o AWS kit de ferramentas para. NET refatoração. Para obter mais informações, consulte o [Accelerate. NET modernização com o AWS Toolkit for. NET](#) Postagem de refatoração no blog Microsoft Workloads on. AWS
- Você pode [acelerar a containerização e a migração do. ASP NET](#) Principais aplicativos a serem [AWS usados AWS App2Container](#).

Contentorizar. NET aplicativos

Visão geral

Os contêineres são uma forma leve e eficiente de empacotar e implantar aplicativos de maneira consistente e reproduzível. Esta seção explica como você pode usar AWS Fargate um serviço de contêiner sem servidor para reduzir os custos do seu. NET aplicativos e, ao mesmo tempo, fornece infraestrutura escalável e confiável.

Impacto nos custos

Alguns fatores que influenciam a eficácia do uso de contêineres para redução de custos incluem o tamanho e a complexidade do aplicativo, o número de aplicativos que precisam ser implantados e o nível de tráfego e demanda dos aplicativos. Para aplicativos pequenos ou simples, os contêineres podem não oferecer economias de custo significativas em comparação com as abordagens tradicionais de infraestrutura, pois a sobrecarga do gerenciamento dos contêineres e dos serviços associados pode, na verdade, aumentar os custos. No entanto, para aplicativos maiores ou mais complexos, o uso de contêineres pode proporcionar economia de custos, melhorando a utilização de recursos e reduzindo o número de instâncias necessárias.

Recomendamos que você tenha em mente o seguinte ao usar contêineres para reduzir custos:

- Tamanho e complexidade do aplicativo — Aplicativos maiores e mais complexos são mais adequados para a containerização porque tendem a exigir mais recursos e podem se beneficiar mais com a melhor utilização dos recursos.
- Número de aplicativos — Quanto mais aplicativos sua organização precisar implantar, maior será a economia de custos por meio da containerização.

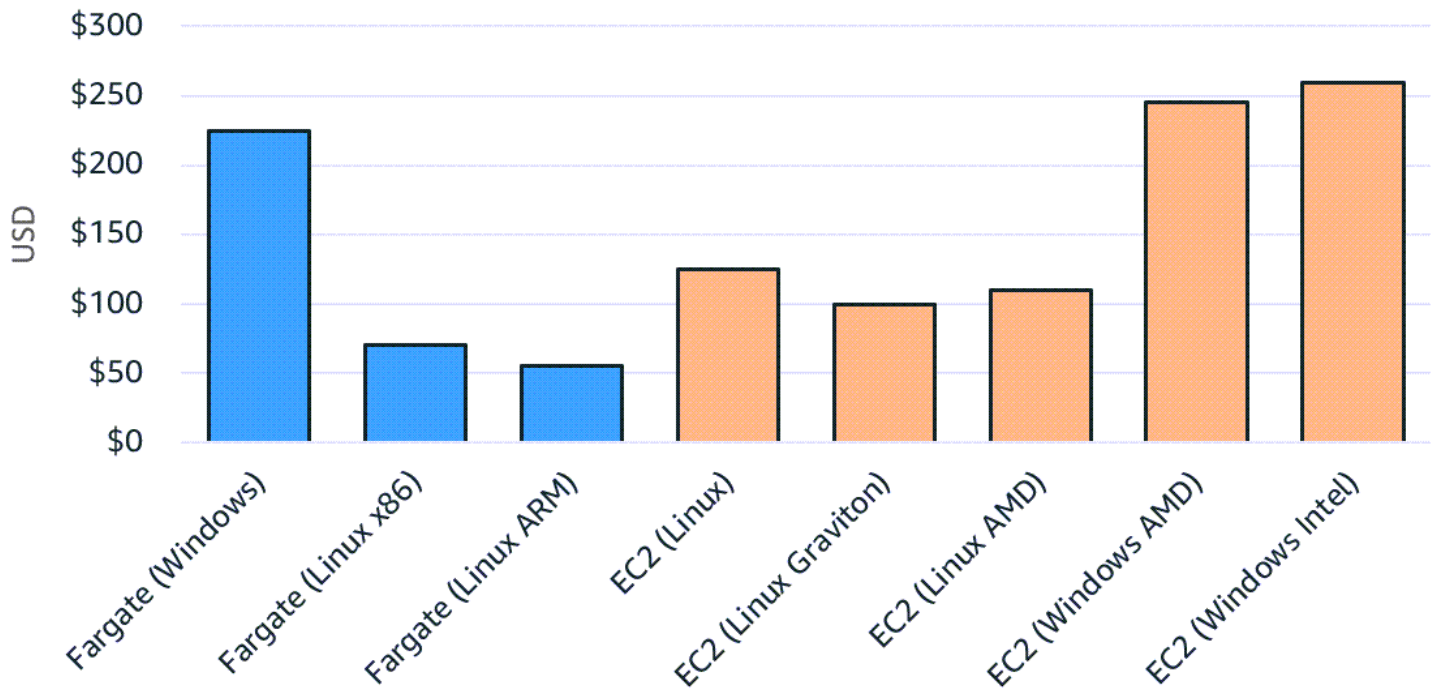
- Tráfego e demanda — Os aplicativos que experimentam alto tráfego e demanda podem se beneficiar da escalabilidade e elasticidade que os contêineres oferecem. Isso pode levar à economia de custos.

Arquiteturas e sistemas operacionais diferentes afetam os custos dos contêineres. Se você estiver usando contêineres do Windows, os custos podem não diminuir devido a considerações de licenciamento. Os custos de licenciamento são menores ou ausentes com os contêineres Linux. O gráfico a seguir usa uma configuração básica AWS Fargate na região Leste dos EUA (Ohio) com as seguintes configurações: 30 tarefas por mês, cada uma sendo executada por 12 horas com 4 vCPUs e 8 GB de memória alocada.

Você pode escolher entre duas plataformas de computação principais para executar seus contêineres AWS: [hosts de contêineres EC2 baseados e sem servidor](#) ou [AWS Fargate](#). Se você usa o Amazon Elastic Container Service (AmazonECS) em vez do Fargate, deve manter a computação em execução (instâncias) para permitir que o mecanismo de posicionamento instancie contêineres quando necessário. Se você usar o Fargate em vez disso, somente a capacidade computacional necessária será provisionada.

O gráfico a seguir mostra a diferença entre contêineres equivalentes usando o Fargate e o Amazon EC2. Devido à flexibilidade do Fargate, as tarefas de um aplicativo podem ser executadas 12 horas por dia, sem utilização fora do horário de expediente. No entanto, para a AmazonECS, você deve controlar a capacidade computacional usando um grupo EC2 de instâncias do [Auto Scaling](#). Isso pode fazer com que a capacidade funcione 24 horas por dia, o que pode, em última análise, aumentar os custos.

Monthly costs of Fargate and Amazon EC2



Recomendações de otimização de custos

Use contêineres Linux em vez de Windows

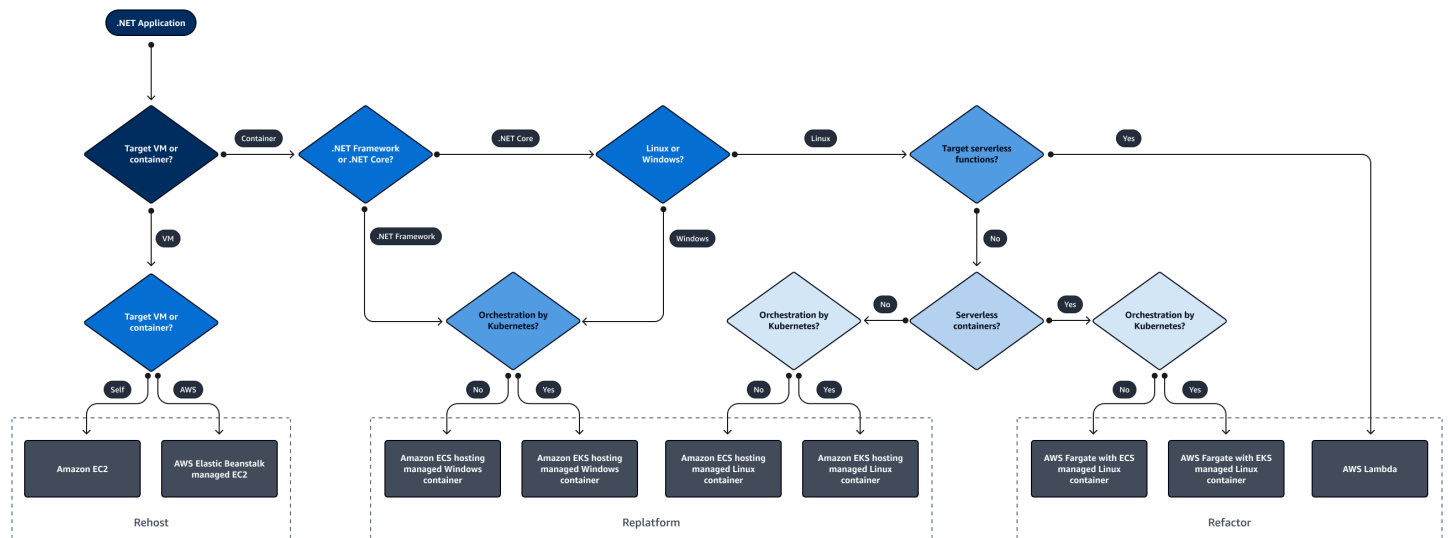
Você pode obter economias significativas se usar contêineres Linux em vez de contêineres Windows. Por exemplo, você pode obter uma economia de aproximadamente 45% nos custos de computação se executar o .NETCore no EC2 Linux em vez de executar o .NETEstrutura no EC2 Windows. Você pode obter uma economia adicional de 40% se usar a ARM arquitetura (AWS Graviton) em vez da x86.

Se você planeja executar contêineres baseados em Linux para os existentes. NETAplicativos de estrutura, você deve portar esses aplicativos para versões modernas e multiplataforma do .NET([como .NET6.0](#)) para usar contêineres Linux. Uma consideração importante é pesar o custo da refatoração em comparação com a economia de custos obtida com a redução do custo dos contêineres Linux. Para obter mais informações sobre como portar seus aplicativos para aplicativos modernos. NET, consulte [Assistente de portabilidade para .NET](#) na AWS documentação.

Outro benefício de mudar para o moderno. NET(ou seja, longe do .NETEstrutura) é que oportunidades adicionais de modernização se tornem disponíveis. Por exemplo, você pode

considerar a reestruturação de seu aplicativo para uma arquitetura baseada em microsserviços que seja mais escalável, ágil e econômica.

O diagrama a seguir ilustra o processo de tomada de decisão para explorar oportunidades de modernização.



Aproveite as vantagens dos Savings Plans

Os contêineres podem ajudar você a aproveitar os [Compute Savings](#) Plans para reduzir seus custos com o Fargate. O modelo de desconto flexível oferece os mesmos descontos que as instâncias reservadas conversíveis. O preço do Fargate é baseado nos recursos v CPU e de memória usados desde o momento em que você começa a baixar a imagem do contêiner até o término da ECS tarefa da Amazon (arredondado para o segundo mais próximo). [Os Savings Plans for Fargate](#) oferecem economia de até 50% no uso do Fargate em troca do compromisso de usar uma quantidade específica de uso de computação (medida em dólares por hora) por um período de um ou três anos. Você pode usar [AWS Cost Explorer](#) para ajudá-lo a escolher um Savings Plan.

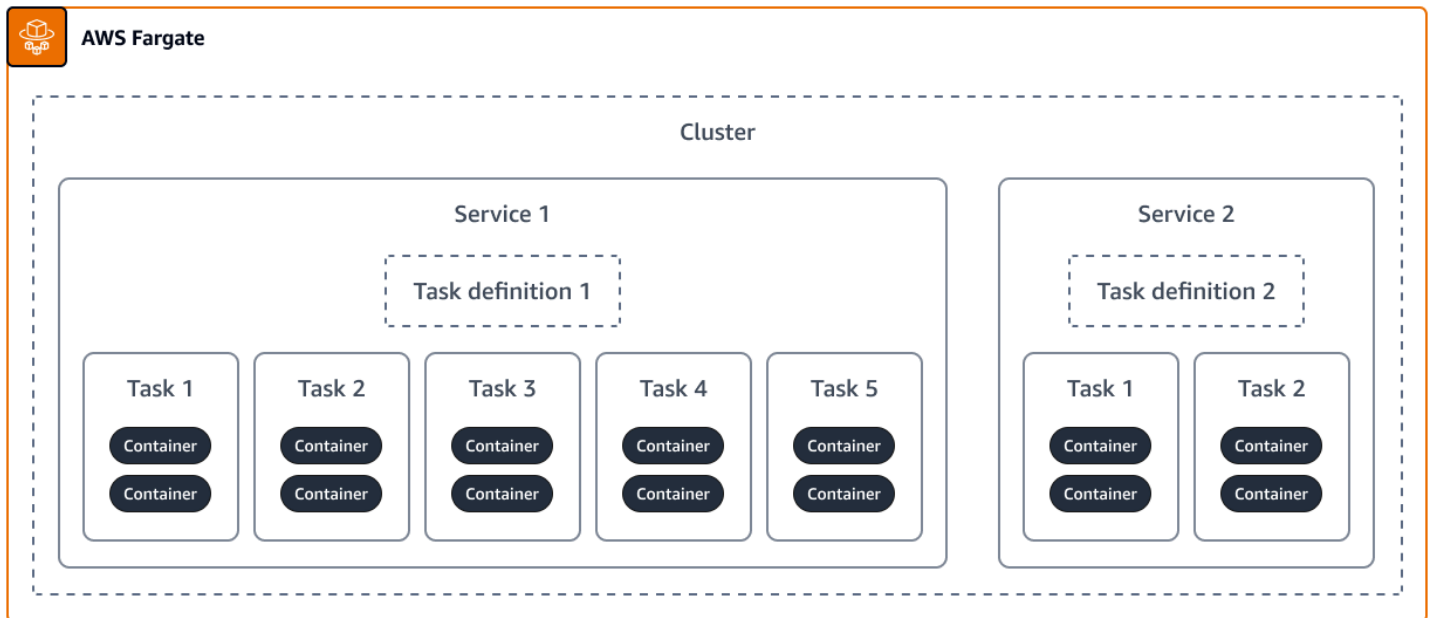
É importante entender que os Compute Savings Plans são aplicados ao uso que gera a maior economia em primeiro lugar. Por exemplo, se você estiver executando uma instância Linux t3.medium us-east-2 e uma instância t3.medium idêntica do Windows, a instância Linux receberá primeiro o benefício do Savings Plan. Isso ocorre porque a instância Linux tem um potencial de economia de 50%, enquanto a mesma instância do Windows tem um potencial de economia de 35%. Se você tiver outros recursos elegíveis do Savings Plan em execução no seu Conta da AWS, como Amazon EC2 ou Lambda, não é necessário que seu Savings Plan seja aplicado primeiro ao Fargate. Para obter mais informações, consulte [Entendendo como os Savings Plans se aplicam ao seu AWS](#)

[uso](#) na documentação dos Savings Plans e [na EC2 seção Otimizar gastos para Windows na Amazon](#) deste guia.

Tarefas do Fargate do tamanho certo

É importante garantir que as tarefas do Fargate sejam dimensionadas corretamente para alcançar o grau máximo de otimização de custos. Frequentemente, os desenvolvedores não têm todas as informações de uso necessárias ao determinar inicialmente as configurações das tarefas do Fargate usadas em seus aplicativos. Isso pode levar ao provisionamento excessivo de tarefas e, em seguida, resultar em gastos desnecessários. Para evitar isso, recomendamos que você carregue os aplicativos de teste em execução no Fargate para entender o desempenho de uma configuração de tarefa específica em diferentes cenários de uso. Você pode usar os resultados do teste de carga, vCPU, alocação de memória das tarefas e políticas de escalonamento automático para encontrar o equilíbrio certo entre desempenho e custo.

O diagrama a seguir mostra como o Compute Optimizer gera recomendações para o tamanho ideal da tarefa e do contêiner.



Uma abordagem é usar uma ferramenta de teste de carga, como a descrita em [Distributed Load Testing on AWS](#), para estabelecer uma linha de base para v CPU e utilização de memória. Depois de executar o teste de carga para simular uma carga típica do aplicativo, você pode ajustar a configuração v CPU e a memória da tarefa até que a utilização básica seja alcançada.

Recursos adicionais

- [Lista de verificação de otimização de custos para Amazon ECS e AWS Fargate](#) (postagem no blog de AWS contêineres)
- [Otimização teórica de custos por tipo de ECS lançamento da Amazon: Fargate vs EC2](#) (postagem no blog da AWS Containers)
- [Assistente de portabilidade para .NET](#) (AWS documentação)
- [Teste de carga distribuída em AWS](#) (Biblioteca de AWS soluções)
- [AWS Compute Optimizer lança suporte para ECS serviços da Amazon em AWS Fargate](#) (postagem no blog AWS Cloud Financial Management)

Use instâncias e contêineres do Graviton

Visão geral

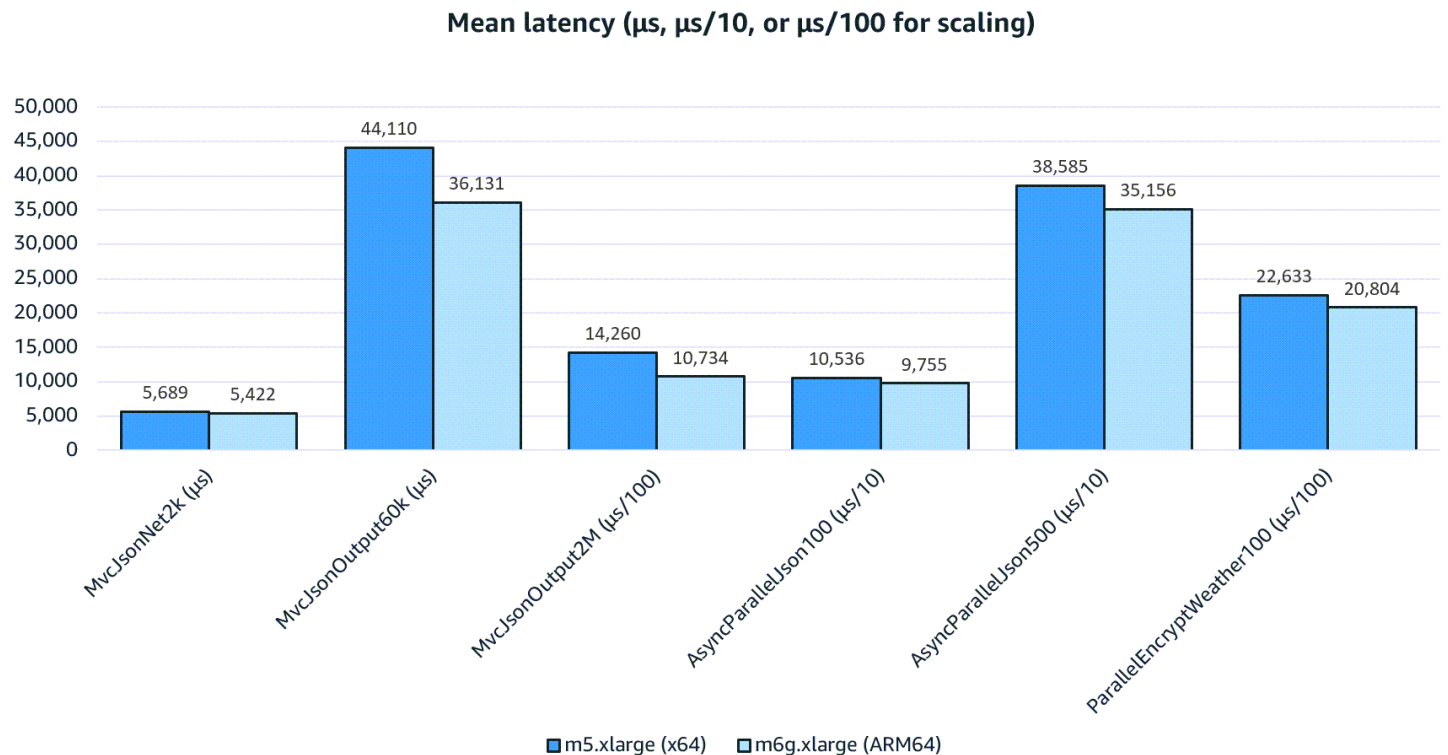
AWS As instâncias do Graviton são alimentadas por ARM processadores projetados AWS para oferecer a melhor relação preço/desempenho para suas cargas de trabalho em nuvem executadas no Amazon Elastic Compute Cloud (AmazonEC2), incluindo contêineres em execução. AWS Atualmente, existem três gerações do Graviton disponíveis para uso na AmazonEC2. Este guia se concentra no uso do Graviton 2 e 3 com .NET aplicativos porque há uma economia significativa quando você usa as versões mais recentes do Graviton. Lembre-se de que as instâncias do Graviton executam somente o sistema operacional Linux. Como resultado, as instâncias Graviton são uma forte oferta para .NET rodando em Linux, mas não são uma opção para o sistema operacional Windows ou antigo .NET aplicativos de estrutura.

O Graviton 3 é 60% mais eficiente do que EC2 instâncias comparáveis, com desempenho até 40% melhor. Este guia se concentra nos benefícios de custo do uso do Graviton, mas é importante observar que o Graviton oferece os benefícios adicionais de melhorias de desempenho e maior sustentabilidade ambiental.

Impacto nos custos

Você pode obter até 45% de economia ao mudar para o Graviton. Depois de refatorar qualquer legado .NET Framework de aplicativos para um moderno .NET versão, você desbloqueia a capacidade de usar instâncias do Graviton. Mudar para Graviton é uma técnica eficaz de otimização de custos para .NET desenvolvedores.

O exemplo na tabela a seguir mostra o potencial de melhorias de desempenho que você pode obter migrando para instâncias do Graviton.



[Para obter uma análise completa e uma explicação da abordagem de benchmarking usada para criar os resultados no diagrama anterior, consulte *Powering .NET5 com AWS Graviton2: Benchmarks in the AWS Compute Blog*.](#)

Uma das razões para a maior eficiência é a diferença no significado de v CPU entre x86 e Graviton. Na arquitetura x86, um v CPU é um núcleo lógico obtido pelo hyperthreading. No Graviton, v CPU equivale a um núcleo físico que permite que o v esteja totalmente comprometido com CPU a carga de trabalho.

O resultado com o Graviton2 é um desempenho de preço 40% melhor do que instâncias x86/x64 comparáveis. Graviton3 oferece o seguinte em relação ao Graviton2:

- Um perfil de desempenho aprimorado com desempenho até 25% melhor
- Desempenho de ponto flutuante até duas vezes maior
- Desempenho da carga de trabalho criptográfica até duas vezes mais rápido
- Desempenho de aprendizado de máquina até três vezes melhor

Além disso, o Graviton3 é a primeira instância na nuvem a apresentar DDR5 memória.

As tabelas a seguir mostram a diferença na economia de custos entre instâncias baseadas em Graviton e instâncias equivalentes baseadas em x86.

Esta tabela mostra uma economia de Graviton de 19,20 por cento.

Tipo de instância	Arquitetura	v CPU	Memória (GB)	Custo por hora (sob demanda)
t4g.xlarge	ARM	4	16	\$0,134
t3.xlarge	x86	4	16	0,164 US\$

Esta tabela mostra uma economia de Graviton de 14,99 por cento.

Tipo de instância	Arquitetura	v CPU	Memória (GB)	Custo por hora (sob demanda)
c7g.4xlarge	ARM	16	32	\$0,5781
c6i.4xlarge	x86	16	32	\$0.6800

É importante testar o perfil de desempenho do seu aplicativo ao considerar o Graviton. O Graviton não substitui práticas sólidas de desenvolvimento de software. Você pode usar os testes para verificar se está aproveitando ao máximo seus recursos computacionais subjacentes.

Recomendações de otimização de custos

Há várias maneiras de aproveitar os processadores/instâncias Graviton. Esta seção mostra as mudanças necessárias para passar do uso de uma máquina com arquitetura x86 para instâncias Graviton (ARM).

Alterar a configuração de tempo de execução no Lambda

Recomendamos que você ative as configurações de tempo de execução AWS Lambda. Para obter mais informações, consulte [Modificação do ambiente de execução na documentação](#) do Lambda.

Desde .NET é uma linguagem compilada, você deve seguir um processo de construção para que isso funcione. Para obter um exemplo de como fazer isso, consulte [.NET em Graviton](#) em GitHub.

Contêineres

Para uma carga de trabalho em contêineres, crie uma imagem de contêiner com várias arquiteturas. Você pode fazer isso especificando várias arquiteturas no comando Docker build. Por exemplo:

```
docker buildx build -t "myImageName:latest" --platform linux/amd64,linux/arm64 --push .
```

Você também pode usar uma ferramenta AWS Cloud Development Kit (AWS CDK) para ajudar a [orquestrar a](#) compilação. Para exemplos do Docker, consulte [Criação de imagens de vários arcos para Arm e x86 com desktops Docker na documentação do Docker](#).

Amazon EC2

Para migrar ARM de x86/x64, defina a ARM arquitetura na etapa de compilação. No Visual Studio, você pode criar um ARM64CPU. Para obter instruções, consulte [Para configurar um projeto para atingir o Arm64 e outras plataformas](#) na documentação da Microsoft.

Se você estiver usando o .NET CLI, em seguida, executar a compilação em uma ARM máquina produz uma compilação compatível com Graviton. Para ver uma demonstração, assista [Accelerate .NET Desempenho 6 com Arm64 no AWS Graviton2 ligado](#). YouTube Problemas de dependência resultarão em erros de tempo de compilação que podem ser resolvidos individualmente. Desde que existam ARM bibliotecas para qualquer dependência, a transição deve ser relativamente simples.

Recursos adicionais

- [Como criar seus contêineres ARM e economizar com instâncias Graviton e Spot na Amazon ECS](#) (AWS blog)
- [AWS Lambda Funções alimentadas pelo processador AWS Graviton2 — Execute suas funções no Arm e obtenha um preço/desempenho até 34% melhor](#) (blog)AWS
- [Migrando AWS Lambda funções para processadores AWS Graviton2 baseados em ARM](#) (blog)AWS
- [Crie e implante .NET aplicativos web para Amazon ECS Clusters AWS Graviton 2 ARM baseados em Graviton 2 usando AWS CDK](#)(blog)AWS
- [Graviton Fast Start — Um novo programa para ajudar a transferir suas cargas de trabalho para o AWS Graviton](#) (blog)AWS

- [Alimentando. NET5 com AWS Graviton2: Benchmarks](#) (blog)AWS

Support a escalabilidade dinâmica para estática. NETAplicativos de estrutura

Visão geral

Um dos principais benefícios de usar a nuvem para aplicativos é a elasticidade, ou seja, a capacidade de ampliar ou reduzir a computação com base na demanda. Isso permite que você pague apenas pela capacidade computacional de que precisa, em vez de provisionar para uso máximo. A Cyber Monday, na qual os varejistas on-line podem obter rapidamente muitas vezes mais tráfego do que o normal (por exemplo, [milhares de por cento em minutos](#)), é um bom exemplo de elasticidade.

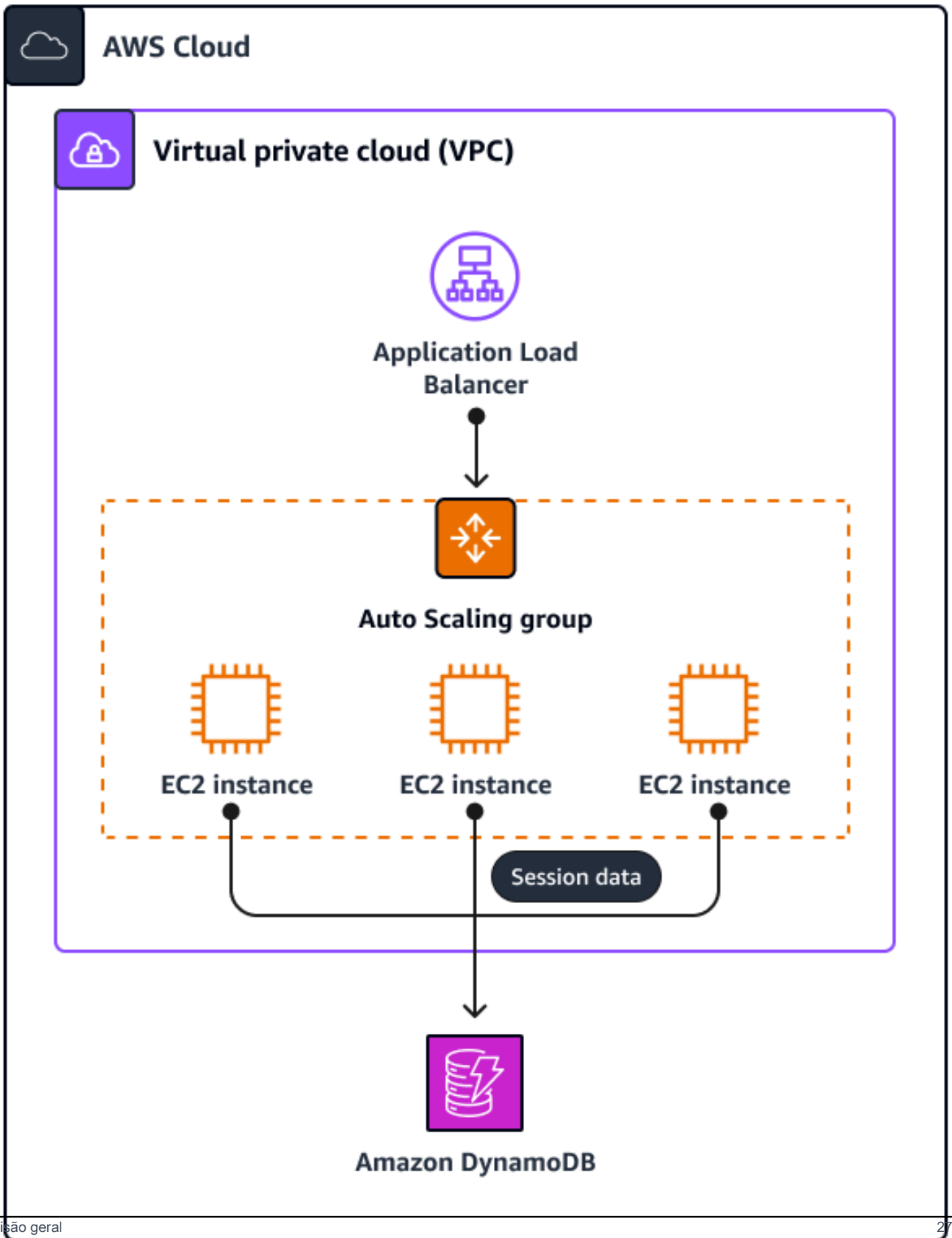
Se você está trazendo um legado. NETaplicativos da web para a nuvem (por exemplo,ASP. NETAplicativos de estrutura em execução emIIS), a capacidade de escalar rapidamente fazendas de servidores com balanceamento de carga pode ser difícil ou impossível devido à natureza estável do aplicativo. Os dados da sessão do usuário são armazenados na memória do aplicativo, geralmente com [ASP. NETestado da sessão](#) ou variáveis estáticas que contêm dados de solicitações cruzadas que devem ser persistidos. A afinidade da sessão do usuário geralmente é mantida por meio de sessões fixas do balanceador de carga.

Isso prova ser um desafio operacional. Quando é necessário aumentar a capacidade, você deve provisionar e adicionar servidores intencionalmente. Isso pode ser um processo lento. Tirar os nós de serviço em caso de aplicação de patches ou falhas inesperadas pode ser problemático para a experiência do usuário final, perdendo o estado de todos os usuários associados aos nós afetados. Na melhor das hipóteses, isso exigiria que os usuários fizessem login novamente.

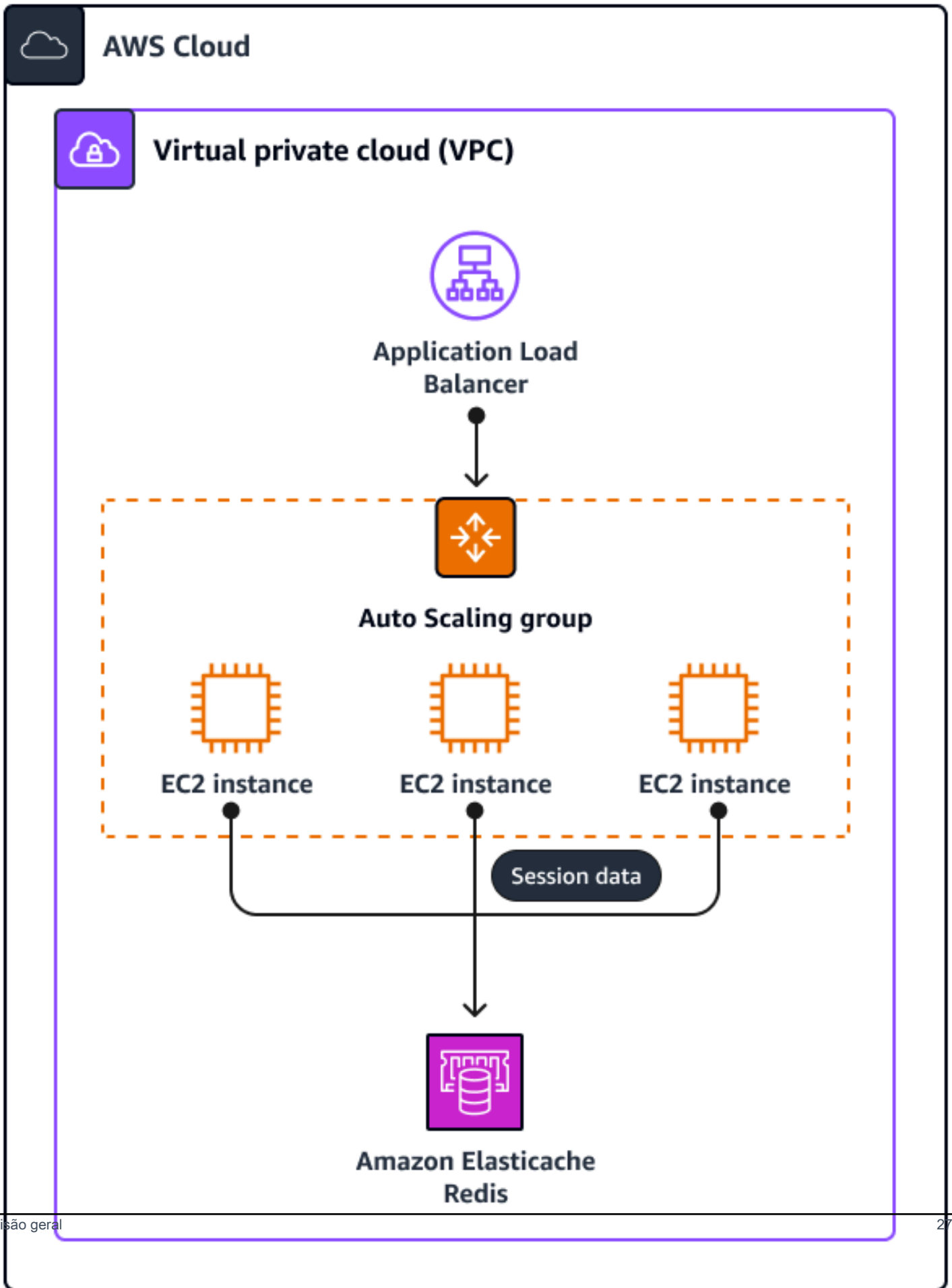
Ao centralizar o estado da sessão paraASP. NETaplicativos e aplicação de regras de escalonamento automático ao legado. ASP NETaplicativos, você pode aproveitar a elasticidade da nuvem e, potencialmente, aproveitar a economia de custos ao executar aplicativos. Por exemplo, você obtém reduções de custo por meio da escalabilidade computacional, mas também pode escolher entre os diferentes modelos de preços disponíveis, como reduzir o [uso de instâncias reservadas e usar os preços da Amazon Spot Instance](#).

Duas técnicas comuns incluem o uso do [Amazon DynamoDB como provedor de estado da sessão e o uso do ElastiCache Amazon \(OSSRedis\) como um. ASP NETarmazenamento de sessões](#).

O diagrama a seguir mostra uma arquitetura que usa o DynamoDB como provedor de estado de sessão.



O diagrama a seguir mostra uma arquitetura que usa ElastiCache (RedisOSS) como provedor de estado de sessão.



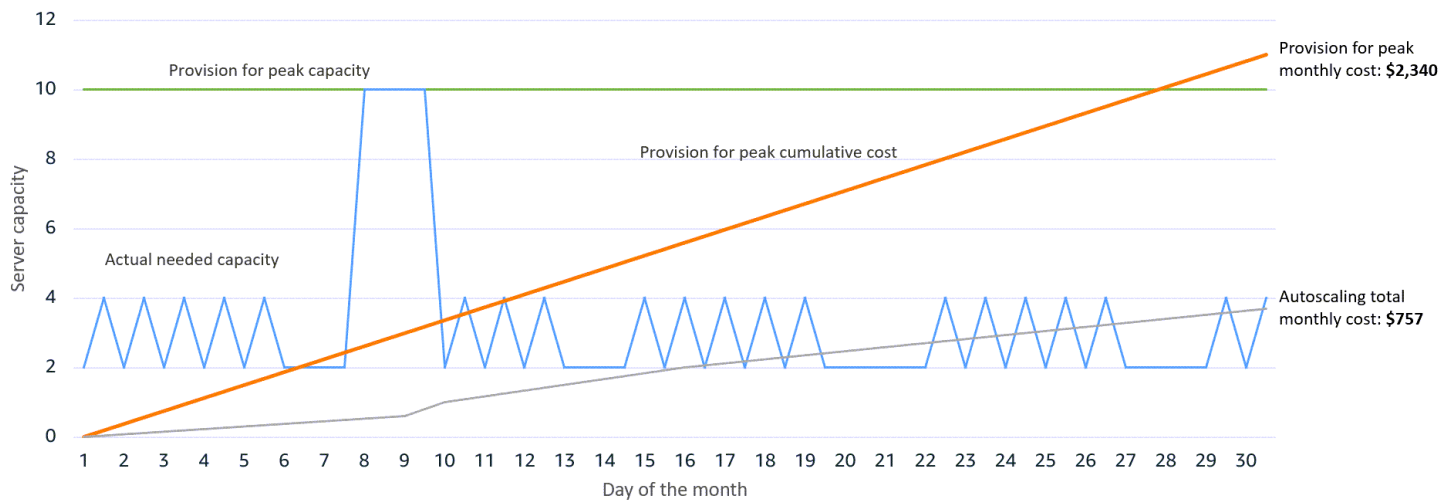
Impacto nos custos

Para determinar os benefícios do dimensionamento para um aplicativo de produção, recomendamos que você modele sua demanda real. Esta seção faz as seguintes suposições para modelar um aplicativo de amostra:

- As instâncias adicionadas e removidas da rotação são idênticas e nenhuma variação no tamanho da instância é introduzida.
- A utilização do servidor nunca cai abaixo de dois servidores ativos para manter a alta disponibilidade do aplicativo.
- A quantidade de servidores é dimensionada linearmente com o tráfego (ou seja, o dobro do tráfego exigirá o dobro da computação).
- O tráfego é modelado ao longo de um mês em incrementos de seis horas, com variação intradiária e um pico anormal de tráfego (por exemplo, uma venda promocional) em um dia de tráfego 10 vezes maior. O tráfego de fim de semana é modelado com base na utilização básica.
- O tráfego noturno é modelado com base na utilização básica, enquanto o tráfego durante a semana é modelado com uma utilização 4x.
- Os preços da Instância Reservada usam preços de um ano, sem adiantamento. O preço diurno normal usa preços sob demanda, enquanto o burst demand usa o preço de instância spot.

O diagrama a seguir ilustra como esse modelo tira proveito da elasticidade em um .NET aplicativo em vez de provisionamento para uso máximo. Isso resulta em uma economia de aproximadamente 68 por cento.

Comparison of cumulative costs for peak provisioning and autoscaling



Se você usa o DynamoDB como um mecanismo de armazenamento do estado da sessão, use os seguintes parâmetros:

```
Storage: 20GB
Session Reads: 40 million
Session Writes: 20 million
Pricing Model: On demand
```

O custo mensal estimado desse serviço é de aproximadamente \$35,00 por mês.

Se você usar ElastiCache (RedisOSS) como um mecanismo de armazenamento do estado da sessão, use os seguintes parâmetros:

```
Number of Nodes: 3
Node size: cache.t4g.medium
Pricing Model: 1y reserved
```

O custo mensal estimado desse serviço é de aproximadamente \$91,00 por mês.

Recomendações de otimização de custos

A primeira etapa é implementar o estado da sessão em um legado. NETaplicativo. Se você estiver usando ElastiCache como seu mecanismo de armazenamento de estado, siga as orientações de [O que é AWS SDK for .NET](#) na AWS SDK for .NET documentação. [Se você estiver usando o](#)

[DynamoDB, siga as orientações ElastiCache de como. ASP NET Armazenamento de sessões](#) no blog de ferramentas para AWS desenvolvedores.

Se o aplicativo usar a InProcessão para começar, certifique-se de que todos os objetos que você planeja armazenar na sessão possam ser serializados. Para fazer isso, use o `SerializableAttribute` atributo para decorar classes cujas instâncias serão armazenadas na sessão. Por exemplo:

```
[Serializable()]
public class TestSimpleObject {
    public string SessionProperty {get;set;}
}
```

Além disso, a `NETMachineKey` deve ser o mesmo entre todos os servidores em uso. Normalmente, esse é o caso quando as instâncias são criadas a partir de uma Amazon Machine Image comum (AMI). Por exemplo:

```
<machineKey
    validationKey="some long hashed value"
    decryptionKey="another long hashed value"
    validation="SHA1"/>
```

No entanto, é importante garantir que, se uma imagem base for alterada, ela seja configurada com a mesma. NETimagem da máquina (configurável no nível do servidor IIS ou no nível do servidor). Para obter mais informações, consulte [SystemWebSectionGroup. MachineKey Propriedade](#) na documentação da Microsoft.

Finalmente, você deve determinar o mecanismo para adicionar servidores a um grupo de Auto Scaling em resposta a um evento de escalabilidade. Há várias maneiras de fazer isso. Recomendamos que os métodos a seguir sejam implantados sem problemas. NETAplicativos de estrutura para uma EC2 instância em um grupo de Auto Scaling:

- Use o [EC2Image Builder](#) para configurar um AMI que contenha o servidor e o aplicativo totalmente configurados. Em seguida, você pode usar esse AMI para configurar o [modelo de lançamento do seu grupo de Auto Scaling](#).
- Use [AWS CodeDeploy](#) para implantar seu aplicativo. CodeDeploy permite a integração diretamente com o [Amazon EC2 Auto Scaling](#). Isso fornece uma alternativa à criação de um novo AMI para cada versão do aplicativo.

Recursos adicionais

- [Crie imagens com o EC2 Image Builder](#) (documentação do EC2 Image Builder)
- [Implantação. NETAplicativos Web usados AWS CodeDeploy com o Visual Studio Team Services](#) (blog de ferramentas para AWS desenvolvedores)

Use o armazenamento em cache para reduzir a demanda do banco de dados

Visão geral

Você pode usar o armazenamento em cache como uma estratégia eficaz para ajudar a reduzir os custos do seu. NETaplicações. Muitos aplicativos usam bancos de dados de back-end, como o SQL Server, quando os aplicativos exigem acesso frequente aos dados. O custo de manter esses serviços de back-end para lidar com a demanda pode ser alto, mas você pode usar uma estratégia eficaz de armazenamento em cache para reduzir a carga nos bancos de dados de back-end, reduzindo os requisitos de dimensionamento e escalabilidade. Isso pode ajudar você a reduzir custos e melhorar o desempenho de seus aplicativos.

O armazenamento em cache é uma técnica útil para economizar nos custos relacionados à leitura de cargas de trabalho pesadas que usam recursos mais caros, como SQL o Server. É importante usar a técnica certa para sua carga de trabalho. Por exemplo, o armazenamento em cache local não é escalável e exige que você mantenha um cache local para cada instância de um aplicativo. Você deve avaliar o impacto no desempenho em comparação com os custos potenciais, para que o menor custo da fonte de dados subjacente compense quaisquer custos adicionais relacionados ao mecanismo de armazenamento em cache.

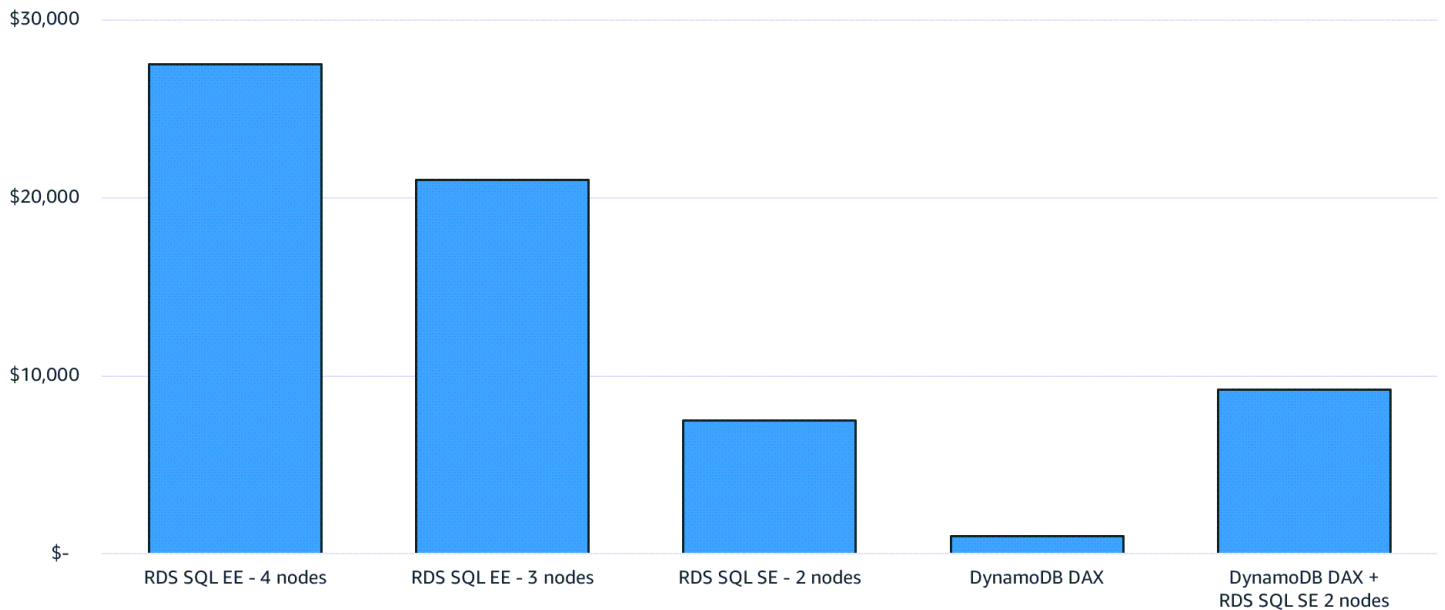
Impacto nos custos

SQLO servidor exige que você leve em consideração as solicitações de leitura ao dimensionar seu banco de dados. Isso pode afetar os custos, pois talvez seja necessário introduzir réplicas de leitura para acomodar a carga. Se você estiver usando réplicas de leitura, é importante entender que elas só estão disponíveis na edição SQL Server Enterprise. Essa edição exige uma licença mais cara do que a edição SQL Server Standard.

O diagrama a seguir foi projetado para ajudar você a entender a eficácia do armazenamento em cache. Ele mostra o Amazon RDS for SQL Server com quatro nós db.m4.2xlarge executando a

edição Server Enterprise. SQL Ele é implantado em uma configuração Multi-AZ com uma réplica de leitura. O tráfego de leitura exclusivo (por exemplo, SELECT consultas) é direcionado às réplicas de leitura. Em comparação, o Amazon DynamoDB usa um cluster r4.2xlarge de dois nós do DynamoDB Accelerator (). DAX

O gráfico a seguir mostra os resultados da eliminação da necessidade de réplicas de leitura dedicadas que lidem com alto tráfego de leitura.



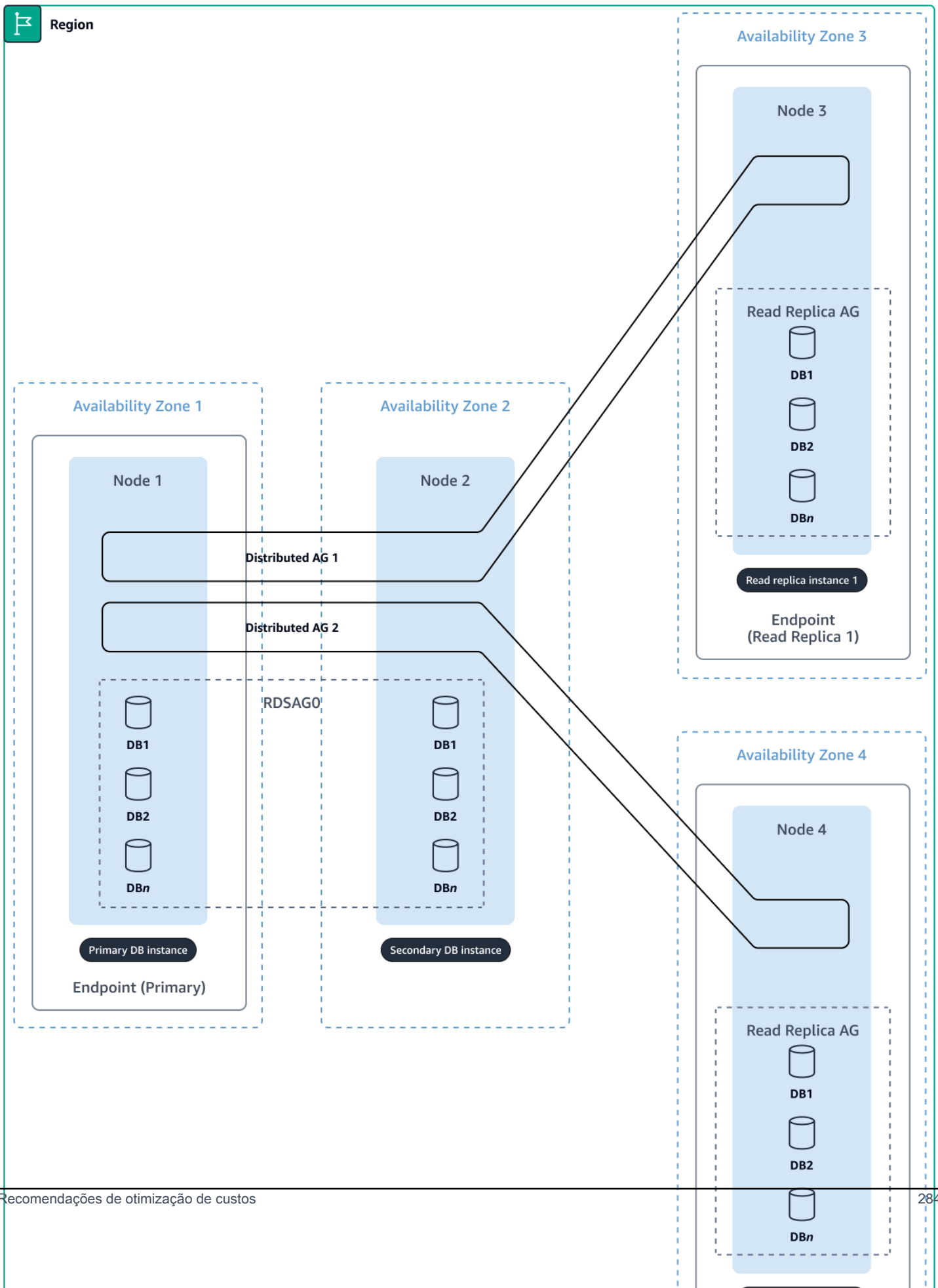
Você pode obter uma economia significativa usando o cache local sem réplicas de leitura ou introduzindo DAX lado a lado o SQL Server on Amazon RDS como uma camada de cache. Essa camada é descarregada do SQL Servidor e reduz o tamanho do SQL Servidor necessário para executar o banco de dados.

Recomendações de otimização de custos

Armazenamento em cache local

O cache local é uma das formas mais usadas de armazenar conteúdo em cache para aplicativos hospedados em ambientes locais ou na nuvem. Isso ocorre porque é relativamente fácil e intuitivo de implementar. O armazenamento em cache local envolve pegar conteúdo de um banco de dados ou de outra fonte e armazená-lo localmente na memória ou no disco para acesso mais rápido. Essa abordagem, embora fácil de implementar, não é ideal para alguns casos de uso. Por exemplo, isso inclui casos de uso em que o conteúdo em cache precisa persistir ao longo do tempo, como preservar o estado do aplicativo ou do usuário. Outro caso de uso é quando o conteúdo em cache precisa ser acessado de outras instâncias do aplicativo.

O diagrama abaixo ilustra um cluster de SQL servidores altamente disponível com quatro nós e duas réplicas de leitura.



Com o armazenamento em cache local, talvez seja necessário balancear a carga do tráfego em várias EC2 instâncias. Cada instância deve manter seu próprio cache local. Se o cache armazenar informações com estado, é necessário que haja confirmações regulares no banco de dados, e talvez os usuários precisem ser encaminhados para a mesma instância em cada solicitação subsequente (sessão fixa). Isso representa um desafio ao tentar escalar aplicativos, pois algumas instâncias podem ser superutilizadas, enquanto outras são subutilizadas devido à distribuição desigual do tráfego.

Você pode usar o cache local, na memória ou usando armazenamento local, para .NET aplicações. Para fazer isso, você pode adicionar funcionalidade para armazenar objetos em disco e recuperá-los quando necessário, ou consultar dados do banco de dados e mantê-los na memória. Para realizar o armazenamento em cache local na memória e no armazenamento local de dados de um SQL servidor em C#, por exemplo, você pode usar uma combinação de `MemoryCache` bibliotecas e `LiteDB`. `MemoryCache` fornece armazenamento em cache na memória, enquanto `LiteDB` é um banco de dados incorporado, sem SQL base em disco, que é rápido e leve.

Para realizar o armazenamento em cache na memória, use o.

`System.Runtime.MemoryCache`. O exemplo de código a seguir mostra como usar a `System.Runtime.Caching.MemoryCache` classe para armazenar dados em cache na memória. Essa classe fornece uma forma de armazenar dados temporariamente na memória do aplicativo. Isso pode ajudar a melhorar o desempenho de um aplicativo, reduzindo a necessidade de buscar dados de um recurso mais caro, como um banco de dados ou um API.

Veja como o código funciona:

1. Uma instância estática privada de `MemoryCache` named `_memoryCache` é criada. O cache recebe um nome (`dataCache`) para identificá-lo. Em seguida, o cache armazena e recupera os dados.
2. O `GetData` método é um método genérico que usa dois argumentos: uma `string` chave e um `Func<T>` delegado chamado `getData`. A chave é usada para identificar os dados em cache, enquanto o `getData` delegado representa a lógica de recuperação de dados que é executada quando os dados não estão presentes no cache.
3. O método primeiro verifica se os dados estão presentes no cache usando o `_memoryCache.Contains(key)` método. Se os dados estiverem no cache, o método recuperará os dados usando `_memoryCache.Get(key)` e os converterá no tipo esperado `T`.
4. Se os dados não estiverem no cache, o método chama o `getData` delegado para buscá-los. Em seguida, ele adiciona os dados ao cache usando `_memoryCache.Add(key, data,`

`DateTimeOffset.Now.AddMinutes(10)`). Essa chamada especifica que a entrada do cache deve expirar após 10 minutos, momento em que os dados são removidos do cache automaticamente.

5. O `ClearCache` método usa uma `string` chave como argumento e remove os dados associados a essa chave do cache usando `_memoryCache.Remove(key)`.

```
using System;
using System.Runtime.Caching;

public class InMemoryCache
{
    private static MemoryCache _memoryCache = new MemoryCache("dataCache");

    public static T GetData<T>(string key, Func<T> getData)
    {
        if (_memoryCache.Contains(key))
        {
            return (T)_memoryCache.Get(key);
        }

        T data = getData();
        _memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10));

        return data;
    }

    public static void ClearCache(string key)
    {
        _memoryCache.Remove(key);
    }
}
```

Você pode usar o seguinte código:

```
public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";

        Func<string> getSampleData = () =>
```



```
{
    // Replace this with your data retrieval logic
    return "Sample data";
};

string data = InMemoryCache.GetData(cacheKey, getSampleData);
Console.WriteLine("Data: " + data);
}
}
```

O exemplo a seguir mostra como usar o [LiteDB](#) para armazenar dados em cache no armazenamento local. Você pode usar o LiteDB como alternativa ou complemento ao cache na memória. O código a seguir demonstra como usar a biblioteca LiteDB para armazenar dados em cache no armazenamento local. A `LocalStorageCache` classe contém as principais funções para gerenciar o cache.

```
using System;
using LiteDB;

public class LocalStorageCache
{
    private static string _liteDbPath = @"Filename=LocalCache.db";

    public static T GetData<T>(string key, Func<T> getData)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
            var item = collection.FindOne(Query.EQ("_id", key));

            if (item != null)
            {
                return item;
            }
        }

        T data = getData();

        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
            collection.Upsert(new BsonValue(key), data);
        }
    }
}
```

```
        return data;
    }

    public static void ClearCache(string key)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection("cache");
            collection.Delete(key);
        }
    }
}

public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";

        Func<string> getSampleData = () =>
        {
            // Replace this with your data retrieval logic
            return "Sample data";
        };

        string data = LocalStorageCache.GetData(cacheKey, getSampleData);
        Console.WriteLine("Data: " + data);
    }
}
```

Se você tiver um cache estático ou arquivos estáticos que não mudam com frequência, você também pode armazenar esses arquivos no armazenamento de objetos do Amazon Simple Storage Service (Amazon S3). O aplicativo pode recuperar o arquivo de cache estático na inicialização para usar localmente. Para obter mais detalhes sobre como recuperar arquivos do Amazon S3 usando o .NET, consulte [Download de objetos](#) na documentação do Amazon S3.

Armazenamento em cache com DAX

Você pode usar uma camada de cache que pode ser compartilhada entre todas as instâncias do aplicativo. O [DynamoDB Accelerator DAX \(\)](#) é um cache de memória totalmente gerenciado e altamente disponível para o DynamoDB que pode oferecer uma melhoria de desempenho dez vezes

maior. Você pode usar DAX para reduzir custos reduzindo a necessidade de provisionar em excesso as unidades de capacidade de leitura nas tabelas do DynamoDB. Isso é especialmente útil para cargas de trabalho com muita leitura e que exigem leituras repetidas para chaves individuais.

O preço do DynamoDB é feito sob demanda ou com capacidade provisionada, portanto, o número de leituras e gravações por mês contribui para o custo. Se você leu cargas de trabalho pesadas, os DAX clusters podem ajudar a reduzir os custos ao reduzir o número de leituras nas tabelas do DynamoDB. Para obter instruções sobre como configurar DAX, consulte [Aceleração na memória com o DynamoDB Accelerator \(DAX\) na documentação do DynamoDB](#). Para obter informações sobre a integração de aplicativos, assista [Integrating DAX Amazon ASP DynamoDB into Your .NET Application](#) no YouTube.

Recursos adicionais

- [Aceleração na memória com o DynamoDB Accelerator \(DAX\) - Amazon DynamoDB \(documentação do DynamoDB\)](#)
- [Integrando o DAX Amazon ASP DynamoDB ao seu .NET Aplicativo \(YouTube\)](#)
- [Baixando objetos](#) (documentação do Amazon S3)

Considere a opção sem servidor. NET

Visão geral

A computação sem servidor se tornou uma abordagem popular para criar e implantar aplicativos. Isso se deve principalmente à escalabilidade e à agilidade que a abordagem sem servidor oferece ao criar uma arquitetura moderna. No entanto, é importante considerar o impacto nos custos da computação sem servidor em alguns cenários.

O Lambda é uma plataforma de computação sem servidor que permite que os desenvolvedores executem códigos sem a necessidade de servidores dedicados. Lambda é uma opção particularmente atraente para desenvolvedores que buscam reduzir os custos de infraestrutura. Com Lambda, os desenvolvedores podem desenvolver e implantar aplicativos altamente escaláveis e potencialmente econômicos. Ao usar uma abordagem sem servidor, os desenvolvedores não provisionam mais servidores para lidar com solicitações de aplicativos. Em vez disso, os desenvolvedores podem criar funções que são executadas sob demanda. Isso torna uma abordagem sem servidor mais escalável, gerenciável e potencialmente mais econômica do que executar, gerenciar e escalar máquinas virtuais. Como resultado, você paga apenas pelos

recursos usados pelo aplicativo, sem precisar se preocupar com recursos subutilizados ou custos de manutenção do servidor.

Os desenvolvedores podem usar plataformas modernas e multiplataforma. NET versões para criar aplicativos sem servidor que sejam rápidos, eficientes e econômicos. As NET versões principais e mais recentes são uma estrutura gratuita e de código aberto que é mais adequada para execução em plataformas sem servidor em relação às anteriores. NET versões do framework. Isso permite que os desenvolvedores reduzam o tempo de desenvolvimento e aumentem o desempenho do aplicativo. Moderno. NET também suporta uma variedade de linguagens de programação, incluindo C# e F#. Por esse motivo, é uma opção atraente para desenvolvedores que desejam criar arquiteturas modernas na nuvem.

Esta seção explica como você pode obter economia de custos usando o Lambda como uma opção sem servidor. [Você pode otimizar ainda mais os custos ajustando os perfis de execução das funções do Lambda, dimensionando corretamente a alocação de memória das funções do Lambda, usando o Native e migrando para funções baseadas em Graviton. AOT](#)

Impacto nos custos

O quanto você pode reduzir custos depende de vários fatores, incluindo quantas execuções suas funções sem servidor executarão, além da quantidade de memória alocada e da duração de cada função. AWS Lambda oferece um nível gratuito, que inclui um milhão de solicitações gratuitas por mês e 400.000 GB de segundos de tempo de computação por mês. Você pode reduzir significativamente os custos mensais de cargas de trabalho que estão dentro ou perto desses limites de nível gratuito.

Também pode haver custos adicionais ao usar um balanceador de carga com funções Lambda como alvo. Isso é calculado como a quantidade de dados processados pelo balanceador de carga para os destinos [Lambda](#).

Recomendações de otimização de custos

Dimensione corretamente suas funções do Lambda

O dimensionamento correto é uma prática essencial para a otimização de custos em NET funções Lambda baseadas em funções. Esse processo envolve a identificação da configuração de memória ideal que equilibra desempenho e economia, sem exigir alterações no código.

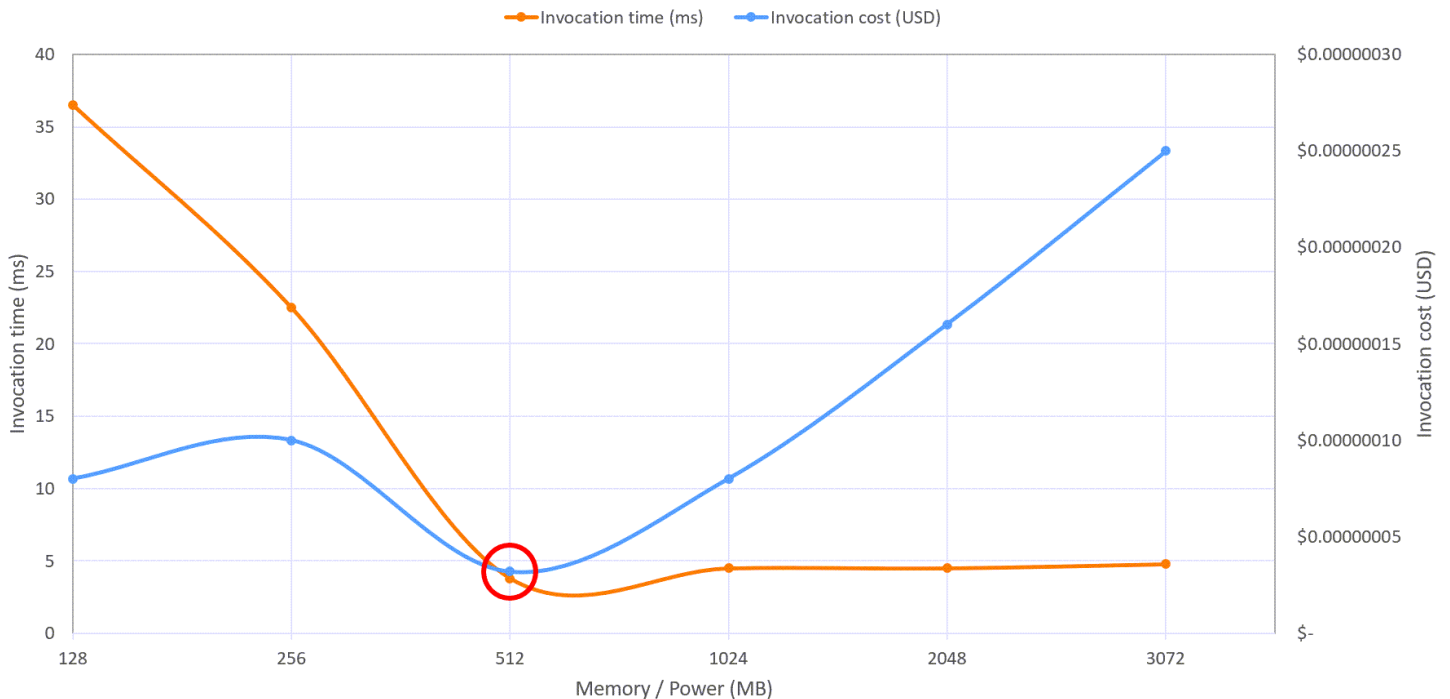
Ao configurar a memória para uma função Lambda, variando de 128 MB a 10.240 MB, você também ajusta a quantidade de v disponível durante a invocação CPU. Isso permite que aplicativos CPU

vinculados à memória ou a aplicativos vinculados acessem recursos adicionais durante a execução, levando a uma possível redução na duração da invocação e no custo geral.

No entanto, identificando a configuração ideal para seu. NETAs funções Lambda baseadas em Lambda podem ser um processo manual e demorado, especialmente se as alterações forem frequentes. A [ferramenta AWS Lambda Power Tuning](#) pode ajudá-lo a identificar a configuração apropriada analisando um conjunto de configurações de memória em relação a um exemplo de carga útil.

Por exemplo, aumentar a memória para um. NETA função Lambda baseada em Lambda pode melhorar o tempo total de invocação e reduzir o custo sem afetar o desempenho. A configuração de memória ideal para uma função pode variar. A ferramenta AWS Lambda Power Tuning pode ajudar a identificar a configuração mais econômica para cada função.

No gráfico de exemplo a seguir, o tempo total de invocação melhora à medida que a memória aumenta para essa função Lambda. Isso leva a uma redução no custo da execução total sem afetar o desempenho original da função. Para essa função, a configuração de memória ideal para a função é de 512 MB, pois é aqui que a utilização de recursos é mais eficiente para o custo total de cada invocação. Isso varia de acordo com a função, e o uso da ferramenta em suas funções do Lambda pode identificar se elas se beneficiam do dimensionamento correto.



Recomendamos que você conclua esse exercício regularmente, como parte de qualquer teste de integração quando novas atualizações forem lançadas. Se for atualizado com pouca frequência, faça

esse exercício periodicamente para garantir que as funções estejam ajustadas e dimensionadas corretamente. Depois de identificar a configuração de memória apropriada para suas funções do Lambda, você pode adicionar o dimensionamento correto aos seus processos. A ferramenta AWS Lambda Power Tuning gera uma saída programática que pode ser usada por seus fluxos de trabalho de CI/CD durante o lançamento do novo código. Isso permite automatizar a configuração da memória.

Você pode baixar a [ferramenta AWS Lambda Power Tuning](#) gratuitamente. Para obter instruções sobre como usar a ferramenta, consulte [Como executar a máquina de estado](#) em GitHub.

O Lambda também oferece suporte nativo AOT, o que permite .NET aplicativos a serem pré-compilados. Isso pode ajudar a reduzir custos ao reduzir os tempos de execução do .NET funções. Para obter mais informações sobre a criação de AOT funções nativas, consulte [.NET funções com AOT compilação nativa na documentação](#) do Lambda.

Evite o tempo de espera ocioso

A duração da função Lambda é uma dimensão usada para calcular o faturamento. Quando o código de função faz uma chamada de bloqueio, você é cobrado pelo tempo que ele espera para receber uma resposta. Esse tempo de espera pode aumentar quando as funções do Lambda são encadeadas ou quando uma função está atuando como orquestradora para outras funções. Se você tiver fluxos de trabalho, como operações em lote ou sistemas de entrega de pedidos, isso aumenta a sobrecarga de gerenciamento. Além disso, talvez não seja possível concluir toda a lógica do fluxo de trabalho e o tratamento de erros dentro do tempo limite máximo do Lambda de 15 minutos.

Em vez de lidar com essa lógica no código da função, recomendamos que você reestruture sua solução para usá-la [AWS Step Functions](#) como orquestradora do fluxo de trabalho. Ao usar um fluxo de trabalho padrão, você é cobrado por cada transição de [estado](#) dentro do fluxo de trabalho, em vez da duração total do fluxo de trabalho. Além disso, você pode mover o suporte para novas tentativas, condições de espera, fluxos de trabalho de erro e retornos de [chamada](#) para a condição de estado para permitir que suas funções do Lambda se concentrem na lógica de negócios. Para obter mais informações, consulte [Otimizando seus AWS Lambda custos — Parte 2](#) no blog sobre AWS computação.

Mude para funções baseadas em Graviton

As funções Lambda alimentadas pelos processadores Graviton2 de próxima geração agora estão disponíveis ao público em geral. As funções do Graviton2, usando uma arquitetura ARM baseada em processador, foram projetadas para oferecer um desempenho até 19% melhor a um custo

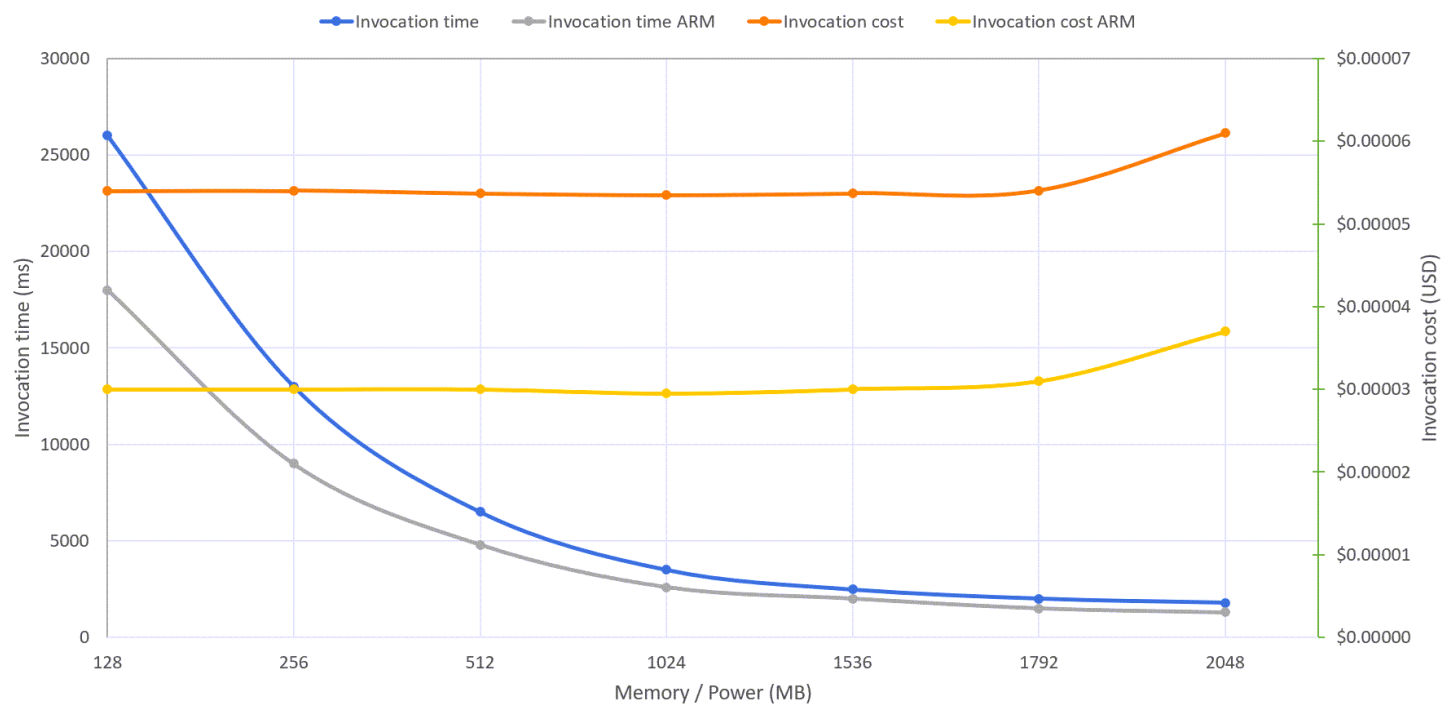
20% menor para uma variedade de cargas de trabalho sem servidor. Com menor latência e melhor desempenho, as funções alimentadas pelos processadores Graviton2 são ideais para alimentar aplicativos sem servidor de missão crítica.

Migrar para funções Lambda baseadas em Graviton pode ser uma opção econômica para .NETdesenvolvedores que desejam otimizar seus custos do Lambda. As funções baseadas em Graviton usam processadores ARM baseados em vez dos processadores x86 tradicionais. Isso pode levar a uma economia significativa de custos sem sacrificar o desempenho.

Embora haja vários benefícios em migrar para funções baseadas em Graviton, também há vários desafios e considerações que recomendamos que você considere. Por exemplo, as funções baseadas em Graviton exigem o uso do Amazon Linux 2, que pode não ser compatível com todos .NETaplicações. Além disso, pode haver problemas de compatibilidade com bibliotecas ou dependências de terceiros que não são compatíveis com processadores ARM baseados.

Se você estiver correndo .NETCrie aplicativos de estrutura e, se quiser aproveitar as vantagens da tecnologia sem servidor com o Lambda, considere a possibilidade de portar os aplicativos para aplicativos modernos .NETusando o [Assistente de portabilidade para .NET](#). Isso pode ajudar você a acelerar a portabilidade do legado .NETaplicações até modernas .NET, permitindo que o aplicativo seja executado no Linux.

O gráfico a seguir compara os resultados das arquiteturas x86 e ARM /Graviton2 para uma função que calcula números primos.



A função está usando um único thread. A duração mais baixa para ambas as arquiteturas é relatada quando a memória é configurada com 1,8 GB. Acima disso, as funções Lambda têm acesso a mais de 1 vCPU, mas, nesse caso, a função não pode usar a energia adicional. Pelo mesmo motivo, os custos são estáveis com memória de até 1,8 GB. Com mais memória, os custos aumentam porque não há benefícios adicionais de desempenho para essa carga de trabalho. O processador Graviton2 está claramente fornecendo melhor desempenho e custos mais baixos para essa função de computação intensiva.

Para configurar sua função para usar e basear o ARM processador com o Graviton, faça o seguinte:

1. Faça login no AWS Management Console e abra o console [Lambda](#).
2. Escolha a opção Criar função.
3. Em Function name (Nome da função), insira um nome.
4. Em Tempo de execução, escolha .NET6 (C#/ PowerShell).
5. Em Arquitetura, selecione arm64.
6. Faça as configurações adicionais necessárias e escolha Criar função.

Recursos adicionais

- [Funções do Lambda como alvos \(documentação\)](#) AWS
- [Otimizando o AWS Lambda custo e o desempenho usando AWS Compute Optimizer](#) (AWS Compute Blog)
- [Otimizando seus AWS Lambda custos — Parte 1](#) (AWS Compute Blog)
- [Otimizando seus AWS Lambda custos — Parte 2](#) (AWS Compute Blog)
- [Construindo sem servidor. NETaplicativos sobre AWS Lambda o uso. NET7](#) (Blog AWS de computação)

Considere bancos de dados criados especificamente

Visão geral

Um dos aspectos mais caros da execução de cargas de trabalho baseadas na Microsoft vem do licenciamento de bancos de dados comerciais, como o Server. SQL As empresas geralmente padronizam o SQL Server como a plataforma de banco de dados preferida e isso se torna arraigado na cultura de desenvolvimento da organização. Os desenvolvedores geralmente escolhem um

modelo relacional SQL baseado em servidor, independentemente do caso de uso. Os motivos para isso incluem:

- A empresa já tem instâncias e/ou licenças de SQL servidor disponíveis.
- As equipes se habituaram ao modelo de SQL programação por meio do uso de bibliotecas compartilhadas e lógica de negócios. ORMs
- A gerência não está ciente das alternativas.
- Os desenvolvedores não estão cientes das alternativas.

Bancos de dados criados especificamente podem acomodar os padrões de acesso aos dados do seu caso de uso. Esses bancos de dados são cada vez mais adotados pelas empresas à medida que adotam arquiteturas mais modernas (como microsserviços) e à medida que o escopo de aplicativos individuais é reduzido.

Um banco de dados criado especificamente não exclui um modelo relacional nem exige um modelo Não (não relacional). SQL Na verdade, um banco de dados relacional é considerado criado especificamente quando selecionado em resposta às necessidades específicas de uma carga de trabalho. O uso de bancos de dados criados especificamente pode ajudar as equipes a reduzir os custos de banco de dados associados aos seus. NETaplicativos, além de obter benefícios padrão de nuvem, como escalabilidade, resiliência e redução do trabalho pesado indiferenciado.

A tabela a seguir mostra os bancos de dados desenvolvidos especificamente oferecidos pelo. AWS

Banco de dados	Tipo	Características
Amazon Aurora Postgre ou Amazon SQL Aurora My SQL	Relacional	<p>Casos de uso em que os dados têm uma estrutura fixa</p> <p>Os bancos de dados relacionais naturalmente mantêm a consistência dos dados por meio de transações ACID</p>
Amazon DynamoDB	Par de valores-chave	Nenhum SQL banco de dados que armazene dados usando uma estrutura de dados de tabela de hash

Banco de dados	Tipo	Características
		<p>Armazenamento e recuperação de dados não estruturados de alto desempenho</p> <p>Os casos de uso incluem perfis de usuário, estado da sessão e dados do carrinho de compras</p>
Amazon ElastiCache	Na memória	<p>Alto desempenho Nenhum SQL banco de dados que armazene dados não estruturados na memória com tempo de acesso inferior a um milissegundo</p> <p>Usado para dados efêmeros acessados com frequência, como sessões de usuário e como uma camada de cache na frente de outros armazenamentos de dados mais lentos</p> <p>Inclui suporte para ElastiCache (RedisOSS) e ElastiCache (Memcached)</p>
Amazon MemoryDB	Memória durável	Banco de dados específico compatível com Redis com armazenamento durável

Banco de dados	Tipo	Características
Amazon Timestream	Séries temporais	<p>Banco de dados projetado para ingestão de dados de alto rendimento em ordem temporal</p> <p>Os casos de uso incluem aplicativos de Internet das Coisas (IoT) e armazenamento de métricas ou dados de telemetria</p>
Amazon DocumentDB	Documento	<p>Nenhum SQL banco de dados que armazene dados sem uma estrutura prescrita ou relações impostas com outros dados</p> <p>Geralmente usado para cargas de trabalho de leitura intensiva, como catálogos de produtos</p>
Amazon Neptune	Gráfico	<p>Nenhum SQL banco de dados que contenha dados e uma representação das conexões entre os itens de dados</p> <p>Os casos de uso incluem detecção de fraudes, mecanismos de recomendação e aplicativos sociais</p>

Banco de dados	Tipo	Características
Banco de dados Amazon Quantum Ledger (AmazonQLDB)	ledger	<p>Armazena dados de transações e fornece verificação criptográfica de cada transação, fornecendo um histórico auditável e imutável</p> <p>Frequentemente usado por aplicativos quando uma única fonte confiável verificável é necessária</p>
Amazon Keyspaces	Coluna larga	<p>Banco de dados distribuído de alto desempenho baseado no Apache Cassandra</p> <p>Os casos de uso incluem aplicativos de IoT, processamento de eventos e aplicativos de jogos</p>

Um fator significativo para a adoção de bancos de dados criados especificamente pode ser atribuído à eliminação do licenciamento comercial. No entanto, a capacidade de escalabilidade automática de bancos de dados como o [DynamoDB](#) (incluindo o modo [sob demanda](#)), [Aurora](#), Amazon [Neptune](#) e [Amazon Keyspaces](#) permite que você provisione capacidade para o caso [médio](#), em vez de para o uso de pico. Bancos de dados criados especificamente, como Timestream ou QLDB Amazon, não têm servidor e são escalados automaticamente para atender à demanda sem qualquer pré-provisionamento.

AWS oferece o [Babelfish para Aurora Postgre SQL](#) se você quiser usar um banco de dados relacional de código aberto e compatível com um propósito específico, mas não puder ou não quiser fazer alterações significativas no código do seu aplicativo. Em alguns casos, o Babelfish permite que você use um código de acesso ao SQL servidor existente, quase sem alterações.

Ao escolher um banco de dados relacional específico para aplicativos, é importante manter os mesmos recursos (ou funcionalmente equivalentes) que você precisa para seus aplicativos. Essa recomendação aborda bancos de dados criados especificamente como um armazenamento de

dados primário para aplicativos. Aplicativos específicos (como armazenamento em cache) são abordados em outras recomendações.

Impacto nos custos

Adotando bancos de dados criados especificamente para. NETembora seja improvável que as cargas de trabalho afetem diretamente o consumo/custo de computação, podem influenciar diretamente o custo dos serviços de banco de dados consumidos pelo. NETaplicações. Na verdade, a economia de custos pode ser uma meta secundária, quando comparada aos benefícios adicionais de agilidade, escalabilidade, resiliência e durabilidade dos dados.

Está fora do escopo deste guia explicar o processo completo de escolha de um banco de dados específico para aplicativos e a arquitetura de uma estratégia de dados para usá-los de forma eficaz. Para obter mais informações, consulte Bancos de [dados criados especificamente](#) no Diretório de tutoriais. AWS

As tabelas a seguir mostram vários exemplos de como a substituição do SQL Server por um banco de dados específico pode alterar os custos do aplicativo. Observe que essas são simplesmente estimativas aproximadas. São necessários benchmarks e otimização das cargas de trabalho reais para calcular o custo exato de produção.

Essas são algumas estimativas de banco de dados comumente usadas para fins específicos que incluem computação sob demanda e bancos de dados de instância única de 100 GB SSD em. us-east-1 Os custos da licença incluem licença de SQL servidor e garantia de software.

A tabela a seguir mostra os custos estimados para exemplos de bancos de dados comerciais.

Mecanismo do banco de dados	Modelo de licenciamento	Tipo/especificações da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
SQL Edição Server Standard na Amazon EC2	Licença incluída	r6i.2xlarge (8/64 GB) CPU RAM	\$1.345,36	\$0,00	\$1.345,36

Mecanismo do banco de dados	Modelo de licenciamento	Tipo/especificações da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
SQL Edição Server Enterprise na Amazon EC2	Licença incluída	r6i.2xlarge (8/64 GB) CPU RAM	\$2.834,56	\$0,00	\$2.834,56
SQL Edição Server Standard na Amazon EC2	BYOL	r6i.2xlarge (8/64 GB) CPU RAM	\$644,56	\$456,00	\$1.100,56
SQL Edição Server Enterprise na Amazon EC2	BYOL	r6i.2xlarge (8/64 GB) CPU RAM	\$644,56	\$1.750,00	\$2.394,56
SQL Edição Server Standard na Amazon RDS		db.r6i.2xlarge (8/64 GB) CPU RAM	\$2.318,30	\$0,00	\$2.318,30
SQL Edição Server Enterprise na Amazon RDS		db.r6i.2xlarge (8/64 GB) CPU RAM	\$3.750,56	\$0,00	\$3.750,56

A tabela a seguir mostra os custos estimados para exemplos específicos.

Mecanismo do banco de dados	Tipo/especificações da instância	AWS custo de computação + armazenamento	Custo da licença	Custo mensal total
Amazon Aurora Postger SQL	r6g.2xlarge (8/64 GB) CPU RAM	\$855,87	\$0,00	\$855,87
DynamoDB	Base provisionada 100/400 WCU RCU	\$72,00		\$72,00
Amazon DocumentDB	db.r6i.2xlarge (8/64 GB) CPU RAM	\$778.60		\$778.60
Amazon QLDB	Gravação de 80M 400M de leitura Diário de 2 TB Índice de 600 GB	\$320,00		\$320,00

Important

A tabela é baseada nos custos estimados de licenciamento SQL do Server with Software Assurance, durante os primeiros três anos da compra. Para a edição SQL Server Standard: \$4.100, pacote de 2 núcleos, 3 anos. Para a edição SQL Server Enterprise: \$15.700, pacote de 2 núcleos, 3 anos.

Recomendamos que você considere as implicações de custo antes de adotar bancos de dados específicos. Por exemplo, o custo de atualizar aplicativos para usar um banco de dados específico está relacionado à complexidade do aplicativo e do banco de dados de origem. Certifique-se de considerar o custo total de propriedade ao planejar essa mudança de arquitetura. Isso inclui refatorar seus aplicativos, aprimorar a equipe em novas tecnologias e planejar cuidadosamente o

desempenho e o consumo previstos para cada carga de trabalho. A partir daí, você pode determinar se o investimento vale a redução de custos. Na maioria dos casos, manter um end-of-support produto é um risco de segurança e conformidade, e o custo de remediá-lo vale o esforço e o investimento inicial.

Recomendações de otimização de custos

Para aplicativos que acessam o SQL Server, existem bibliotecas substitutas para bancos de dados relacionais criados especificamente. Você pode implementar essas bibliotecas em seu aplicativo para substituir uma funcionalidade similar SQL do aplicativo Server.

A tabela a seguir destaca algumas bibliotecas que podem ser usadas em muitos cenários comuns.

Ferramentas	Banco de dados	Substituto para	Compatibilidade de estrutura
Provedor principal do Npgsql Entity Framework	Amazon Aurora Postger SQL	Provedor de SQL servidor principal do Entity Fram	Moderno. NET
Provedor Npgsql Entity Framework 6	Amazon Aurora Postger SQL	Provedor de SQL servidor Entity Framework 6.0	. NET Estrutura
Npgsql (.ADO NET(SQL biblioteca Postgre compatível))	Amazon Aurora Postger SQL	ADO.NET	. NET Estrutura/Moder na. NET
Meu provedor principal do SQL Entity Framework	Amazon Aurora My SQL	Provedor de SQL servidor principal do Entity Fram	Moderno. NET
Pomelo. EntityFrameworkCore. MySql	Amazon Aurora My SQL	Provedor de SQL servidor principal do Entity Fram	Moderno. NET

[Conectar-se ao Amazon Aurora Postgre SQL usando o Babelfish](#) não requer nenhuma codificação especial para se conectar. No entanto, todo código deve ser exaustivamente testado antes de ser usado.

Outros bancos de dados criados especificamente têm bibliotecas para acesso. NETbibliotecas compatíveis que permitem acessar bancos de dados criados especificamente. Os exemplos incluem:

- [Usando o Amazon DynamoDB SQL Sem](#) bancos de dados (documentação)AWS SDK for .NET
- Driver [MongoDB C# \(documentação do MongoDB\)](#)
- [.NET](#)(Documentação do Timestream)
- [QLDBDriver da Amazon para .NET](#)(QLDBDocumentação da Amazon)
- [Usando uma Cassandra. NETDriver principal do cliente para acessar o Amazon Keyspaces programaticamente](#) (documentação do Amazon Keyspaces)
- [Usando .NETpara se conectar a uma instância de banco de dados Neptune \(documentação do Neptune\)](#)

Se você migrar para bancos de dados criados especificamente, poderá usar essas ferramentas AWS para ajudar no processo de migração:

- [AWS Schema Conversion Tool \(AWS SCT\)](#) pode ajudá-lo a transformar esquemas de SQL servidor no Amazon Aurora e no Amazon DynamoDB.
- [AWS Database Migration Service \(AWS DMS\)](#) podem ajudá-lo a migrar dados, uma vez ou continuamente, do SQL Server para o Aurora ou o DynamoDB.
- O [Babelfish Compass](#) pode ajudá-lo a verificar a compatibilidade do banco de dados do seu SQL servidor para uso com o Babelfish para Aurora Postgre. SQL

Recursos adicionais

- [Orientação para migrar o SQL servidor para o Amazon Aurora SQL Postgre AWS](#) (blog do banco de dados)
- [.NETWorkshop de modernização \(AWS Workshop Studio\)](#)
- [Dia de Imersão na APP Modernização do Babelfish](#) (Workshop Studio)AWS
- [.NETDia de imersão](#) (estúdio de AWS oficina)
- [Comece a usar o Amazon Timestream com .NET](#)(GitHub)

- [QLDBDriver da Amazon para .NET— Tutorial de início rápido](#) (QLDBdocumentação da Amazon)
- [Bancos de dados personalizados para uso moderno. NETinscrições em AWS](#) (AWS apresentação)

Próximas etapas

Depois de concluir a revisão deste guia, recomendamos que você execute as próximas etapas a seguir para implementar o MACO:

1. Entre em contato com um especialista em MACO. Um especialista em MACO pode ajudar a responder suas perguntas e resolver suas preocupações. Se você já estiver trabalhando com uma equipe de AWS contas, entre em contato com a equipe e solicite ajuda de um especialista em MACO. Se você não tiver uma equipe de contas, entre em contato com optimize-microsoft@amazon.com.
2. Aplique as recomendações. Aplique as recomendações, as melhores práticas e as estratégias que você aprendeu neste guia e ao falar com um especialista em MACO.
3. Acompanhe as mudanças de custos. Marque suas cargas de trabalho e use serviços como AWS Cost Explorer e AWS Budgets para acompanhamento, monitoramento e controle detalhados de custos.

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
Atualizações do SQL Server e do Container	Adicionamos as seções Otimizar o dimensionamento do SQL Server usando o Otimizador de Computação , as recomendações de Trusted Advisor revisão para cargas de trabalho do SQL Server e a seção Replataforma de aplicativos do Windows com o App2Container .	29 de junho de 2024
Otimização do licenciamento do SQL Server	Adicionamos a seção Otimizar o licenciamento do SQL Server usando o Compute Optimizer .	22 de maio de 2024
Publicação inicial	—	21 de dezembro 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 instância do EC2 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 [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 as AIOps são usadas na estratégia de migração para a AWS, 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.

Availability Zone (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 stressam 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 da 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 [postagens do CCoE no blog](#) de estratégia Nuvem AWS corporativa.

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: realizar investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma zona de pouso, definir um CCoE, 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 AWS CodeCommit. 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, AWS Panorama oferece dispositivos que adicionam CV às redes de câmeras locais, e a Amazon SageMaker 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. O CI/CD é comumente descrito como um pipeline. O CI/CD pode ajudar você 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

Consulte [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](#).

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.

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).

environment (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

Consulte [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 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.

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.

G

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.

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 gerenciar recursos, políticas e conformidade em todas as unidades organizacionais (UOs). 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 AH 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.

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.

I

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 segurança da AWS](#) recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação 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

I

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 das Coisas Industrial (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 [Construir uma estratégia de transformação digital para a Internet das Coisas Industrial \(IIoT\)](#).

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 segurança da AWS](#) recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação 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 machine learning com a 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](#).

migração de grande porte

Uma migração de 300 servidores ou mais.

LBAC

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

privilégio 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](#).

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-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.](#)

microsserviço

Um serviço pequeno e independente que se comunica por meio de APIs bem definidas e normalmente pertence a equipes pequenas e autônomas. Por exemplo, um sistema de seguradora pode incluir microsserviços 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 APIs leves. 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: rehoste a migração para o Amazon EC2 AWS com o 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 todos 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.

OU

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

NÃO

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 segurança da AWS](#) recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação 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 [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 microsserviço com base em padrões de acesso a dados e outros requisitos. Se seus microsserviços tiverem a mesma tecnologia de armazenamento de dados, eles poderão enfrentar desafios de implementação ou apresentar baixa performance. Os microsserviços 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 microsserviços](#).

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 em engenharia de sistemas que leva em consideração a privacidade em todo o processo de engenharia.

zonas hospedadas privadas

Um contêiner que armazena informações sobre como você quer que o Amazon Route 53 responda a consultas ao DNS para um domínio e seus subdomínios dentro de uma 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.

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.

publicar/assinar (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 \(RACI\)](#).

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](#).

rotação

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.

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 do Amazon EC2 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 no AWS Organizations. As SCPs definem barreiras de proteção ou estabelecem limites para as ações que um administrador pode delegar a usuários ou perfis. É possível usar SCPs como listas de permissão ou 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.

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 pode ser usado para interconectar as VPCs e as redes on-premises. 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 de VPC

Uma conexão entre duas VPCs que permite rotear 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 o [AWS Workload Qualification Framework](#).

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.

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.