

Cloud Computing

Prof: Rodrigo da Cruz Fujioka

Lattes: <http://lattes.cnpq.br/0843668802633139>

Versão: 1.0 – 23/02/2023



Amazon Elastic Compute Cloud (Amazon EC2)

O [Amazon Elastic Compute Cloud \(Amazon EC2\)](#) fornece capacidade computacional segura e redimensionável na nuvem como instâncias do Amazon EC2.

Imagine que você é responsável pela arquitetura dos recursos de sua empresa e precisa ser compatível com novos sites. Com recursos locais tradicionais, você precisa fazer o seguinte:

- Gastar dinheiro adiantadamente para comprar hardware.
- Aguardar até que os servidores sejam entregues.
- Instalar os servidores em seu data center físico.
- Fazer todas as configurações necessárias.

Em comparação, com uma instância do Amazon EC2, você pode usar um servidor virtual para executar aplicativos na nuvem AWS.

- Você pode provisionar e executar uma instância do Amazon EC2 em minutos.
- Você pode parar de usar a instância quando terminar de executar uma carga de trabalho.
- Você paga apenas pelo tempo de computação em que uma instância estiver em execução, não quando ela é interrompida ou encerrada.
- Você pode economizar custos pagando apenas pela capacidade do servidor necessária ou desejada.

Como o Amazon EC2 funciona

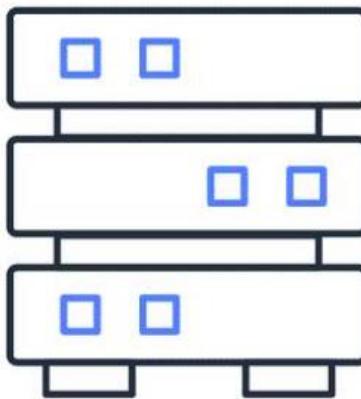
Para saber mais, selecione cada marcador.

1



Iniciar

2



Conectar

3



Usar



1

Execução



Iniciar

Primeiro, você executa uma instância. Comece selecionando um modelo com configurações básicas para sua instância. Essas configurações incluem o sistema operacional, o servidor de aplicativos ou os aplicativos. Você também precisa selecionar o tipo de instância, que é a configuração de hardware específica da sua instância.

À medida que se prepara para iniciar uma instância, você deve especificar as configurações de segurança para controlar o tráfego de rede que pode fluir para dentro e para fora da instância. Mais adiante neste curso, exploraremos os recursos de segurança do Amazon EC2 com mais detalhes.

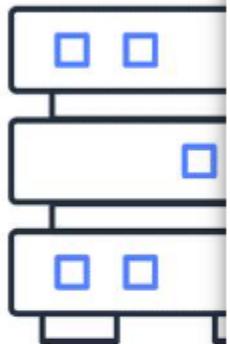
< >

3



Usar

2



Conexão

< >

Em seguida, conecte-se à instância. Você pode se conectar à instância de várias maneiras. Seus programas e aplicativos têm vários métodos diferentes para se conectar diretamente à instância e trocar dados. Os usuários também podem se conectar à instância iniciando a sessão e acessando a área de trabalho do computador.

Conecta...

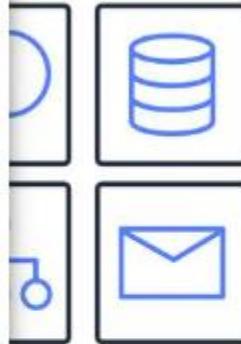


Uso

< >

3

Depois de se conectar à instância, você pode começar a usá-la. Você pode executar comandos para instalar software, adicionar armazenamento, copiar e organizar arquivos e muito mais.



Conectar

Usar

Os tipos de instâncias do Amazon EC2

[Os tipos de instâncias do Amazon EC2](#) são otimizados para tarefas diferentes. Ao selecionar um tipo de instância, considere as necessidades específicas de suas cargas de trabalho e seus aplicativos. Isso pode incluir requisitos para recursos de computação, memória ou armazenamento.

Para saber mais, clique no símbolo + ao lado de cada categoria.

Instâncias de uso geral	+
Instâncias otimizadas para computação	+
Instâncias otimizadas para memória	+
Instâncias de computação acelerada	+
Instâncias otimizadas para armazenamento	+

Instâncias de uso geral

Instâncias de uso geral equilibram os recursos de computação, memória e rede. Você pode usá-las para diversas cargas de trabalho, como:

- servidores de aplicativos
- servidores de jogos
- servidores de back-end para aplicativos empresariais
- bancos de dados pequenos e médios

Suponha que você tenha um aplicativo no qual as necessidades de recursos para computação, memória e rede sejam praticamente equivalentes. Você pode executar esse aplicativo em uma instância de uso geral porque ele não precisa de otimização em nenhuma área de recurso único.

Instâncias otimizadas para computação

Instâncias otimizadas para computação são ideais para aplicativos vinculados à computação que se beneficiam de processadores de alto desempenho. Assim como instâncias de uso geral, você pode usar instâncias otimizadas para computação para cargas de trabalho, como servidores web, de aplicativos e de jogos.

No entanto, a diferença é que aplicativos otimizados para computação são ideais para servidores web de alto desempenho, servidores de aplicativos de computação intensiva e servidores de jogos dedicados. Você também pode usar instâncias otimizadas para computação para cargas de trabalho de processamento em lote, com o processamento de muitas transações em um único grupo.

Instâncias otimizadas para memória

Instâncias otimizadas para memória

são projetadas para fornecer desempenho rápido para cargas de trabalho que processam grandes conjuntos de dados na memória. Na computação, a memória é uma área de armazenamento temporário. Ela contém todos os dados e instruções de que uma unidade central de processamento (CPU) precisa para conseguir realizar ações. Antes que um programa de computador ou aplicativo possa ser executado, ele é carregado do armazenamento para a memória. Esse processo de pré-carregamento dá à CPU acesso direto ao programa de computador.

Suponha que você tenha uma carga de trabalho que precise que grandes quantidades de dados sejam pré-carregados antes de executar um aplicativo. Esse cenário pode ser de um banco de dados de alto desempenho ou uma carga de trabalho que envolva a execução de processamento em tempo real de uma grande quantidade de dados não estruturados. Nesses tipos de casos de uso, considere usar uma instância otimizada para memória. As instâncias otimizadas para memória permitem que você execute cargas de trabalho com altas necessidades de memória e tenha um ótimo desempenho.

Instâncias de computação acelerada

Instâncias de computação acelerada usam aceleradores de hardware, ou coprocessadores, para executar algumas funções de forma mais eficiente do que é possível em um software executado em CPUs. Exemplos dessas funções são cálculos de números com vírgula flutuante, processamento de gráficos e correspondência de padrões de dados.

Na computação, um acelerador de hardware é um componente que pode agilizar o processamento de dados. As instâncias de computação acelerada são ideais para cargas de trabalho, como aplicativos gráficos e streaming de jogos e de aplicativos.

Instâncias otimizadas para armazenamento

As **instâncias otimizadas para armazenamento** são projetadas para cargas de trabalho que exigem alto acesso sequencial de leitura e gravação a grandes conjuntos de dados no armazenamento local. Exemplos de cargas de trabalho adequadas para instâncias otimizadas para armazenamento são sistemas de arquivos distribuídos, aplicativos de data warehouse e sistemas de processamento de transações on-line de alta frequência (OLTP).

Na computação, o termo operações de entrada/saída por segundo (IOPS) é uma métrica que mensura o desempenho de um dispositivo de armazenamento. Ela indica quantas operações diferentes de entrada ou saída um dispositivo pode executar em um segundo. As instâncias otimizadas para armazenamento foram projetadas para fornecer dezenas de milhares de IOPS aleatórias e de baixa latência para aplicativos.

Imagine as operações de entrada como dados colocados em um sistema, como registros inseridos em um banco de dados. Uma operação de saída são dados gerados por um servidor. Um exemplo de saída pode ser a análise realizada nos registros em um banco de dados. Se você tiver um aplicativo com alto requisito de IOPS, uma instância otimizada para armazenamento poderá fornecer melhor desempenho em relação a outros tipos de instâncias não otimizados para esse tipo de uso.

Faça a correspondência de cada descrição com um tipo de instância do Amazon EC2.

☰ Equilibra recursos de computação, memória e rede

Otimizada para memória

☰ Ideal para bancos de dados de alto desempenho

Otimizada para armazenamento

☰ Adequada para aplicativos de data warehouse

Uso geral

☰ Oferece processadores de alto desempenho

Otimizada para computação

Faça a correspondência de cada descrição com um tipo de instância do Amazon EC2.

-  Ideal para bancos de dados de alto desempenho → Otimizada para memória
-  Adequada para aplicativos de data warehouse → Otimizada para armazenamento
-  Equilibra recursos de computação, memória e rede → Uso geral
-  Oferece processadores de alto desempenho → Otimizada para computação

Definição de preços do Amazon EC2

Com o Amazon EC2, você paga apenas pelo tempo de computação que usar. O Amazon EC2 oferece diversas opções de preço para diferentes casos de uso. Por exemplo, se o seu caso de uso tolerar interrupções, você poderá economizar com instâncias spot. Você também pode economizar assumindo um compromisso antecipadamente e bloqueando um nível mínimo de uso com instâncias reservadas.

Para saber mais, clique no símbolo + ao lado de cada categoria.

Instâncias sob demanda



Savings Plans do Amazon EC2



Instâncias reservadas



Instâncias spot



Hosts dedicados



Teste de conhecimento

Qual é a diferença entre o Savings Plans do Amazon EC2 e as instâncias spot?

Vire o flashcard clicando nele.

Savings Plans do Amazon
EC2



Instâncias spot



Savings Plans do Amazon EC2 são ideais para cargas de trabalho que envolvem uma quantidade consistente de uso de computação em um período de um ou três anos.

Com o Savings Plans do Amazon EC2, você pode reduzir seus custos de computação em até 72% em relação aos custos de instâncias sob demanda.

As **instâncias spot** são ideais para cargas de trabalho com horários de início e término flexíveis ou que toleram interrupções. Com as instâncias spot, você pode reduzir seus custos de computação em até 90% em relação aos custos de instâncias sob demanda.

Diferentemente do Savings Plans do Amazon EC2, as instâncias spot *não* exigem contratos ou o compromisso com uma quantidade consistente de uso de computação.

Scaling do Amazon EC2

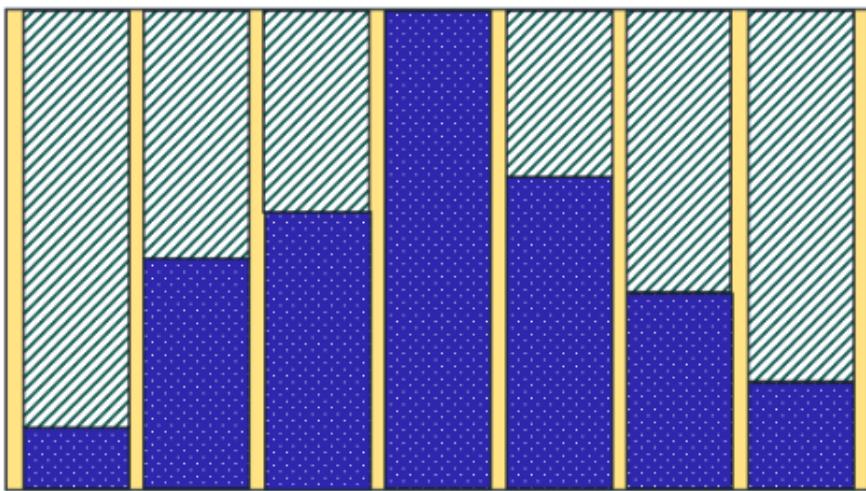
Escalabilidade

A **escalabilidade** envolve começar apenas com os recursos de que você precisa e projetar sua arquitetura para responder automaticamente às alterações de demanda, fazendo aumentos ou reduções. Como resultado, você paga apenas pelos recursos que usa. Você não precisa se preocupar com a falta de capacidade de computação para atender às necessidades de seus clientes.

Se você quisesse que o processo de scaling acontecesse automaticamente, qual serviço AWS você usaria? O serviço AWS que fornece essa funcionalidade para instâncias do Amazon EC2 é o **Amazon EC2 Auto Scaling**.

Amazon EC2 Auto Scaling

Se você já tentou acessar um site que não carregava e atingiu o tempo limite algumas vezes, ele pode ter recebido mais solicitações do que conseguia atender. Essa situação é semelhante a esperar em uma longa fila em uma cafeteria quando há apenas um barista disponível para registrar os pedidos dos clientes.



- Dom. Seg. Ter. Qua. Qui. Sex. Sáb.
 - Demanda
 - Capacidade não utilizada

O Amazon EC2 Auto Scaling permite que você adicione ou remova automaticamente instâncias do Amazon EC2 em resposta à alteração da demanda do aplicativo. Ao fazer auto scaling de suas instâncias, aumentando ou reduzindo conforme a necessidade, você consegue manter uma sensação maior de disponibilidade de aplicativos.

No Amazon EC2 Auto Scaling, há duas abordagens disponíveis: scaling dinâmico e scaling preditivo.

- O *scaling dinâmico* responde às alterações na demanda.
- O *scaling preditivo* programa automaticamente o número correto de instância do Amazon EC2 com base na demanda prevista.

Em seguida, você pode definir a **capacidade desejada** como duas instâncias do Amazon EC2, mesmo que o aplicativo precise de um mínimo de uma única instância do Amazon EC2 para que seja executado.



Se você não especificar o número desejado de instâncias do Amazon EC2 em um grupo do Auto Scaling, a capacidade desejada se tornará a capacidade mínima regular.

A terceira configuração que você pode definir em um grupo do Auto Scaling é a **capacidade máxima**. Por exemplo, você pode configurar o grupo do Auto Scaling para aumentar em resposta à demanda elevada, mas apenas para um máximo de quatro instâncias do Amazon EC2.

Como o Amazon EC2 Auto Scaling usa instâncias do Amazon EC2, você paga apenas pelas instâncias que usar, e somente quando elas forem usadas. Você agora tem uma arquitetura econômica que fornece a melhor experiência do cliente e ao mesmo tempo reduz custos.

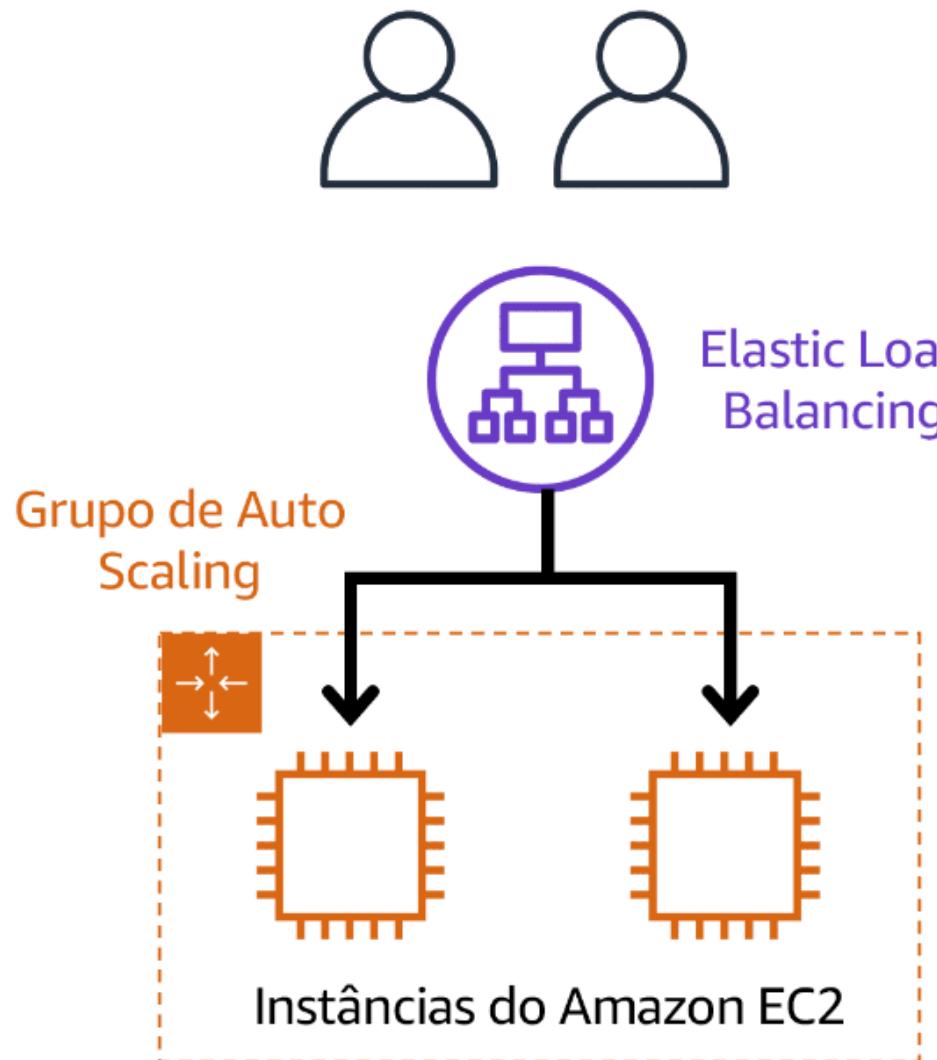
Elastic Load Balancing

O **Elastic Load Balancing** é o serviço AWS que distribui automaticamente o tráfego de entrada de aplicativos entre vários recursos, como instâncias do Amazon EC2.

Um balanceador de carga atua como um ponto único de contato para todo o tráfego da web de entrada no seu grupo do Auto Scaling. Isso significa que, à medida que você adiciona ou remove instâncias do Amazon EC2 em resposta à quantidade de tráfego de entrada, essas solicitações são direcionadas para o balanceador de carga primeiro. Em seguida, as solicitações se espalham por vários recursos que lidarão com elas. Por exemplo, se você tiver várias instâncias do Amazon EC2, o Elastic Load Balancing distribuirá a carga de trabalho entre elas para que nenhuma instância tenha que carregar a maior parte.

Embora o Elastic Load Balancing e o Amazon EC2 Auto Scaling sejam serviços separados, eles trabalham juntos para que os aplicativos executados no Amazon EC2 possam fornecer alto desempenho e disponibilidade.

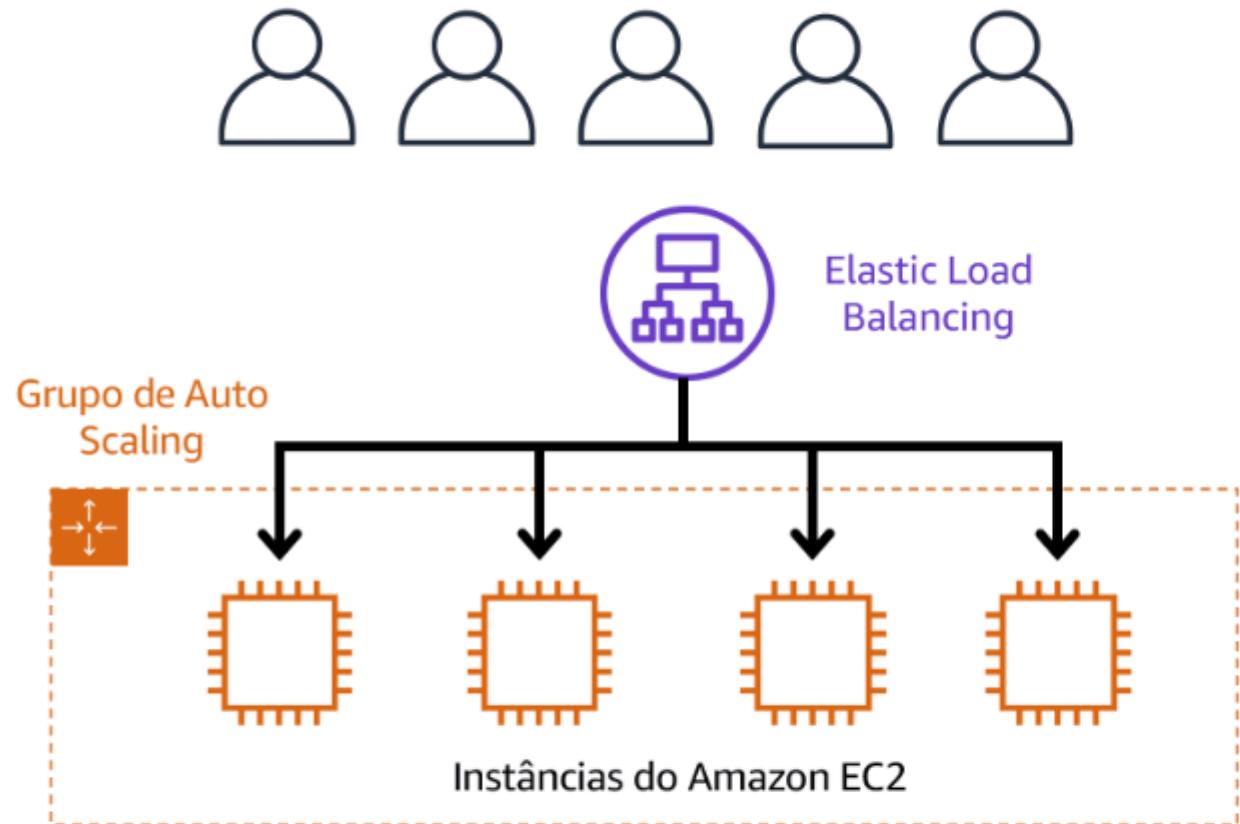
Exemplo: Elastic Load Balancing



Período de baixa demanda

Aqui está um exemplo de como o Elastic Load Balancing funciona. Suponha que alguns clientes vieram à cafeteria e estão prontos para fazer seus pedidos.

Se apenas algumas caixas registradoras estiverem abertas, isso corresponde à demanda dos clientes que precisam do serviço. A cafeteria tem menos probabilidade de ter caixas registradoras abertas sem clientes. Nesse exemplo, você pode pensar nas caixas registradoras como instâncias do Amazon EC2.



Período de alta demanda

Ao longo do dia, à medida que o número de clientes aumenta, a cafeteria abre mais caixas registradoras para acomodá-los. No diagrama, o grupo do Auto Scaling representa isso.

Além disso, um funcionário da cafeteria direciona os clientes para a caixa registradora mais adequada para que o número de solicitações possa ser distribuído uniformemente entre as caixas abertas. Você pode pensar nesse funcionário da cafeteria como um balanceador de carga.



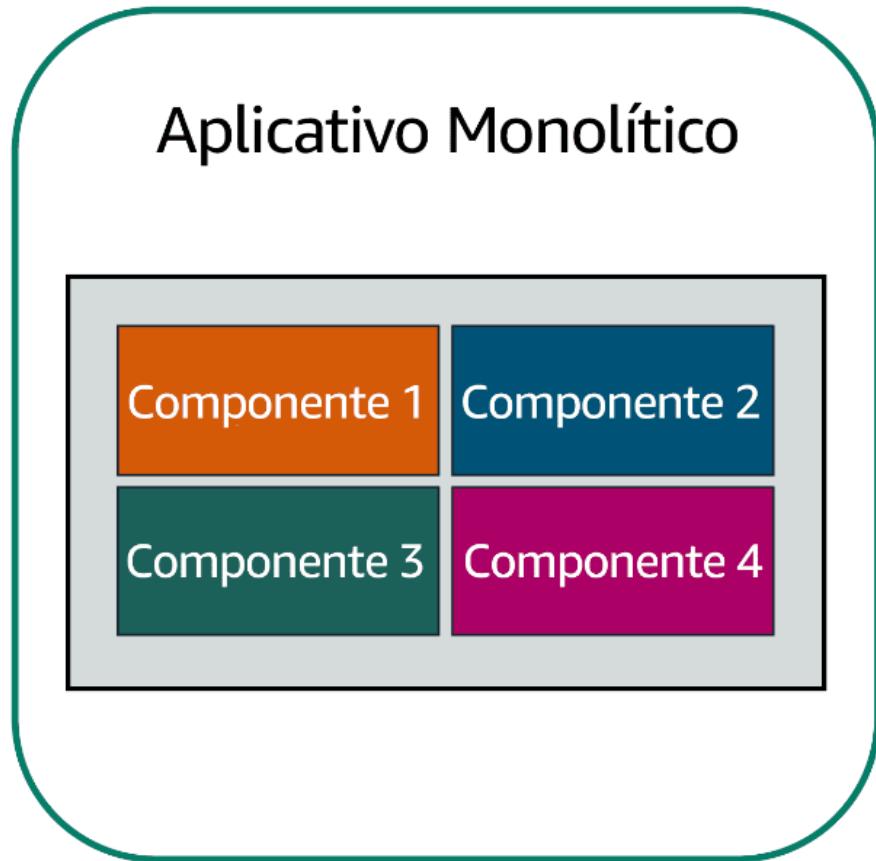


Amazon Simple Notification Service
(Amazon SNS)



Amazon Simple Queue Service
(Amazon SQS)

Aplicativos monolíticos e microserviços



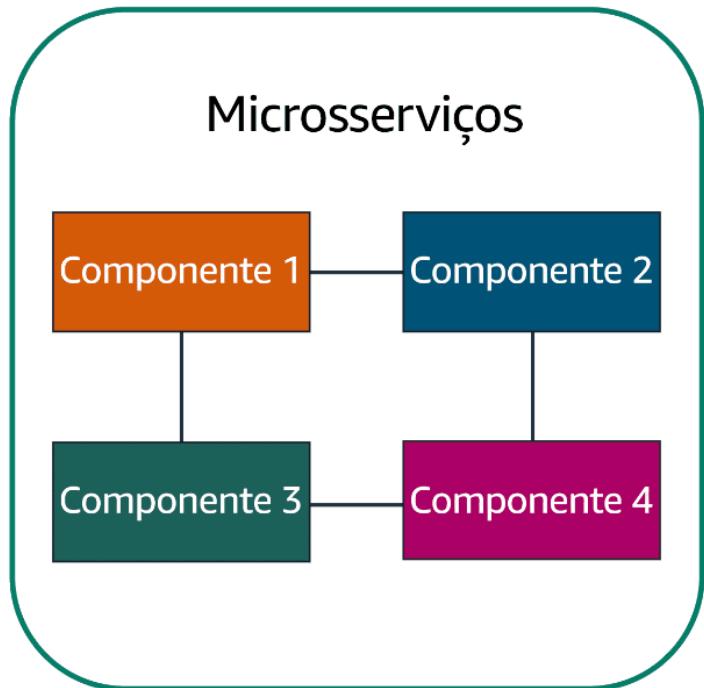
Os aplicativos são formados por vários componentes.

Os componentes se comunicam entre si para transmitir dados, atender solicitações e manter o aplicativo em execução.

Suponha que você tenha um aplicativo com componentes com acoplamento forte. Esses componentes podem ser bancos de dados, servidores, interface do usuário, lógica de negócios e assim por diante. Esse tipo de arquitetura pode ser considerado um **aplicativo monolítico**.

Nessa abordagem à arquitetura do aplicativo, se um único componente falhar, outros componentes falharão e possivelmente todo o aplicativo.

Para ajudar a manter a disponibilidade do aplicativo quando um único componente falha, você pode projetar esse aplicativo por uma abordagem de **microsserviços**.



Em uma abordagem de microsserviços, os componentes do aplicativo têm um acoplamento fraco. Neste caso, se um único componente falhar, os outros componentes continuarão funcionando porque estarão em comunicação uns com os outros. O acoplamento fraco evita a falha completa do aplicativo.

Ao projetar aplicativos na AWS, você pode adotar uma abordagem de microsserviços com serviços e componentes que cumprem funções diferentes. Dois serviços facilitam a integração de aplicativos: Amazon Simple Notification Service (Amazon SNS) e Amazon Simple Queue Service (Amazon SQS).

Amazon Simple Notification Service (Amazon SNS)

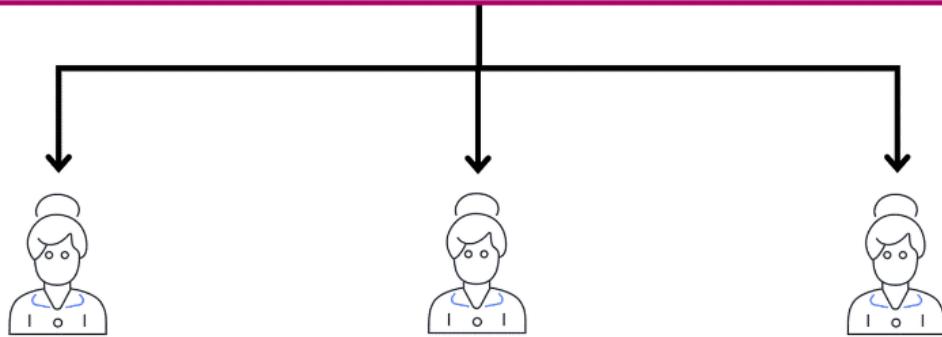
O **Amazon Simple Notification Service (Amazon SNS)** é um serviço de publicação/assinatura.

Usando tópicos do Amazon SNS, um editor publica mensagens para assinantes. Isso se parece com a cafeteria: o operador de caixa entrega os pedidos ao barista que, por sua vez, prepara as bebidas.

No Amazon SNS, os assinantes podem ser servidores web, endereços de e-mail, funções do AWS Lambda ou várias outras opções.

Publicação de atualizações de um único tópico

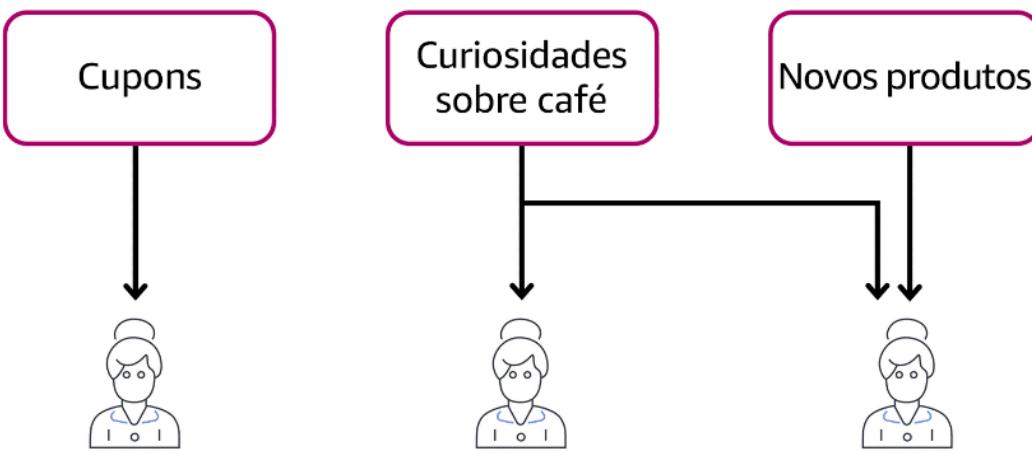
Cupons, curiosidades sobre café e novos produtos



Suponha que a cafeteria tenha uma única newsletter com atualizações de todas as áreas de seu negócio. Ela tem como tópicos cupons, curiosidades sobre café e novos produtos. Todos esses tópicos são agrupados por ser uma única newsletter. Todos os clientes inscritos na newsletter recebem atualizações sobre cupons, curiosidades sobre café e novos produtos.

Depois de algum tempo, alguns clientes informam que preferem receber newsletters separadas apenas para os tópicos específicos de interesse. Os proprietários da cafeteria decidem tentar essa abordagem.

Publicação de atualizações de vários tópicos



Agora, em vez de ter uma única newsletter para todos os tópicos, a cafeteria dividiu-a em três newsletters separadas. Cada newsletter é dedicada a um tópico específico: cupons, curiosidades sobre café e novos produtos.

Os assinantes receberão atualizações imediatamente apenas para os tópicos específicos aos quais se inscreveram.

É possível que os assinantes se inscrevam em um ou mais tópicos. Por exemplo, o primeiro cliente se inscreve apenas no tópico de cupons e o segundo assinante se inscreve apenas no tópico de curiosidades sobre café. O terceiro cliente se inscreve tanto em curiosidades sobre café quanto no tópico de novos produtos.



Suponha que a cafeteria tenha um processo de pedido em que um funcionário na caixa registradora recebe os pedidos e um barista prepara as bebidas. Pense nesses dois funcionários como dois componentes separados de um aplicativo.

Primeiro, o operador de caixa recebe o pedido e o anota em um pedaço de papel. Em seguida, ele entrega o papel ao barista. Por fim, o barista faz a bebida e a serve ao cliente.

Quando o próximo pedido chegar, o processo se repete. Esse processo funciona sem problemas, desde que o operador de caixa e o barista sejam coordenados.

O que poderia acontecer se o operador de caixa anotasse um pedido e o entregasse ao barista, mas esse último estivesse em uma pausa ou ocupado com outro pedido? O operador de caixa precisaria esperar até que o barista estivesse pronto para aceitar o pedido. Isso causaria atrasos no processo de pedido e faria com que os clientes tivessem que esperar mais para receber seus pedidos.

Como a cafeteria se tornou mais popular e a linha de pedidos está se movendo mais lentamente, os proprietários percebem que o processo atual é demorado e ineficiente. Eles decidem tentar uma abordagem diferente que usa uma fila.

Exemplo: Pedidos em uma fila



Qual serviço AWS é a melhor opção para publicar mensagens para assinantes?

- Amazon Simple Queue Service (Amazon SQS)
- Amazon EC2 Auto Scaling
- Amazon Simple Notification Service (Amazon SNS)
- Elastic Load Balancing

A resposta correta é **Amazon Simple Notification Service (Amazon SNS)**.

O Amazon SNS é um serviço de publicação/assinatura. Usando tópicos do Amazon SNS, um editor publica mensagens para assinantes.

As outras respostas estão incorretas porque:

- O Amazon Simple Queue Service (Amazon SQS) é um serviço de enfileiramento de mensagens. Ele não usa a assinatura de mensagem e o modelo de tópicos envolvidos no Amazon SNS.
- O Amazon EC2 Auto Scaling permite que você adicione ou remova automaticamente instâncias do Amazon EC2 em resposta à alteração da demanda do aplicativo.
- O Elastic Load Balancing é o serviço AWS que distribui automaticamente o tráfego de entrada de aplicativos entre vários recursos, como instâncias do Amazon EC2.

Qual serviço AWS é a melhor opção para publicar mensagens para assinantes?

Amazon Simple Queue Service (Amazon SQS)

Amazon EC2 Auto Scaling

Amazon Simple Notification Service (Amazon SNS)

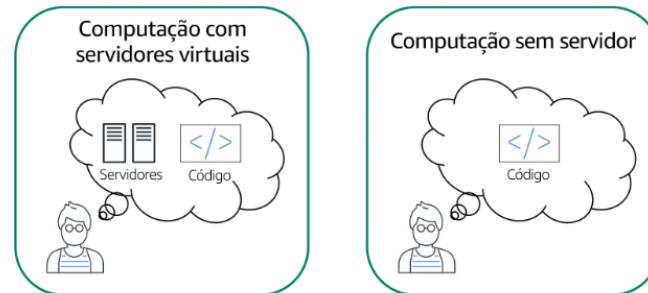
Elastic Load Balancing

Outros serviços de computação

Computação sem servidor

No início deste módulo, você conheceu o Amazon EC2, um serviço que permite executar servidores virtuais na nuvem. Se você quiser executar aplicativos no Amazon EC2, faça o seguinte:

- 1 Provisione as instâncias (servidores virtuais).
- 2 Faça upload do código.
- 3 Continue gerenciando as instâncias enquanto o aplicativo está em execução.



O termo “sem servidor” significa que o código é executado em servidores, sem que você precise provisionar ou gerenciar esses servidores. Com a computação sem servidor, você pode se concentrar na inovação de novos produtos e recursos em vez de manter servidores.

Outro benefício da computação sem servidor é a flexibilidade de dimensionar aplicativos sem servidor automaticamente. A computação sem servidor pode ajustar a capacidade de aplicativos modificando as unidades de consumo, como taxa de transferência e memória.

Um serviço AWS para computação sem servidor é o **AWS Lambda**.

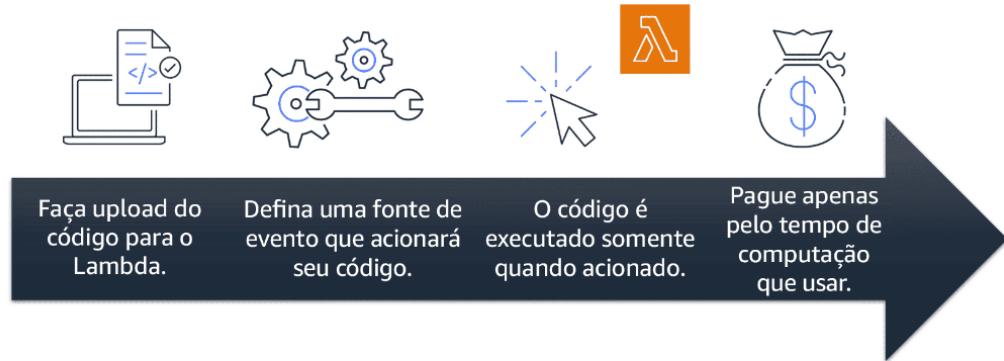
AWS Lambda

O [AWS Lambda](#) é um serviço que permite a execução de códigos sem a necessidade de provisionar ou gerenciar servidores.

Ao usar o AWS Lambda, você paga apenas pelo tempo de computação que consumir. As cobranças se aplicam ao tempo em que o código fica em execução. Você pode executar códigos para praticamente qualquer tipo de aplicativo ou serviço de back-end sem a necessidade de qualquer gerenciamento.

Por exemplo, uma função simples do Lambda é o redimensionamento automático de imagens com o upload feito na nuvem AWS. Nesse caso, a função é acionada ao fazer upload de uma nova imagem.

Como o AWS Lambda funciona



- 1 Você envia o código para o Lambda.
- 2 Você define que o código seja acionado a partir de uma origem de evento, como serviços AWS, aplicativos móveis ou endpoints HTTP.
- 3 O Lambda executa o código somente quando acionado.
- 4 Você paga apenas pelo tempo de computação que usar. No exemplo anterior de redimensionamento de imagens, você pagaria apenas pelo tempo de computação usado ao fazer upload de novas imagens. Fazer upload das imagens aciona o Lambda a executar o código da função de redimensionamento de imagem.

Na AWS, você também pode criar e executar aplicativos **em contêineres**.

Contêineres

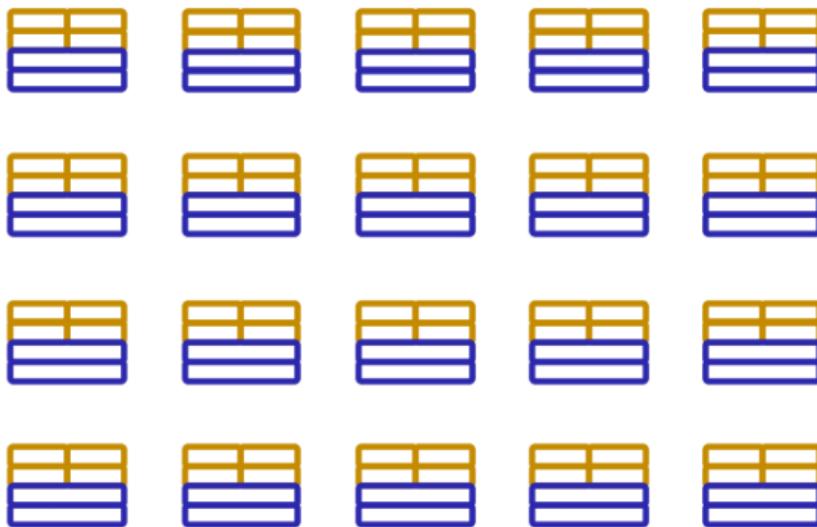
Os **contêineres** são uma maneira comum de empacotar códigos, configurações e dependências de aplicativo em um único objeto. Você também pode usar contêineres para processos e fluxos de trabalho nos quais há requisitos essenciais de segurança, confiabilidade e escalabilidade.

Um host com vários contêineres



Suponha que o computador do desenvolvedor de aplicativos de uma empresa tenha um ambiente diferente do ambiente dos computadores usados pela equipe de operações de TI. O desenvolvedor quer garantir que o ambiente do aplicativo permaneça consistente, independentemente da implantação, para usar uma abordagem em contêineres. Isso ajuda a reduzir o tempo gasto na depuração de aplicativos e no diagnóstico de diferenças em ambientes de computação.

Dezenas de hosts com centenas de contêineres



Ao executar aplicativos em contêineres, é importante considerar a escalabilidade. Suponha que, em vez de um único host com vários contêineres, você precise gerenciar dezenas de hosts com centenas de contêineres. Alternativamente, você precisa gerenciar possivelmente centenas de hosts com milhares de contêineres. Em grande escala, imagine quanto tempo pode demorar para monitorar o uso da memória, a segurança, o registro em log e assim por diante.

Amazon Elastic Container Service (Amazon ECS)

O [Amazon Elastic Container Service \(Amazon ECS\)](#) é um sistema de gerenciamento de contêineres altamente dimensionável e de alto desempenho que permite executar e dimensionar aplicativos em contêineres na AWS.

O Amazon ECS é compatível com contêineres Docker. O [Docker](#) é uma plataforma de software que permite criar, testar e implantar aplicativos rapidamente. A AWS é compatível com a Docker Community Edition de código aberto e do Docker Enterprise Edition baseado em assinatura. Com o Amazon ECS, você pode usar chamadas de API para iniciar e interromper aplicativos ativados pelo Docker.

Amazon Elastic Kubernetes Service (Amazon EKS)

O [Amazon Elastic Kubernetes Service \(Amazon EKS\)](#) é um serviço totalmente gerenciado que você pode usar para executar o Kubernetes na AWS.

O [Kubernetes](#) é um software de código aberto que permite implantar e gerenciar aplicativos em contêineres em grande escala. Uma grande comunidade de voluntários mantém o Kubernetes, e a AWS trabalha ativamente em conjunto com essa comunidade Kubernetes. Conforme novos recursos e funcionalidades são lançados para aplicativos Kubernetes, você pode facilmente aplicar essas atualizações aos aplicativos gerenciados pelo Amazon EKS.

Você deseja usar uma instância do Amazon EC2 para uma carga de trabalho de processamento em lote. Qual seria o melhor tipo de instância do Amazon EC2 a ser usado?

- Uso geral
- Otimizada para memória
- Otimizada para computação
- Otimizada para armazenamento

Você deseja usar uma instância do Amazon EC2 para uma carga de trabalho de processamento em lote. Qual seria o melhor tipo de instância do Amazon EC2 a ser usado?

Uso geral

Otimizada para memória

Otimizada para computação

Otimizada para armazenamento

A resposta correta é **Otimizada para computação**.

As outras respostas estão incorretas porque:

- instâncias de uso geral equilibram os recursos de computação, memória e rede. Essa família de instâncias não seria a melhor escolha para o aplicativo nesse cenário. As instâncias otimizadas para computação são mais adequadas para cargas de trabalho de processamento em lote do que as instâncias de uso geral.
- As instâncias otimizadas para memória são ideais para cargas de trabalho que processam grandes conjuntos de dados na memória, como bancos de dados de alto desempenho.
- As instâncias otimizadas para armazenamento foram projetadas para cargas de trabalho que exigem alto acesso sequencial de leitura e gravação a grandes conjuntos de dados no armazenamento local. A questão não especifica o tamanho dos dados que serão processados. O processamento em lote envolve o processamento de dados em grupos. Uma instância otimizada para computação é ideal para esse tipo de carga de trabalho, que se beneficiaria de um processador de alto desempenho.

Você tem uma carga de trabalho que será executada por um total de seis meses e consegue suportar interrupções. Qual seria a opção de compra mais econômica do Amazon EC2?

- Instância reservada
- Instância spot
- Instância dedicada
- Instância sob demanda

Você tem uma carga de trabalho que será executada por um total de seis meses e consegue suportar interrupções. Qual seria a opção de compra mais econômica do Amazon EC2?



Instância reservada



Instância spot



Instância dedicada



Instância sob demanda



Correto

A resposta correta é **Instância spot**.

As outras respostas estão incorretas porque:

- Instâncias reservadas exigem um contrato com duração de 1 ou 3 anos.
A carga de trabalho neste cenário será executada apenas por 6 meses.
- Instâncias dedicadas são executadas em um Virtual Private Cloud (VPC) em hardware dedicado a um único cliente. Elas têm um custo mais alto do que outras respostas, que são executadas em hardware compartilhado.
- As instâncias sob demanda cumprem os requisitos de execução por apenas seis meses e são resistentes a interrupções. No entanto, uma instância spot seria a melhor escolha porque não requer contrato com duração mínima, consegue suportar interrupções e custa menos do que uma instância sob demanda.

Você deseja implantar e gerenciar aplicativos em contêineres. Qual serviço deve usar?

- AWS Lambda
- Amazon Simple Notification Service (Amazon SNS)
- Amazon Simple Queue Service (Amazon SQS)
- Amazon Elastic Kubernetes Service (Amazon EKS)

AWS Fargate

O [AWS Fargate](#) é um mecanismo de computação sem servidor para contêineres. Ele funciona com o Amazon ECS e o Amazon EKS.

Com o AWS Fargate, você não precisa provisionar ou gerenciar servidores. O AWS Fargate gerencia sua infraestrutura de servidor para você. Você pode se concentrar em inovar e desenvolver seus aplicativos, pagando apenas pelos recursos necessários para executar os contêineres.



Cadastrar-se na AWS

Verificação segura

Você não será cobrado pelo uso abaixo dos limites do nível gratuito da AWS. Podemos armazenar temporariamente até USD 1 (ou um valor equivalente em moeda local) como uma transação pendente por 3 a 5 dias para verificar sua identidade.



Informações de faturamento

Número do cartão de crédito



A AWS aceita a maioria dos principais cartões de crédito. Para saber mais sobre as opções de pagamento, consulte nossas [perguntas frequentes](#)

Data de expiração

Mês ▾ Ano ▾

Nome do titular do cartão

Endereço de faturamento

Usar meu endereço de contato

Rua Otto feio da Silveira, 31
João Pessoa Paraíba 58031010
BR

Usar um novo endereço

Tipo de registro fiscal

CNPJ

CPF

Número do registro fiscal

XXXXXX-XXXX/XXXX-XX



Cadastrar-se na AWS

Confirme sua identidade

Antes que você possa usar a conta da AWS, é necessário confirmar seu número de telefone. Ao continuar, o sistema automatizado da AWS entrará em contato com você com um código de verificação.

Código do país ou região

Brasil (+55) ▾

Número de telefone

[REDACTED]

Ramal

[REDACTED]

Verificação de segurança



Digite os caracteres como mostrado acima

hxz4z3

Ligar para mim agora (etapa 4 de 5)



Cadastrar-se na AWS

Selecionar um plano de suporte

Escolha um plano de suporte para sua conta comercial ou pessoal. [Compare planos e exemplos de definição de preço](#). Você pode alterar seu plano a qualquer momento no Console de Gerenciamento da AWS.

Suporte Basic - gratuito

- Recomendado para novos usuários que estão começando a usar a AWS
- Acesso por autoatendimento 24 horas por dia/7 dias por semana a todos os recursos da AWS
- Apenas para questões de conta e faturamento
- Acesso ao Personal Health Dashboard e ao Trusted Advisor



Suporte Developer - a partir de 29 USD/mês

- Recomendado para desenvolvedores que estão avaliando a AWS
- Acesso por e-mail ao AWS Support durante o horário comercial
- Tempo de resposta de 12 horas (horário comercial)



Suporte Business - a partir de 100 USD/mês

- Recomendado para executar cargas de trabalho de produção na AWS
- Suporte técnico 24 horas por dia/7 dias por semana por e-mail, telefone e chat
- Tempo de resposta de 1 hora
- Conjunto completo de recomendações de melhores práticas do Trusted Advisor



Precisa de suporte de nível Enterprise?

A partir de 15.000 USD por mês, você receberá tempo de resposta de 15 minutos e experiência ao estilo de concierge com um Gerente técnico da conta designado. [Saiba mais](#)

[Concluir cadastramento](#)



Parabéns

Agradecemos por atualizar sua assinatura da Amazon Web Services (AWS). Seu console de gerenciamento agora será atualizado com todos os serviços da AWS.

[Acesse o Console de Gerenciamento da AWS](#)

Página inicial do console [Informações](#)

[Redefinir para layout padrão](#)[+ Adicionar widgets](#)

Visitado recentemente [Informações](#)



Não há serviços visitados recentemente

Explore um desses serviços da AWS comumente visitados.

[IAM](#) [EC2](#) [S3](#) [RDS](#) [Lambda](#)

[Ver todos os serviços](#)

Bem-vindo à AWS



[Primeiros passos com a AWS](#)  Aprenda os fundamentos e encontre informações valiosas para aproveitar ao máximo a AWS.



[Treinamento e certificação](#)  Aprenda com especialistas da AWS e desenvolva suas habilidades e conhecimentos.



[Quais são as novidades da AWS?](#)  Descubra novos serviços, recursos e regiões da AWS.

AWS Health [Informações](#)

Problemas em aberto

0 Últimos 7 dias

Alterações programadas

0 Próximos e últimos 7 dias

Outras notificações

0 Últimos 7 dias

[Acessar o AWS Health](#)

Custo e uso [Informações](#)

 Falha ao carregar o widget. Tente atualizar a página.



Construir uma solução [Informações](#)

Comece a construir assistentes simples e fluxos de trabalho automatizados.

 Executar uma máquina virtual Com EC2 (2 mins)	 Registrar um domínio Com Route 53 (3 mins)
 Iniciar um projeto de desenvolvimento Com CodeStar (5 mins)	 Construir uma aplicação Web Com AWS App Runner (5 mins)
 Conectar um dispositivo IoT Com AWS IoT (5 mins)	 Implantar um microsserviço sem servidor Com API Gateway (2 mins)
 Construir usando servidores virtuais Com Lightsail (2 mins)	 Começar a migrar para a AWS Com AWS MGN (2 mins)
 Hospedar uma aplicação Web estática Com AWS Amplify Console (2 mins)	 Criar um SQL Server na AWS Com alta disponibilidade (HA e FCI) (2 mins)

Trusted Advisor [Informações](#)



Não há recomendações

Isso pode acontecer porque você não executou verificações do Trusted Advisor ou não tem planos AWS Business ou AWS Enterprise Support.



Services

Search

[Alt+S]

 New EC2 Experience
Tell us what you think X

Recursos

Visualização global do EC2 ×

Você está usando os seguintes recursos do Amazon EC2 na Região Oeste dos EUA (Oregon):

Instâncias (em execução)

0

Auto Scaling Groups

0

Grupos de posicionamento

0

Grupos de segurança

1

Hosts dedicados

0

Instâncias

0

IPs elásticos

0

Load balancers

0

Pares de chaves

0

Sprints

0

Volumes

0

ⓘ Dimensione, configure e implante facilmente grupos de disponibilidade Always On do Microsoft SQL Server na AWS usando o AWS Launch Wizard for SQL Server. [Saiba mais](#) ×

Executar instância

Para começar, execute uma instância do Amazon EC2, que é um servidor virtual na nuvem.

Executar instância ▼Migrar um servidor ↗

Observação: suas instâncias serão executadas na Região Oeste dos EUA (Oregon)

Eventos agendados

Oeste dos EUA (Oregon)

Nenhum evento programado

Migrar um servidor

Use o AWS Application Migration Service para simplificar e acelerar a migração de infraestruturas físicas, virtuais e de nuvem para a AWS.

[Comece a usar o AWS Application Migration Service](#) ↗

Integridade do serviço

AWS Health Dashboard ↗

Região

Oeste dos EUA (Oregon)

Status

🟢 Este serviço está funcionando normalmente

Zonas

Nome da zona

ID da zona

us-west-2a

usw2-az2

us-west-2b

usw2-az1

us-west-2c

usw2-az3

us-west-2d

usw2-az4

[Habilitar zonas locais adicionais](#)

Atributos da conta

C

Plataformas com suporte

- VPC

VPC padrão



Configurações

Criptografia do EBS

Zonas

Console serial do EC2

Especificação de crédito padrão

Experimentos com o console

Informações adicionais

[Guia de conceitos básicos](#)

[Documentação](#)

[Todos os recursos do EC2](#)

[Fóruns](#)

[Definição de preço](#)

[Entre em contato conosco](#)

Tópicos de ajuda

X

[Como associar um endereço IP público estático à minha instância Windows ou Linux do EC2?](#)

[Adicionar contas de usuário com SSH a uma instância Linux do EC2](#)

[Quais etapas preciso seguir antes de alterar o tipo de instância da minha instância Linux do EC2?](#)

[Usar endereços IP elásticos no Amazon EC2](#)

[Como excluir ou terminar meus recursos do Amazon EC2?](#)

Recursos

Visualização global do EC2



Você está usando os seguintes recursos do Amazon EC2 na Região Oeste dos EUA (Oregon):

Instâncias (em execução)	0	Auto Scaling Groups	0	Grupos de posicionamento	0	Grupos de segurança	1
Hosts dedicados	0	Instâncias	0	IPs elásticos	0	Load balancers	0
Pares de chaves	0	Snapshots	0	Volumes	0		

Instâncias		Informações	C	Conectar	Estado da instância	Ações	Executar instâncias	...															
<input type="text"/> Localizar instância por atributo ou tag (case-sensitive)																							
Estado da instância = running	X	<button>Limpar filtros</button>																					
Name	▼	ID de instância	▼	Estado da inst... ▼	Tipo de inst... ▼	Verificação de s... ▼	Status do al... ▼	Zona de dispon... ▼	DNS IPv4 público ▼	▼	Endereço IP... ▼	▼	IP elástico ▼	▼	IPs IPv6 ▼	▼	Monitoram... ▼	▼	Nome do grupo de s... ▼	▼	Nome da c... ▼	▼	Data de lançamento

Nenhuma instância correspondente encontrada

Nome

aulaunipesp

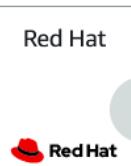
[Adicionar mais tags](#)

▼ Imagens de aplicação e de sistema operacional (imagem de máquina da Amazon) [Informações](#)

Uma AMI é um modelo que contém a configuração do software (sistema operacional, servidor de aplicações e aplicações) necessária para executar a instância. Pesquise ou navegue pelas AMIs se você não estiver vendo o que está buscando abaixo

Pesquise nosso catálogo completo, incluindo milhares de imagens de aplicações e sistemas operacionais

Início rápido

[Procurar mais AMIs](#)

Incluindo AMIs da AWS, do Marketplace e da comunidade

Imagen de máquina da Amazon (AMI)

Ubuntu Server 22.04 LTS (HVM), SSD Volume Type

ami-0735c191cf914754d (64 bits (x86)) / ami-079f51a7bcc65b92 (64 bits (Arm))

Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs

Qualificado para o nível gratuito



Descrição

Canonical, Ubuntu, 22.04 LTS, amd64 jammy image build on 2023-02-08

Arquitetura

64 bits (x86) ▾

ID da AMI

ami-0735c191cf914754d

Provedor verificado

<input type="checkbox"/>	
P	Ubuntu Server 22.04 LTS (HVM), SSD Volume Type ami-0735c191cf914754d (64 bits (x86)) / ami-079f51a7bccca65b92 (64 bits (Arm)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Qualificado para o nível gratuito
	Ubuntu Server 20.04 LTS (HVM), SSD Volume Type ami-00712dae9a53f8c15 (64 bits (x86)) / ami-0d43d68939ee00e60 (64 bits (Arm)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Qualificado para o nível gratuito
	Ubuntu Server 18.04 LTS (HVM), SSD Volume Type ami-0a97be4c4be6d6cc4 (64 bits (x86)) / ami-01f3b7c55a20e46c5 (64 bits (Arm)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Qualificado para o nível gratuito
	Ubuntu Server 20.04 LTS (HVM) with SQL Server 2022 Standard ami-0e8aab9a7a911cdc0 (64 bits (x86)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Deep Learning AMI GPU PyTorch 1.13.1 (Ubuntu 20.04) 20230309 ami-03a8a98906e5f7f9e (64 bits (x86)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Deep Learning AMI Neuron PyTorch 1.13.0 (Ubuntu 20.04) 20230309 ami-0ced4e5f9b96a793e (64 bits (x86)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Deep Learning AMI Habana PyTorch 1.12.0 SynapseAI 1.6.0 (Ubuntu 20.04) 20220928 ami-06b8d5099f3a8d79d (64 bits (x86)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Deep Learning AMI GPU TensorFlow 2.11.0 (Ubuntu 20.04) 20221220 ami-066a708ff36b35801 (64 bits (x86)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Ubuntu Server 22.04 LTS (HVM), SSD Volume Type ami-0735c191cf914754d (64 bits (x86)) / ami-079f51a7bccca65b92 (64 bits (Arm)) Virtualização: hvm ENA habilitado: true Tipo de dispositivo raiz: ebs
	Qualificado para o nível gratuito

▼ Tipo de instância [Informações](#)

Tipo de instância

t2.micro Qualificado para o nível gratuito

Família: t2 1 vCPU 1 GiB Memória
Sob demanda Linux definição de preço: 0.0116 USD por hora
Sob demanda SUSE definição de preço: 0.0116 USD por hora
Sob demanda Windows definição de preço: 0.0162 USD por hora
Sob demanda RHEL definição de preço: 0.0716 USD por hora

 | **t1.micro** Qualificado para o nível gratuito

Família: t1 1 vCPU 0.612 GiB Memória
Sob demanda SUSE definição de preço: 0.02 USD por hora
Sob demanda Windows definição de preço: 0.02 USD por hora
Sob demanda Linux definição de preço: 0.02 USD por hora
Sob demanda RHEL definição de preço: 0.08 USD por hora

t2.nano
Família: t2 1 vCPU 0.5 GiB Memória
Sob demanda Linux definição de preço: 0.0058 USD por hora
Sob demanda Windows definição de preço: 0.0081 USD por hora
Sob demanda SUSE definição de preço: 0.0058 USD por hora

t2.micro Qualificado para o nível gratuito

Família: t2 1 vCPU 1 GiB Memória
Sob demanda Linux definição de preço: 0.0116 USD por hora
Sob demanda SUSE definição de preço: 0.0116 USD por hora
Sob demanda Windows definição de preço: 0.0162 USD por hora
Sob demanda RHEL definição de preço: 0.0716 USD por hora

[Mais informações](#)

vpc-0117e46cdbe570e6e

z1d.12xlarge

Família: z1d 48 vCPU 384 GiB Memória

Sob demanda Windows definição de preço: 6.672 USD por hora

Sob demanda SUSE definição de preço: 4.589 USD por hora

Sob demanda RHEL definição de preço: 4.594 USD por hora

Sob demanda Linux definição de preço: 4.464 USD por hora

z1d.6xlarge

Família: z1d 24 vCPU 192 GiB Memória

Sob demanda Windows definição de preço: 3.336 USD por hora

Sob demanda SUSE definição de preço: 2.357 USD por hora

Sob demanda Linux definição de preço: 2.232 USD por hora

Sob demanda RHEL definição de preço: 2.362 USD por hora

z1d.3xlarge

Família: z1d 12 vCPU 96 GiB Memória

Sob demanda Windows definição de preço: 1.668 USD por hora

Sob demanda SUSE definição de preço: 1.241 USD por hora

Sob demanda Linux definição de preço: 1.116 USD por hora

Sob demanda RHEL definição de preço: 1.246 USD por hora

z1d.metal

Família: z1d 48 vCPU 384 GiB Memória

Sob demanda RHEL definição de preço: 4.594 USD por hora

Sob demanda Windows definição de preço: 6.672 USD por hora

Sob demanda Linux definição de preço: 4.464 USD por hora

Sob demanda SUSE definição de preço: 4.589 USD por hora



Tipo de instância

t1.micro

Família: t1 1 vCPU 0.612 GiB Memória
Sob demanda SUSE definição de preço: 0.02 USD por hora
Sob demanda Windows definição de preço: 0.02 USD por hora
Sob demanda Linux definição de preço: 0.02 USD por hora
Sob demanda RHEL definição de preço: 0.08 USD por hora

Qualificado para o nível gratuito

Resumo

Comparar tipos de instância

Número de instâncias

Criar par de chaves



Os pares de chaves permitem que você se conecte à sua instância com segurança.

Insira o nome do par de chaves abaixo. Quando solicitado, armazene a chave privada em um local seguro e acessível no seu computador. **Você precisará dele mais tarde para se conectar à sua instância.** [Saiba mais](#)

Par de chaves (login) Informações

Você pode usar um par de chaves para se conectar ao seu servidor selecionado antes de executar a instância.

Nome do par de chaves - **obrigatório**

Selecionar

Nome do par de chaves

iesp

O nome pode incluir até 255 caracteres ASCII. Ele não pode incluir espaços iniciais ou finais.

Configurações de rede Informações

Rede Informações

vpc-0117e46cdbe570e6e

Sub-rede Informações

Sem preferência (sub-rede padrão em qualquer rede)

Atribuir IP público automaticamente Informações

Habilitar

Firewall (grupos de segurança) Informações

Um grupo de segurança é um conjunto de regras de firewall que permitirão que o tráfego específico alcance sua instância.

Criar grupo de segurança

Tipo de par de chaves

RSA

Par de chaves públicas e privadas criptografadas por RSA

ED25519

Par de chaves ED25519 públicas e privadas criptografadas (não compatível com instâncias do Windows)

Formato de arquivo de chave privada

.pem

Para uso com OpenSSH

.ppk

Para uso com PuTTY

[Cancelar](#)

[Criar par de chaves](#)

existente

Criaremos um novo grupo de segurança chamado "launch-wizard-1" com as seguintes regras:

Habilitar**Firewall (grupos de segurança) [Informações](#)**

Um grupo de segurança é um conjunto de regras de firewall que controlam o tráfego para sua instância. Adicione regras para permitir que o tráfego específico alcance sua instância.

 Criar grupo de segurança

 Selecionar grupo de segurança existente

Criaremos um novo grupo de segurança chamado “**launch-wizard-1**” com as seguintes regras:

Permitir tráfego SSH de

Ajuda você a se conectar à sua instância

Qualquer lugar

0.0.0.0/0

Permitir tráfego HTTPS da Internet

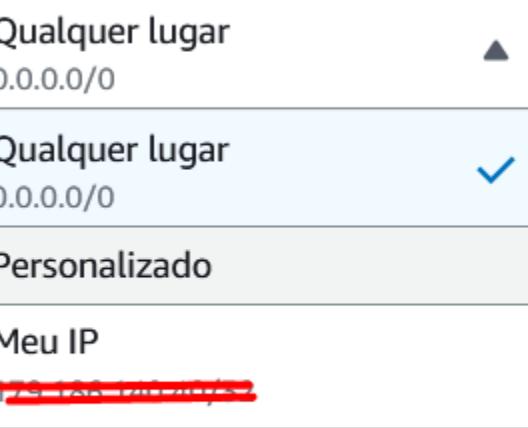
Para configurar um endpoint, por exemplo, ao criar um servidor Web

Permitir tráfego HTTP da Internet

Para configurar um endpoint, por exemplo, ao criar um servidor Web

⚠️ Regras com origem 0.0.0.0/0 permitem que todos os endereços IP accessem sua instância.

Recomendamos configurar regras de grupo de segurança para permitir o acesso apenas de endereços IP conhecidos.



) permitem que todos os endereços IP accessem egras de grupo de segurança para permitir o acce



▼ Resumo

Número de instâncias [Informações](#)

1

[Imagen do software \(AMI\)](#)

Canonical, Ubuntu, 22.04 LTS, ...[Ler mais](#)

ami-0735c191cf914754d

[Tipo de servidor virtual \(tipo de instância\)](#)

t1.micro

[Firewall \(grupo de segurança\)](#)

Novo grupo de segurança

[Armazenamento \(volumes\)](#)

1 volume(s) - 8 GiB



Nível gratuito: No primeiro ano, inclui 750 horas de uso de instâncias t2.micro (ou t3.micro nas regiões em que o t2.micro está indisponível) em AMIs de nível gratuito por mês, 30 GiB de armazenamento do EBS, 2 milhões de E/S, 1 GB de snapshots e 100 GB de largura de banda para a Internet



[Cancelar](#)

[Executar instância](#)



Êxito

Execução da instância iniciada com êxito ([1051a77100f75601](#))

▼ Log de execução

Inicializando solicitações	Com êxito
Criando grupos de segurança	Com êxito
Criando regras do grupo de segurança	Com êxito
Iniciar inicialização	Com êxito

Próximas etapas

Crie alertas de faturamento e uso de nível gratuito

Para gerenciar custos e evitar surpresas nas faturas, configure notificações por e-mail para faturamento e limites de uso de nível gratuito.

[Criar alertas de faturamento](#)

Conectar à sua instância

Assim que a instância estiver em execução, faça login nela usando seu computador local.

[Conectar-se à instância](#)

[Saiba mais](#)

Conectar um banco de dados do RDS

Configure a conexão entre uma instância do EC2 e um banco de dados para permitir o fluxo de tráfego entre eles.

[Conectar um banco de dados do RDS](#)

[Criar um novo banco de dados do RDS](#)

[Saiba mais](#)

Próximas etapas

Crie alertas de faturamento e uso de nível gratuito

Para gerenciar custos e evitar surpresas nas faturas, configure notificações por e-mail para faturamento e limites de uso de nível gratuito.

[Criar alertas de faturamento](#)

Conectar à sua instância

Assim que a instância estiver em execução, faça login nela usando seu computador local.

[Conectar-se à instância](#)

[Saiba mais](#)

Conectar um banco de dados do RDS

Configure a conexão entre uma instância do EC2 e um banco de dados para permitir o fluxo de tráfego entre eles.

[Conectar um banco de dados do RDS](#)

[Criar um novo banco de dados do RDS](#)

[Saiba mais](#)