



Introdução AWS

Introdução Prática a Computação em Nuvem

André Almeida

andre.almeida@ifrn.edu.br

Minicurso – Introdução a Computação em Nuvem - AWS

Tópicos

- Conceitos Gerais
- Infraestrutura Global
- Noções de Redes
- Amazon VPC e conceitos vinculados
- Recursos de Computação com EC2
- Serviços de Banco de Dados
- Balanceamento de Carga

Atividades

- Implantação de Aplicação Django Python em múltiplos cenários

Após o Minicurso

Depois de concluir este módulo, você será capaz de:

- Reconhecer os aspectos gerais de computação em nuvem e AWS
- Entender a estrutura básica de funcionamento de uma VPC
- Conhecer e criar instâncias EC2(VMs) com AWS
- Criar banco de dados gerenciados com RDS
- Configurar mecanismos de balanceamento de carga
- Iniciar a jornada para certificação AWS em Cloud Foundations

Conhecendo mais da AWS

- Inscrição nos cursos online da AWS Academy
- AWS Cloud Foundations
- AWS Web LAB
 - Permite o uso de 100 dólares para experimentar diversos produtos/serviços da AWS, até o final de 2025.
- Cada concurso concluído emite um Badge que pode ser compartilhado em perfis profissionais, em especial com o LinkedIn.

Seção 1: Visão geral e objetivos do curso

Entendo o curso de Cloud Foundations

Pré-requisitos do curso

- Conhecimento geral necessário
 - Conhecimento técnico de TI
 - Conhecimento empresarial de TI
- Conhecimento desejável
 - Familiaridade com conceitos de computação em nuvem
 - Conhecimento prático de sistemas distribuídos
 - Familiaridade com conceitos básicos de rede
 - Conhecimento prático de arquiteturas multicamada



Descrição do curso

- Módulo 1: Visão geral dos conceitos de nuvem
- Módulo 2: Economia e faturamento da nuvem
- Módulo 3: Visão geral da infraestrutura global da AWS
- Módulo 4: Segurança na Nuvem AWS
- Módulo 5: Redes e entrega de conteúdo
- Módulo 6: Computação
- Módulo 7: Armazenamento
- Módulo 8: Bancos de dados
- Módulo 9: Arquitetura de nuvem
- Módulo 10: Auto scaling e monitoramento



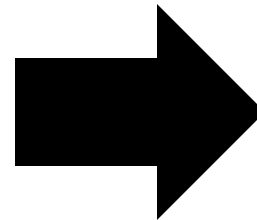
Definição de computação em nuvem

Computação em nuvem é a entrega **sob demanda** de poder computacional, banco de dados, armazenamento, aplicativos e outros recursos de TI **pela Internet** com uma definição de preço **conforme o uso**.



Infraestrutura como software

A computação em nuvem permite que você **deixe de pensar em sua infraestrutura como hardware** e passe a **pensar nela (e usá-la) como software**.

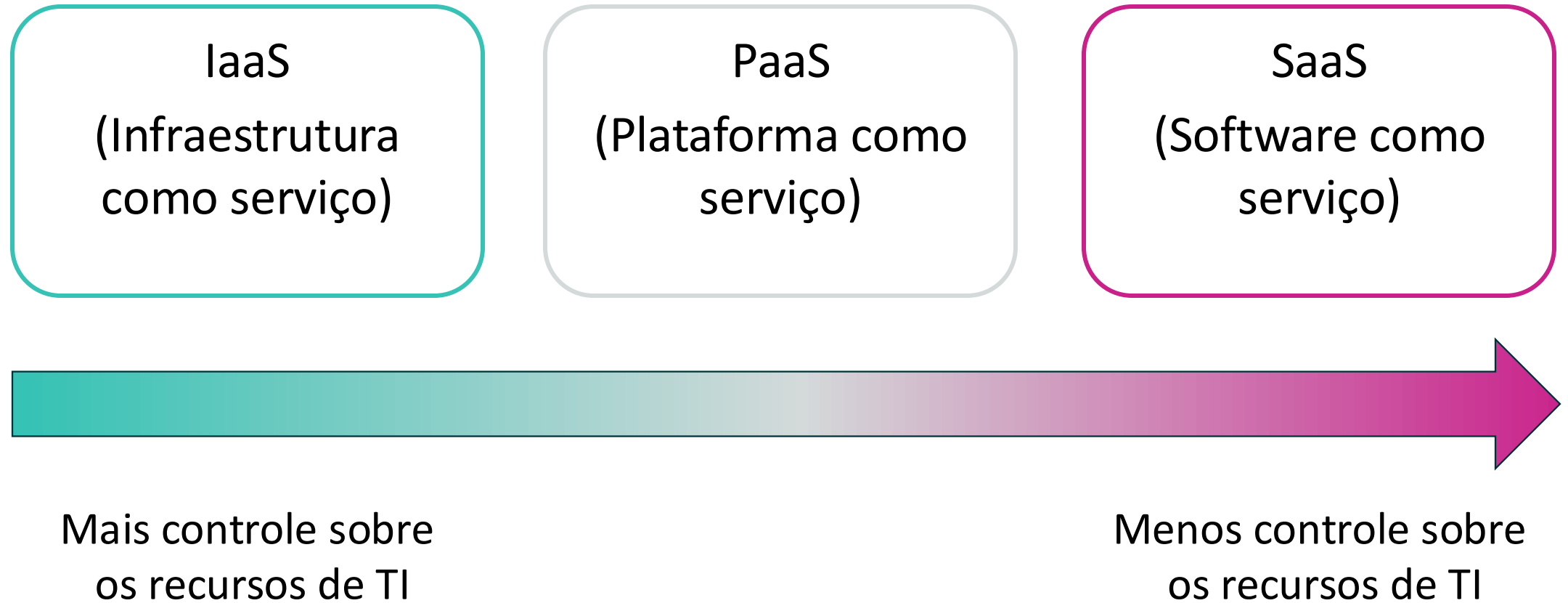


Modelo de computação tradicional



- Infraestrutura como hardware
- Soluções de hardware:
 - Exigem espaço, equipe, segurança física, planejamento, despesas de capital
 - Têm um ciclo longo de aquisição de hardware
 - Exigem provisionamento de capacidade por meio da tentativa de adivinhar os picos máximos teóricos

Modelos de serviço em nuvem

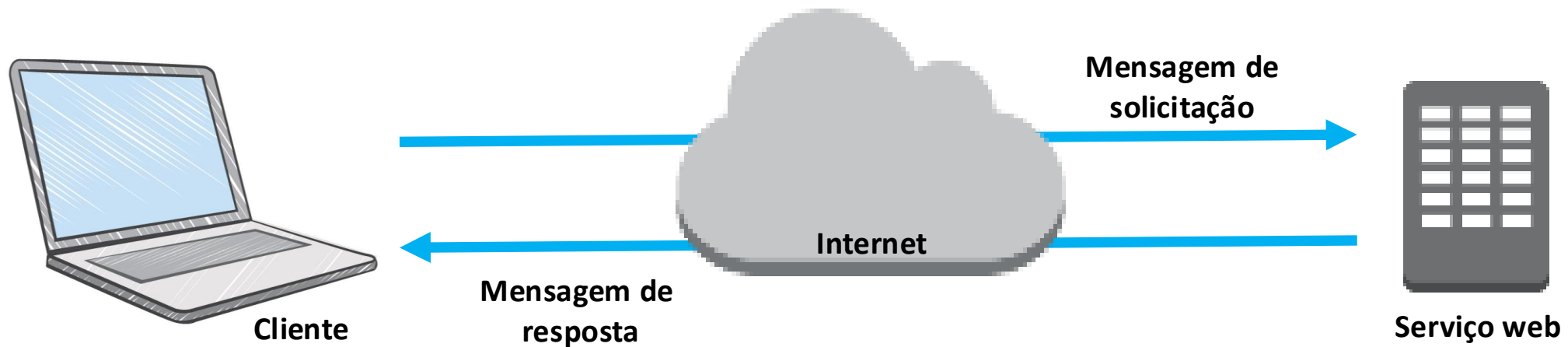


Semelhanças entre a AWS e a TI tradicional



O que são serviços web?

Um **serviço web** é qualquer software disponibilizado pela Internet que usa um **formato padronizado**, como Extensible Markup Language (XML) ou JavaScript Object Notation (JSON), para a solicitação e resposta de uma **interação de Application Programming Interface (API)**.



O que é a AWS?

- A AWS é uma **plataforma de nuvem segura** que oferece um **amplo conjunto de produtos globais baseados na nuvem**.
- A AWS oferece **acesso sob demanda** a recursos de computação, armazenamento, rede, banco de dados e outros recursos de TI e ferramentas de gerenciamento.
- A AWS oferece **flexibilidade**.
- Você **paga apenas pelos serviços individuais de que precisa**, pelo **tempo que os utilizar**.
- Os serviços da AWS **funcionam juntos** como componentes básicos.

Seção 2: Informações sobre o exame de certificação AWS

Introdução ao curso



Exames de certificação da AWS

Available AWS Certifications

aws  certified
Updated May 2019

Professional

Two years of comprehensive experience designing, operating, and troubleshooting solutions using the AWS Cloud



Associate

One year of experience solving problems and implementing solutions using the AWS Cloud



Specialty

Technical AWS Cloud experience in the Specialty domain as specified in the [exam guide](#)



Foundational

Six months of fundamental AWS Cloud and industry knowledge



O curso Cloud Foundations ajuda a preparar você para o exame de certificação AWS Cloud Practitioner



Exame AWS Certified Cloud Practitioner

- Para ver detalhes sobre o exame, inclusive informações sobre como se inscrever, acesse: <https://aws.amazon.com/certification/certified-cloud-practitioner/>
- Baixe e leia atentamente o guia para o exame AWS Certified Cloud Practitioner em <https://d1.awsstatic.com/training-and-certification/Docs - Cloud Practitioner/AWS Certified Cloud Practitioner-Exam Guide EN v1.6.pdf>
- Baixe as perguntas de exemplo do exame em <https://d1.awsstatic.com/training-and-certification/Docs - Cloud Practitioner/AWS Certified Cloud Practitioner Sample Questions v1.1 FINAL.PDF>
- Consulte o caminho recomendado para obter a certificação em <https://aws.amazon.com/training/path-cloudpractitioner/>
- O AWS Academy Cloud Foundations abrange grande parte do material encontrado no curso Cloud Practitioner Essentials, porém mais detalhadamente.
- Há ainda um treinamento digital gratuito, disponível em <https://www.aws.training/>.



Seção 3: Infraestrutura global da AWS

Visão geral da infraestrutura global da AWS

Infraestrutura Global da AWS

- A **Infraestrutura global da AWS** foi projetada e criada para entregar um ambiente de computação em nuvem **flexível, confiável, dimensionável**, e **seguro** com alta qualidade **global em desempenho de rede**.
- A AWS atualiza constantemente sua presença global de infraestrutura. Acesse uma das páginas a seguir para obter informações atualizadas sobre a infraestrutura:

- Mapa da infraestrutura global da AWS:
https://aws.amazon.com/about-aws/global-infrastructure/#AWS_Global_Infrastructure_Map
Escolha um círculo no mapa para ver um resumo de informações sobre a Região representada nesse círculo.
- Regiões e Zonas de Disponibilidade:
https://aws.amazon.com/about-aws/global-infrastructure/regions_az/
Escolha uma guia para ver um mapa da geografia selecionada e uma lista de Regiões, locais de borda, zonas locais e caches regionais.



Regiões AWS

- Uma **Região AWS** é uma área geográfica.
- A **replicação de dados** entre Regiões é controlada por você.
- A **comunicação** entre regiões usa a infraestrutura de rede backbone da AWS.
- Cada região fornece redundância total e conectividade com a rede.
- Uma Região normalmente consiste em duas ou mais **Zonas de Disponibilidade**.



Exemplo: Região de Londres

Seleção de uma Região

Determine a região certa
para serviços, aplicativos
e dados com base
nesses fatores



Governança de dados,
requisitos legais



Proximidade com
os clientes (latência)



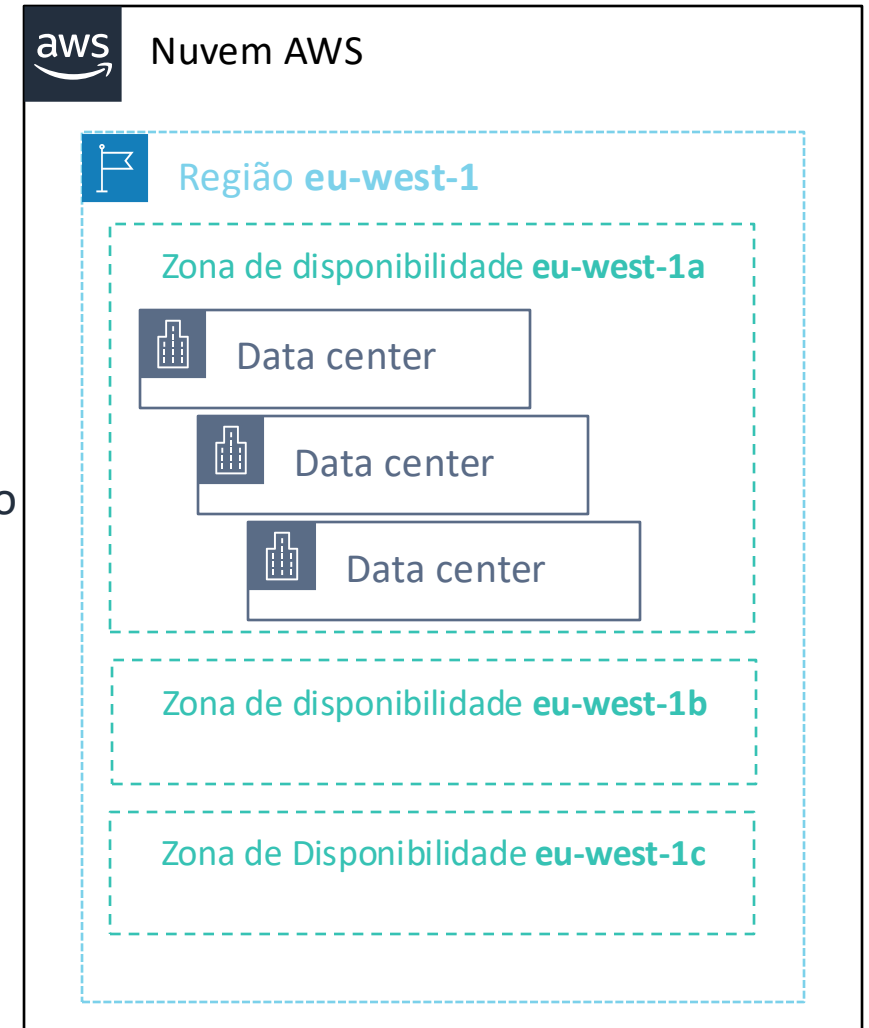
Serviços disponíveis na
região



Custos (variam por Região)

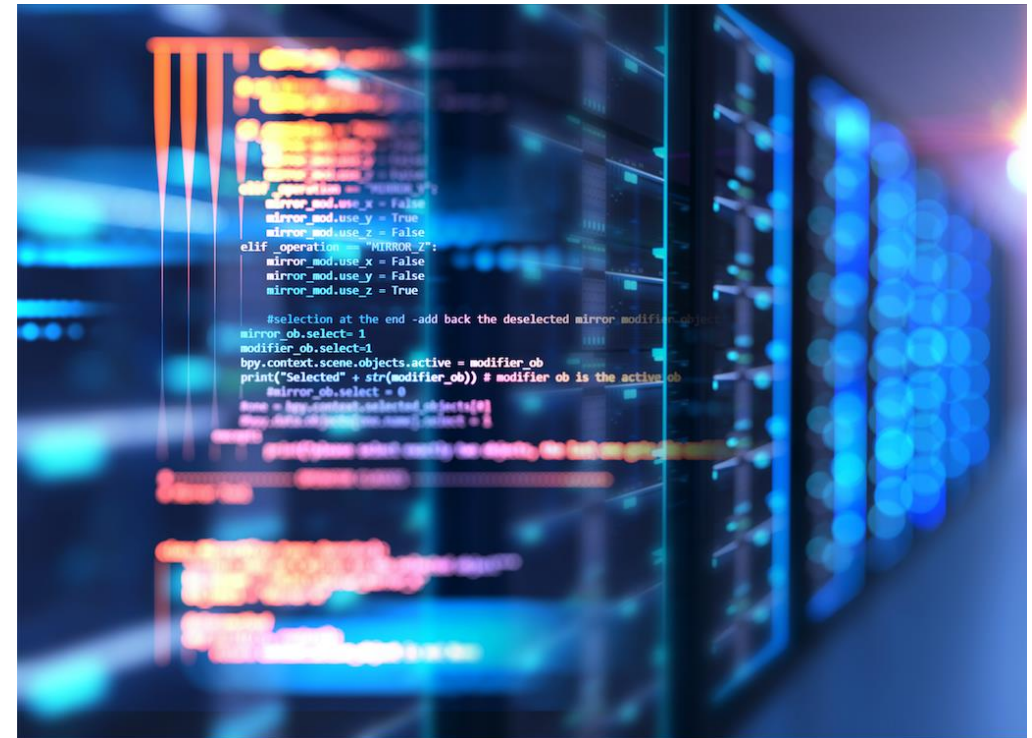
Zonas de Disponibilidade

- Cada **Região** tem várias Zonas de Disponibilidade.
- Cada **Zona de Disponibilidade** é uma partição totalmente isolada da infraestrutura da AWS.
 - As Zonas de Disponibilidade consistem em **data centers** distintos
 - Elas são projetadas para isolamento de falhas
 - Elas são interconectadas a outras zonas de disponibilidade usando redes privadas de alta velocidade
 - Você escolhe suas zonas de disponibilidade
 - **A AWS recomenda a replicação de dados e recursos entre Zonas de Disponibilidade** para fins de resiliência



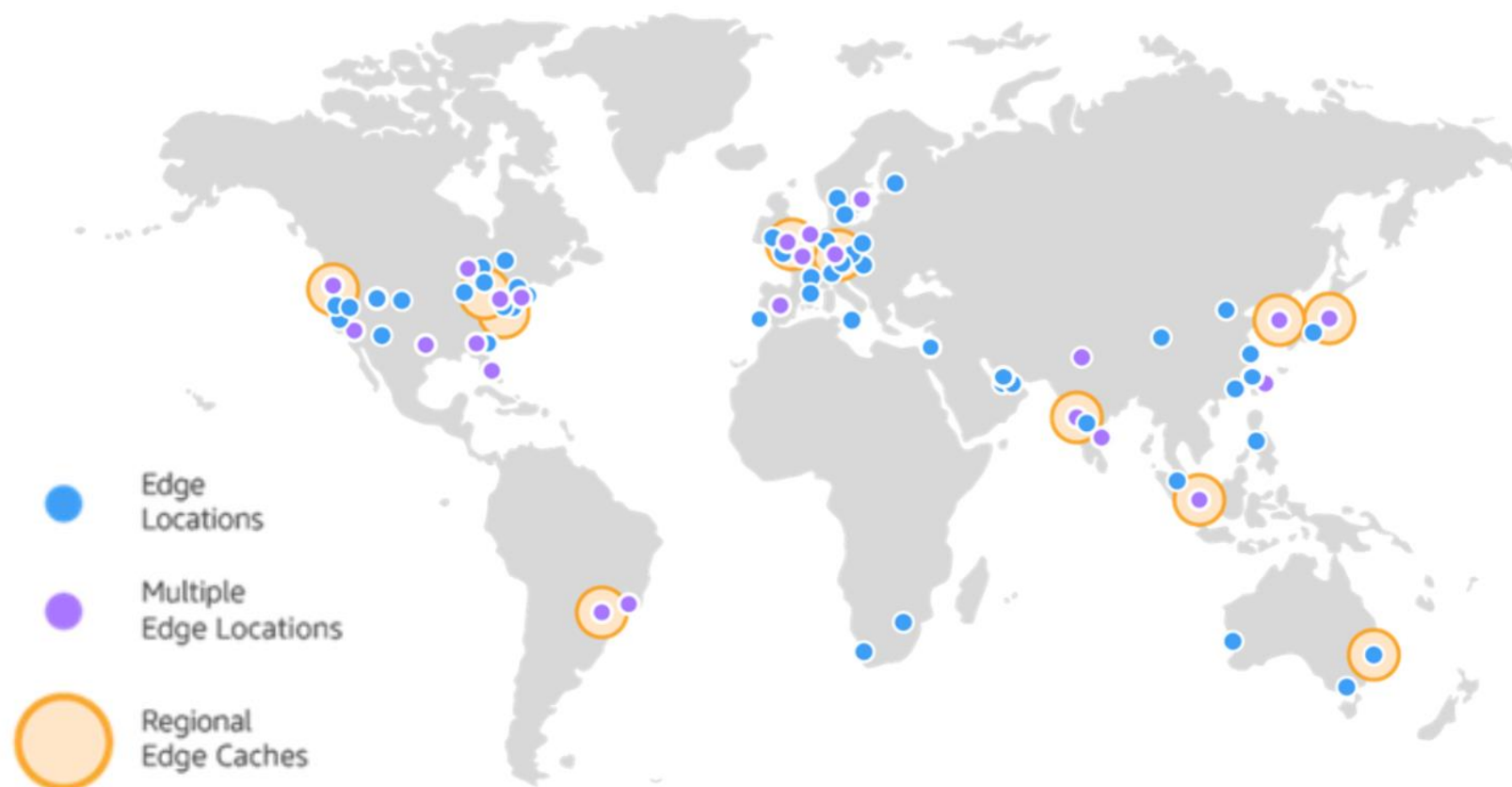
Data centers da AWS

- Os data centers da AWS são **projetados para segurança**.
- Os data centers são onde os dados residem e o processamento de dados ocorre.
- Cada data center tem energia, redes e conectividade redundantes e está hospedado em uma instalação separada.
- Normalmente, um data center tem de 50.000 a 80.000 servidores físicos.



Pontos de presença

- A AWS fornece uma rede global de locais de **pontos de presença**
- Consiste em **locais de borda** e um número muito menor de **caches de borda regionais**
- Usada com o Amazon CloudFront
 - Uma Content Delivery Network (CDN - Rede de entrega de conteúdo) global que entrega conteúdo aos usuários finais com **latência reduzida**
- Os pontos de presença de caches regionais usados para conteúdo com acesso pouco frequente



Serviços fundamentais da AWS

Aplicativos



Desktops virtuais



Colaboração e compartilhamento

Serviços de plataforma

Bancos de dados

Relacional

NoSQL

Armazenamento em cache

Analytics

Computação em clusters

Tempo real

Data warehouse

Fluxos de trabalho de dados

Serviços de aplicações

Enfileiramento

Orquestração

Streaming de aplicativos

Transcodificação

E-mail

Pesquisa

Implantação e gerenciamento

Contêineres

Ferramentas de DevOps

Modelos de recursos

Controle de uso

Monitoramento e logs

Serviços móveis

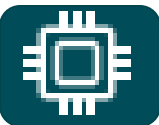
Identidade

Sincronização

Mobile Analytics

Notificações

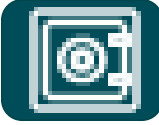
Serviços básicos



Computação (virtual, auto scaling e balanceamento de carga)



Redes



Armazenamento (objeto, bloco e arquivo)

Infraestrutura

Regiões

Zonas de Disponibilidade



Locais de borda

Categorias de serviços da AWS



Analytics



Integração
de aplicações



Realidade aumentada
e realidade virtual



Blockchain



Aplicações
empresariais



Computação



Gerenciamento
de custos



Envolvimento
de clientes



Banco de dados



Ferramentas de
desenvolvedor



Computação
de usuário final



Tecnologia
de jogo



Internet
das Coisas



Machine
Learning



Gerenciamento
e governança



Serviços de mídia



Migração
e transferência



Dispositivos
móveis



Redes e entrega
de conteúdo



Robótica



Satélite



Segurança, identidade
e conformidade



Armazenamento



Parte Prática

Roteiro prático

Roteiro Prático

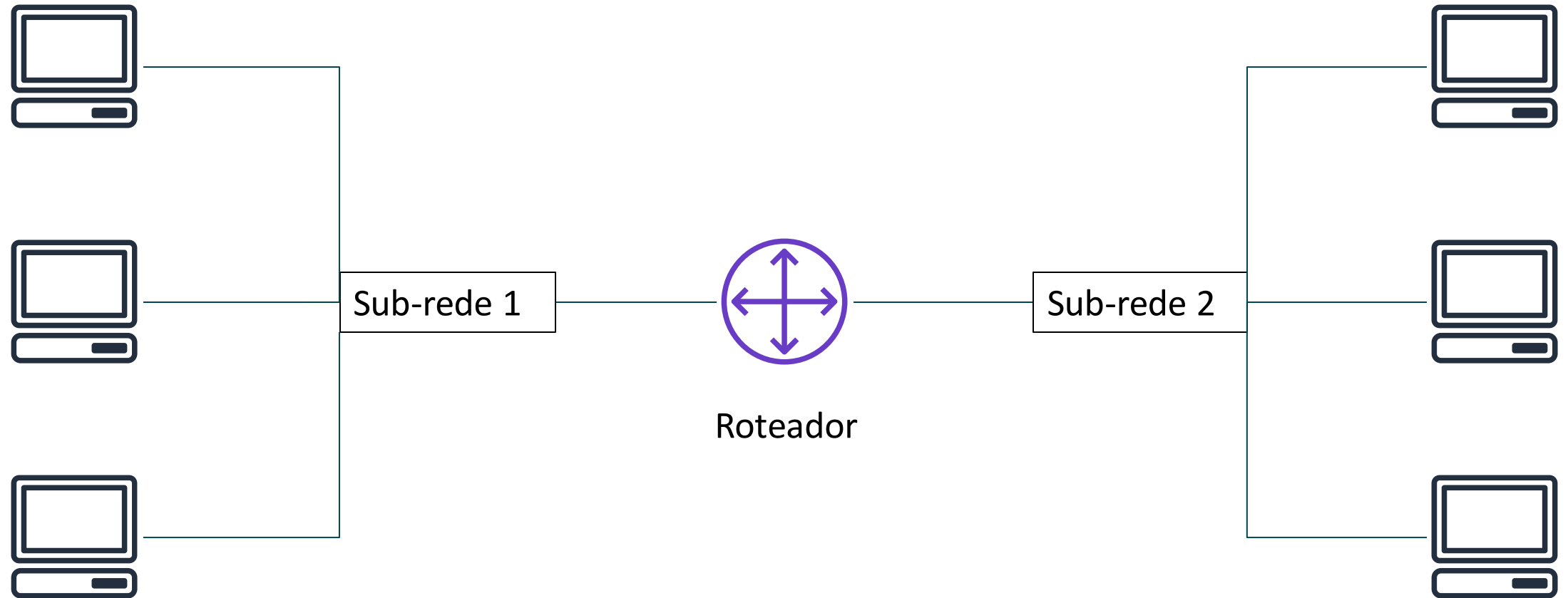


- Repositório do Github
- <https://tinyurl.com/wtads-aws>
- Sugestão é que vocês façam o fork do repositório, caso queiram modificar algum aspecto da aplicação
- Dentro do repositório temos um roteiro completo da parte prática
- Porém antes de iniciar, vocês devem acessar o ambiente do AWS Academy, através do link enviado no e-mail usado para inscrição no Minicurso
- Vamos iniciar o ambiente e dar uma visão geral

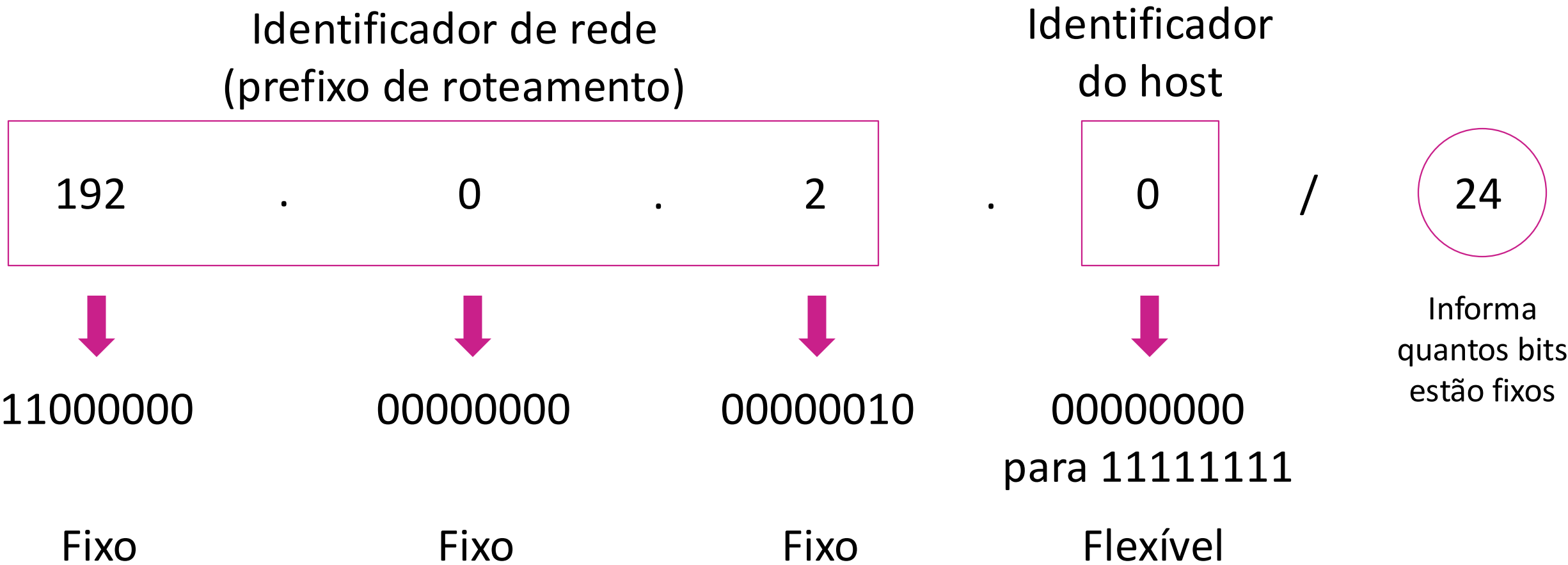
Seção 4: Noções básicas de redes

Redes VPC

Redes



Roteamento sem classe entre domínios (CIDR)



Seção 5: Amazon VPC

Redes VPC



Amazon VPC

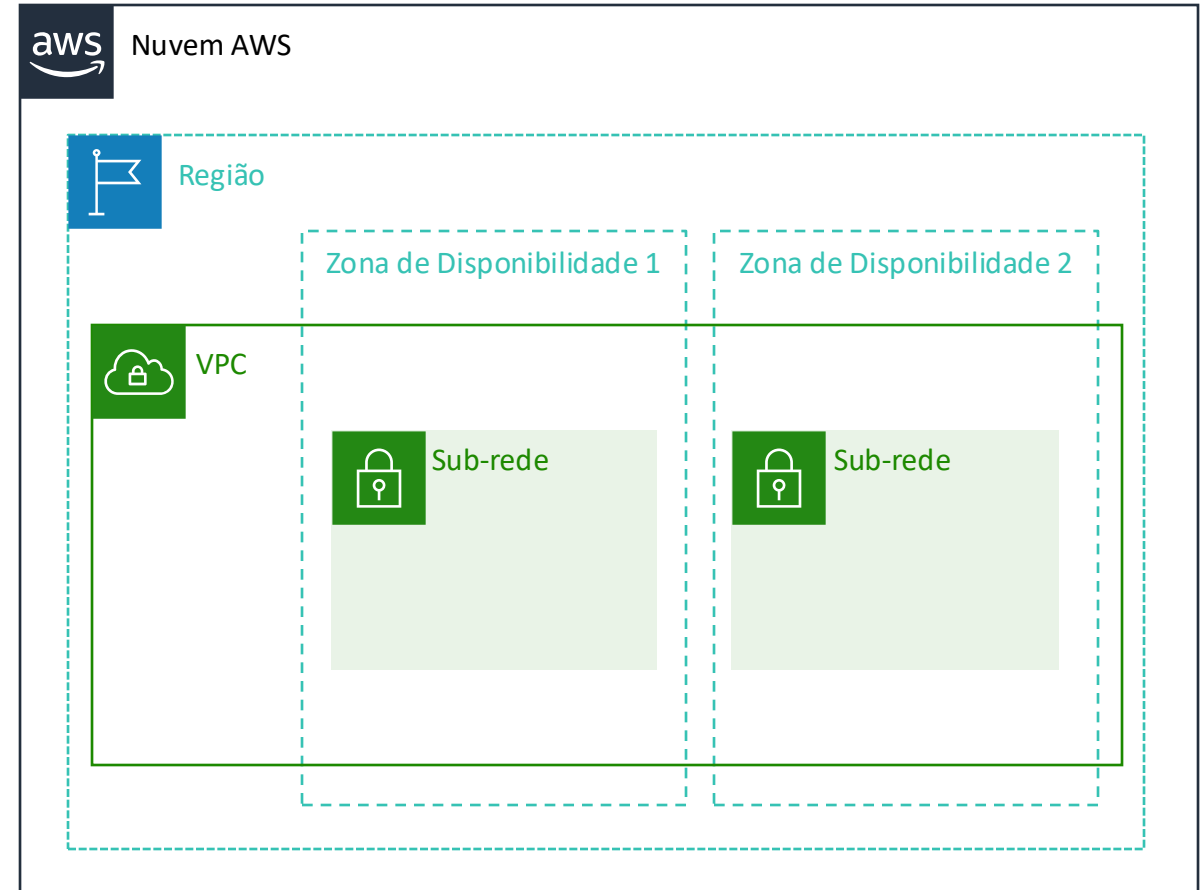


Amazon
VPC

- Permite provisionar uma seção **isolada logicamente** da Nuvem AWS onde você pode iniciar recursos da AWS em uma rede virtual que você mesmo define
- Fornece **controle sobre seus recursos de rede virtual**, incluindo:
 - Seleção do intervalo de endereços IP
 - Criação de sub-redes
 - Configuração de tabelas de rotas e gateways de rede
- Permite **personalizar a configuração de rede** para sua VPC
- Permite usar **várias camadas de segurança**


VPCs e sub-redes

- VPCs:
 - Logicamente isoladas de outras VPCs
 - Dedicadas à sua conta da AWS
 - Pertencem a uma única Região da AWS e podem abranger várias Zonas de Disponibilidade
- Sub-redes:
 - Intervalo de endereços IP que dividem uma VPC
 - Pertencem a uma única Zona de Disponibilidade
 - Classificadas como públicas ou privadas



Endereçamento IP

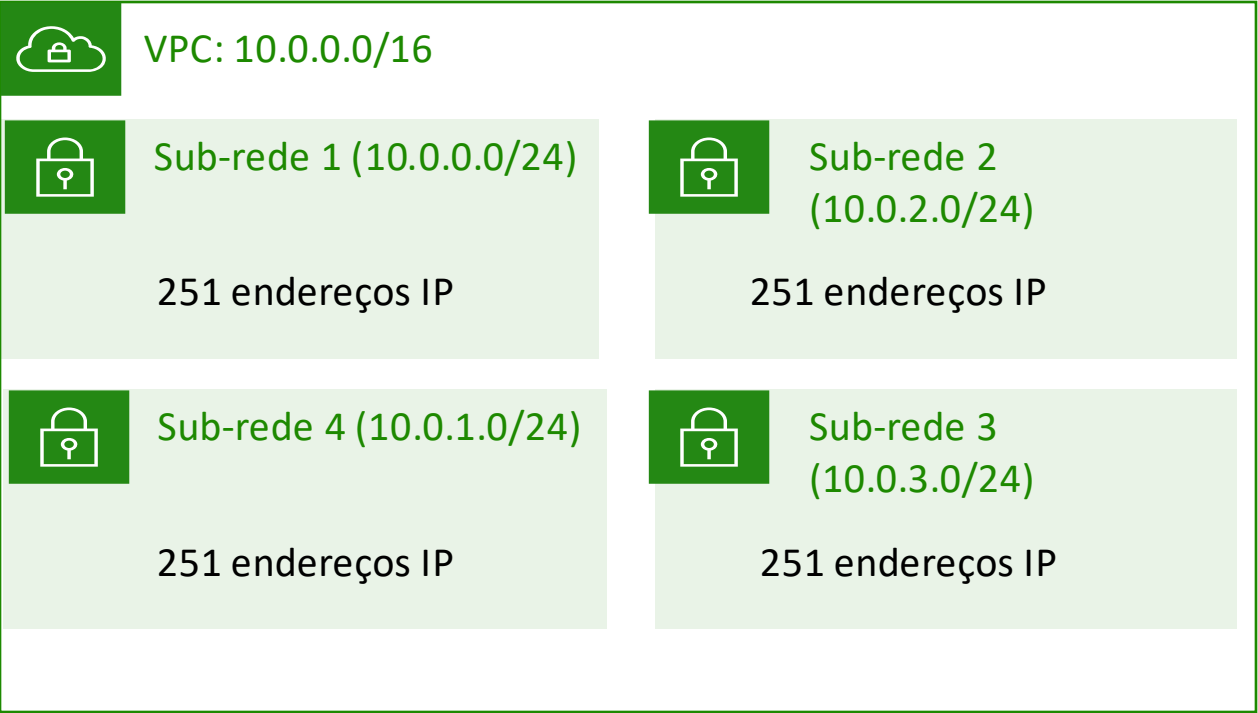
- Ao criar uma VPC, você a atribui a um **bloco CIDR IPv4** (intervalo de endereços IPv4 **privados**).
- Você **não pode alterar o intervalo de endereços** depois de criar a VPC.
- O **maior** tamanho de bloco CIDR IPv4 é **/16**.
- O **menor** tamanho do bloco CIDR IPv4 é **/28**.
- O IPv6 também é compatível (com um limite de tamanho de bloco diferente).
- Os blocos CIDR de sub-redes **não podem se sobrepor**.

 VPC

x.x.x.x/16 ou 65.536 endereços (máximo)
para
x.x.x.x/28 ou 16 endereços (mínimo)

Endereços IP reservados

Exemplo: uma VPC com um bloco CIDR IPv4 de 10.0.0.0/16 tem 65.536 endereços IP no total. A VPC tem quatro sub-redes de tamanho igual. Somente 251 endereços IP estão disponíveis para uso por cada sub-rede.



Endereços IP para o bloco CIDR 10.0.0.0/24	Reservado para
10.0.0.0	Endereço de rede
10.0.0.1	Comunicação interna
10.0.0.2	Resolução do sistema de nomes de domínio (DNS)
10.0.0.3	Uso futuro
10.0.0.255	Endereço de transmissão de rede

Tipos de endereços IP públicos

Endereço IPv4 público

- Atribuído manualmente por meio de um endereço IP elástico
- Atribuído automaticamente por meio das configurações de endereço IP público de atribuição automática no nível da sub-rede

Endereço IP elástico

- Associado a uma conta da AWS
- Pode ser alocado e remapeado a qualquer momento
- Custos adicionais podem ser aplicados

Interface de rede elástica

- Uma interface de rede elástica é uma **interface de rede virtual** que você pode:
 - Anexar a uma instância.
 - Separar da instância e anexar a outra instância para redirecionar o tráfego de rede.
- Os **atributos dela são mantidos** quando são reanexadas a uma nova instância.
- Cada instância em sua VPC tem uma **interface de rede padrão** que recebe um endereço IPv4 privado do intervalo de endereços IPv4 de sua VPC.



Rotas e tabelas de rotas

- Uma **tabela de rota** contém um conjunto de **regras (ou rotas)** que você pode configurar para direcionar o tráfego de rede da sub-rede.
- Cada **rota** especifica um destino e um alvo.
- Por padrão, toda tabela de rota contém uma **rota local** para comunicação na VPC.
- Cada **sub-rede** deve estar associada a uma **tabela de rota** (no máximo uma).

Tabela de rota principal (padrão)

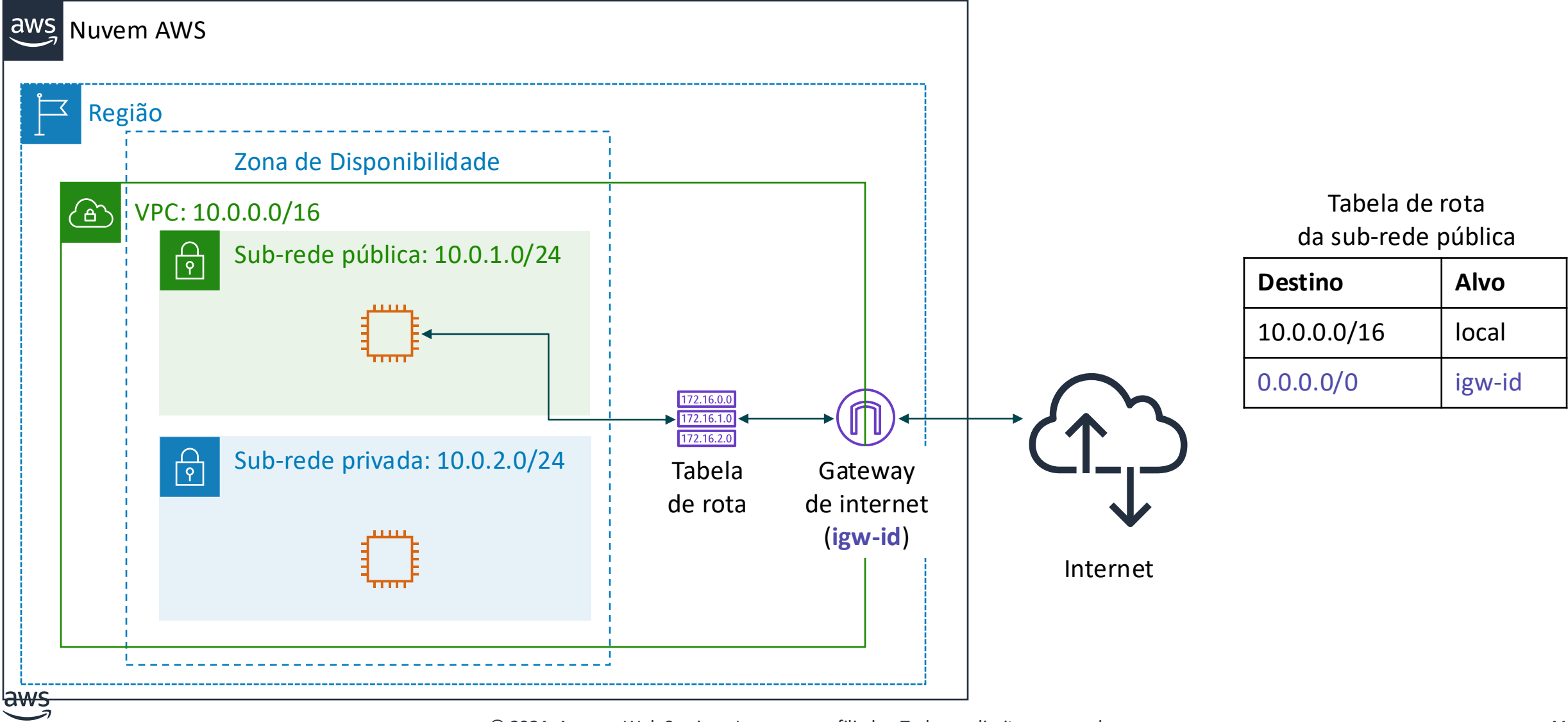
Destino	Alvo
10.0.0.0/16	local

Bloco CIDR da VPC

Seção 6: Gateway de Internet

Redes VPC

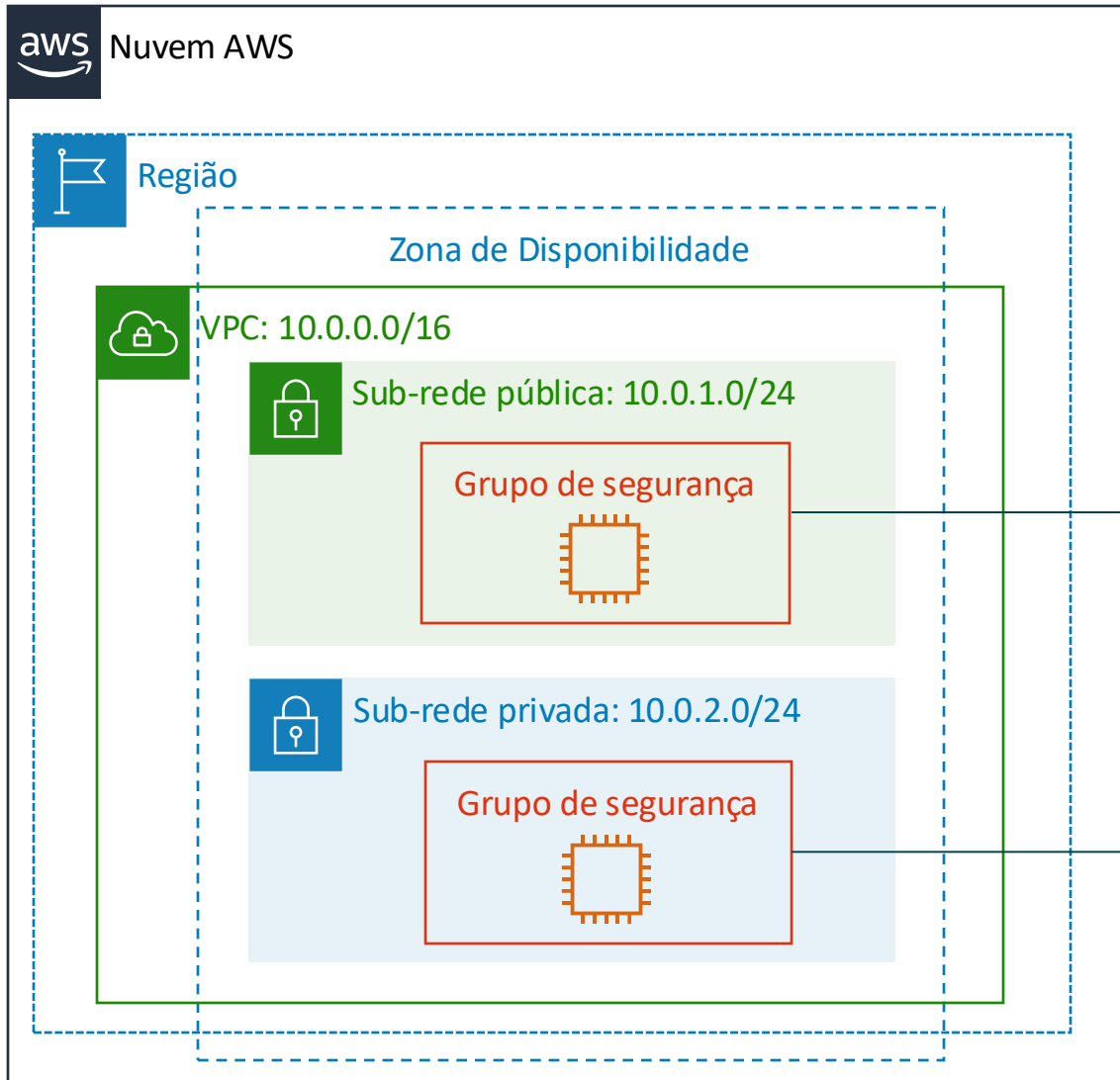
Gateway de internet



Seção 7: Segurança da VPC

Redes VPC

Grupos de segurança (1 de 2)



Grupos de segurança atuam no nível da instância.

Grupos de segurança (2 de 2)

- Os grupos de segurança têm **regras** que controlam o tráfego de entrada e de saída da instância.
- Os grupos de segurança padrão **negam todo o tráfego de entrada** e **permitem todo o tráfego de saída**.
- Os grupos de segurança são **stateful**.

Entrada			
Origem	Protocolo	Intervalo de portas	Descrição
sg-xxxxxxxx	Tudo	Tudo	Permitir tráfego de entrada de interfaces de rede atribuídas ao mesmo grupo de segurança.

Saída			
Destino	Protocolo	Intervalo de portas	Descrição
0.0.0.0/0	Tudo	Tudo	Permitir todo o tráfego IPv4 de saída.
::/0	Tudo	Tudo	Permitir todo o tráfego IPv6 de saída.



Exemplos personalizados de grupos de segurança

- Você pode **especificar regras de permissão**, mas não de negação.
- **Todas as regras são avaliadas** antes da decisão de permitir o tráfego.

Entrada			
Origem	Protocolo	Intervalo de portas	Descrição
0.0.0.0/0	TCP	80	Permitir acesso HTTP de entrada de todos os endereços IPv4
0.0.0.0/0	TCP	443	Permitir acesso HTTPS de entrada de todos os endereços IPv4
Intervalo de endereços IP públicos da sua rede	TCP	22	Permitir acesso SSH de entrada para instâncias Linux de endereços IP IPv4 na sua rede (pelo gateway de internet)
Saída			
Destino	Protocolo	Intervalo de portas	Descrição
O ID do grupo de segurança para os servidores do banco de dados do Microsoft SQL Server	TCP	1433	Permitir o acesso do Microsoft SQL Server de saída a instâncias no grupo de segurança especificado

Seção 8: Amazon EC2

Serviço de Computação

Amazon Elastic Compute Cloud (Amazon EC2)

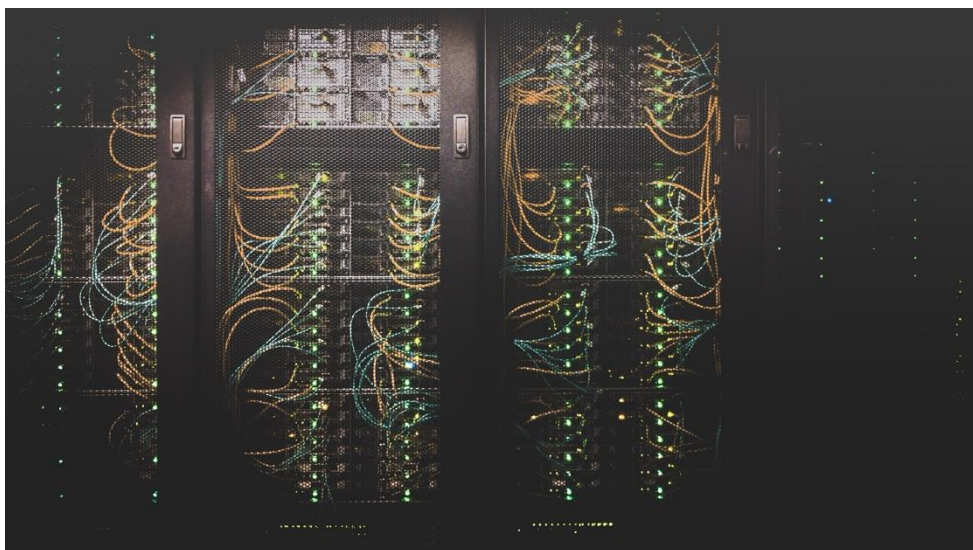
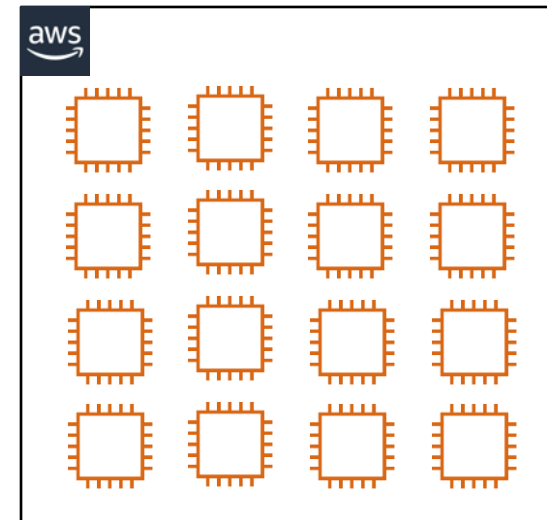


Foto de Taylor Vick no Unsplash

Servidores locais

Exemplos de uso de instâncias do Amazon EC2

- ✓ Servidor de aplicações
- ✓ Servidor web
- ✓ Servidor de banco de dados
- ✓ Servidor de jogos
- ✓ Servidor de e-mail
- ✓ Servidor de mídia
- ✓ Servidor de catálogo
- ✓ Servidor de arquivos
- ✓ Servidor de computação
- ✓ Servidor de proxy



Instâncias do Amazon EC2

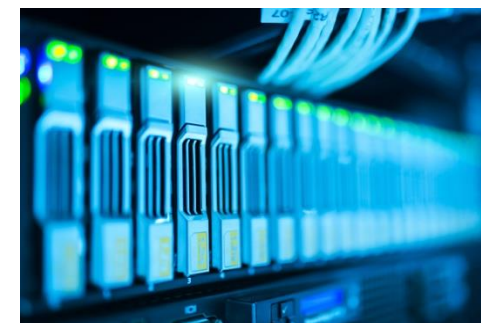


Foto de panumas nikhomkhai do Pexels

Visão geral do Amazon EC2



Amazon
EC2

- **Amazon Elastic Compute Cloud (Amazon EC2)**
 - Fornece **máquinas virtuais**, chamadas de **instâncias do EC2**, na nuvem.
 - Oferece *controle total* sobre o sistema operacional convidado (Windows ou Linux) de cada instância.
- É possível iniciar instâncias de qualquer tamanho em uma Zona de Disponibilidade em qualquer lugar do mundo.
 - Inicie instâncias em **imagens de máquina da Amazon (AMIs)**.
 - Inicie instâncias com apenas alguns cliques ou uma linha de código e elas estarão prontas em minutos.
- Você pode controlar o tráfego de e para instâncias.

Denominação e tamanhos de tipo de instância do EC2

Denominação de tipo de instância

- Exemplo: **t3.large**
 - T é o nome da família
 - 3 é o número da geração
 - Large é o tamanho

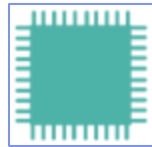
Exemplos de tamanhos de instância

Nome da instância	vCPU	Memória (GB)	Armazenamento
t3.nano	2	0,5	Somente EBS
t3.micro	2	1	Somente EBS
t3.small	2	2	Somente EBS
t3.medium	2	4	Somente EBS
t3.large	2	8	Somente EBS
t3.xlarge	4	16	Somente EBS
t3.2xlarge	8	32	Somente EBS

Selecionar tipo de instância: com base no caso de uso



Uso geral



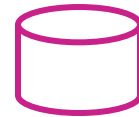
Otimizada para
computação



Otimizada para
memória



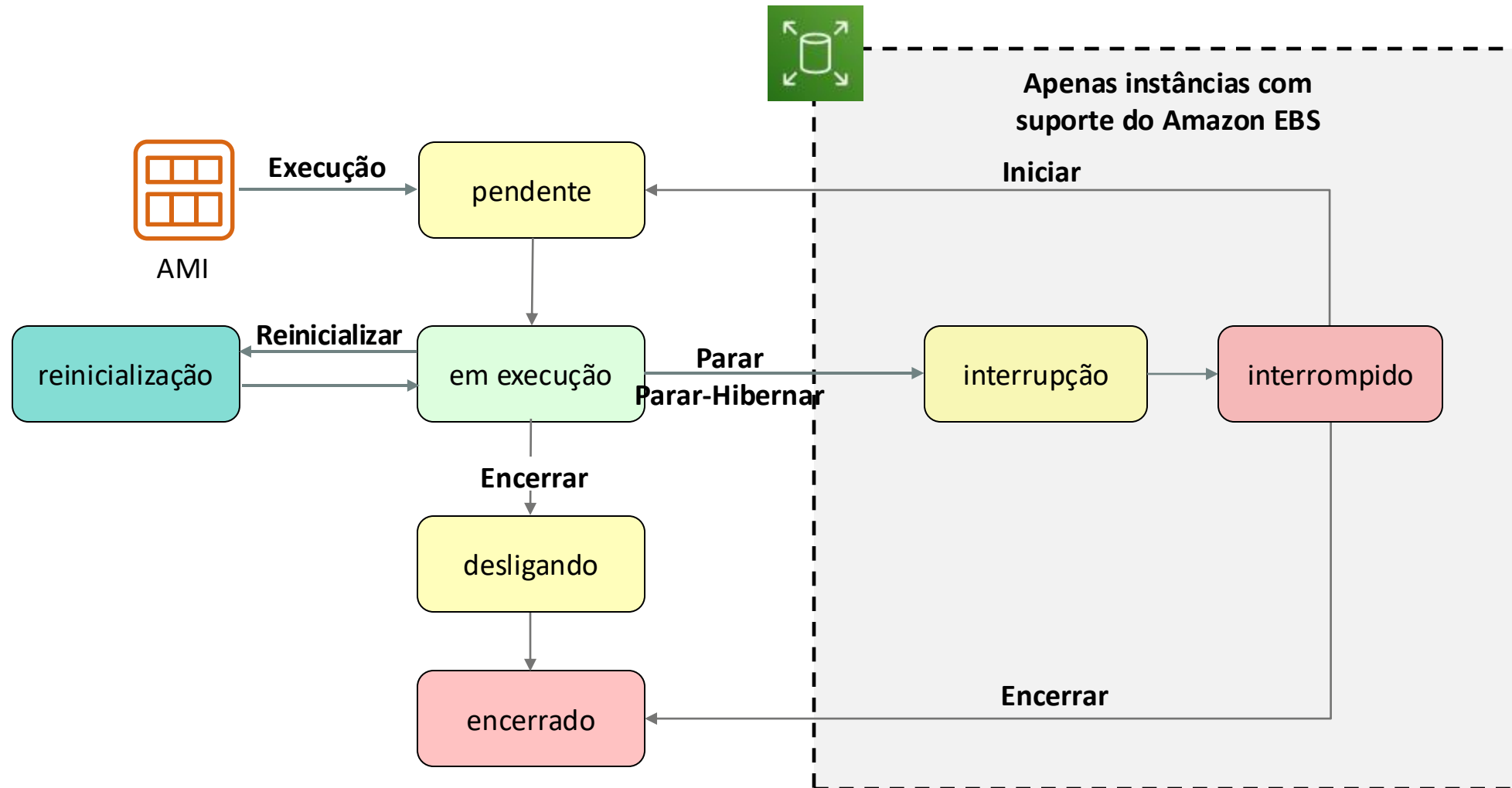
Computação
acelerada



Otimizada para
armazenamento

Tipos de instância	a1, m4, m5, t2, t3	c4, c5	r4, r5, x1, z1	f1, g3, g4, p2, p3	d2, h1, i3
Caso de uso	Ampla	Alto desempenho	Bancos de dados na memória	Machine learning	Sistemas de arquivos distribuídos

Ciclo de vida da instância do Amazon EC2



Considere o uso de um endereço IP elástico

- A **reinicialização** de uma instância *não* alterará endereços IP ou nomes de host DNS.
- Quando uma instância é **interrompida** e, em seguida, **reiniciada**:
 - O endereço IPv4 *público* e o nome do host DNS *externo* serão alterados.
 - O endereço IPv4 *privado* e o nome do host DNS interno *não são* alterados.
- Se você precisar de um endereço IP público persistente:
 - Associe um **endereço IP elástico** à instância.
- Características do endereço IP elástico:
 - Pode ser associado a instâncias da região, conforme necessário.
 - Permanece alocado à sua conta até que você opte por liberá-lo.



Endereço IP
elástico

Amazon CloudWatch para monitoramento

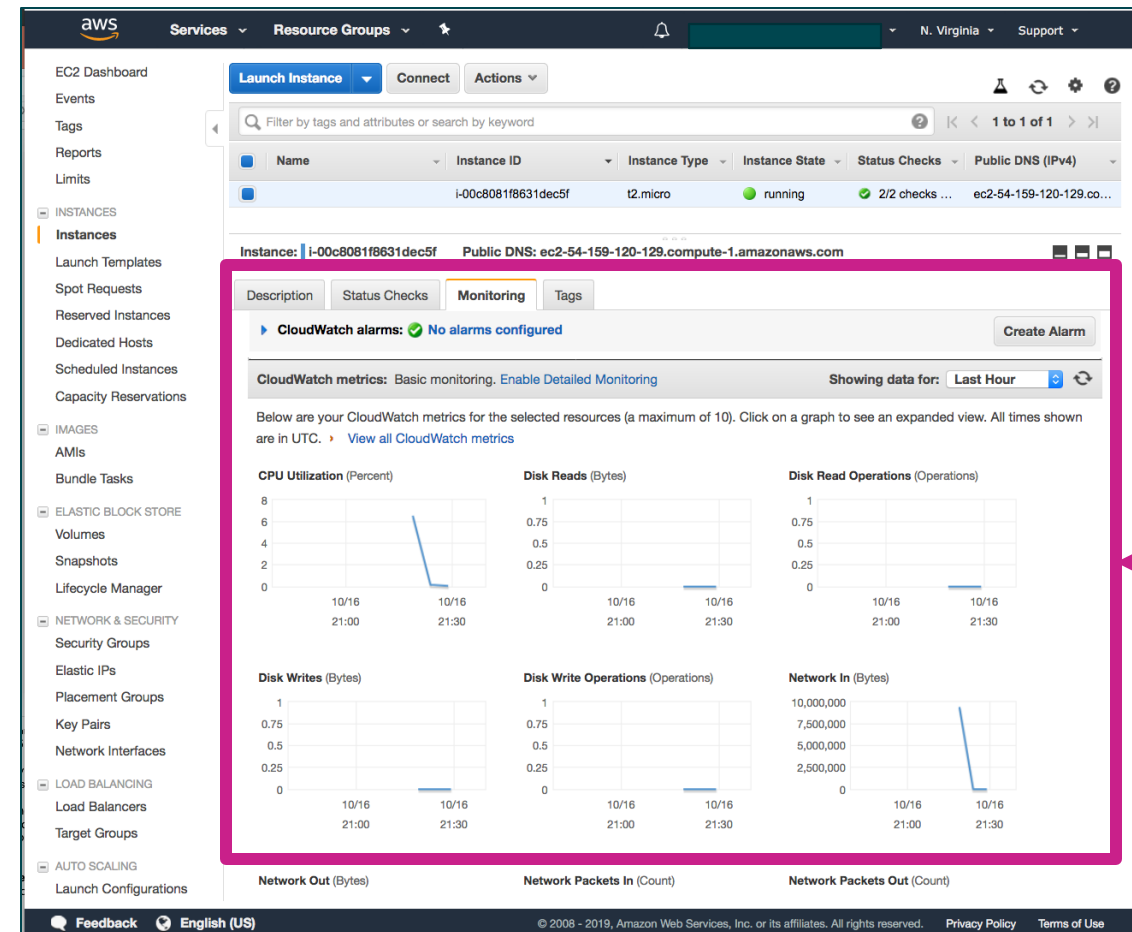
- Use o **Amazon CloudWatch** para monitorar instâncias do EC2
 - Fornece métricas quase em tempo real
 - Fornece gráficos na guia **Monitoramento** do console do Amazon EC2 que você pode visualizar
 - Mantém 15 meses de dados históricos
- **Monitoramento básico**
 - Padrão, sem custo adicional
 - Dados de métrica enviados para o CloudWatch a cada 5 minutos
- **Monitoramento detalhado**
 - Taxa mensal fixa para sete métricas pré-selecionadas
 - Dados de métricas entregues a cada 1 minuto



Amazon CloudWatch



Instância com o CloudWatch



Seção 9: Amazon Relational Database Service

Banco de Dados



Amazon Relational Database Service



Amazon Relational Database Service (Amazon RDS)

Serviços não gerenciados versus serviços gerenciados

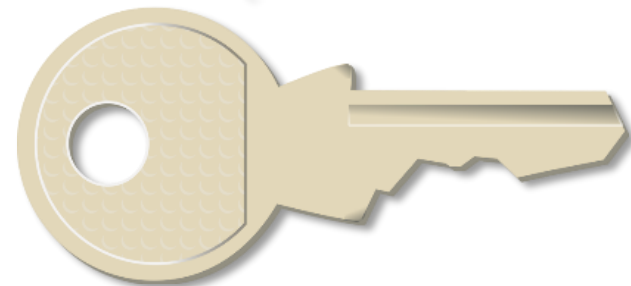
Não gerenciado:

Você gerencia o scaling, a tolerância a falhas e a disponibilidade.

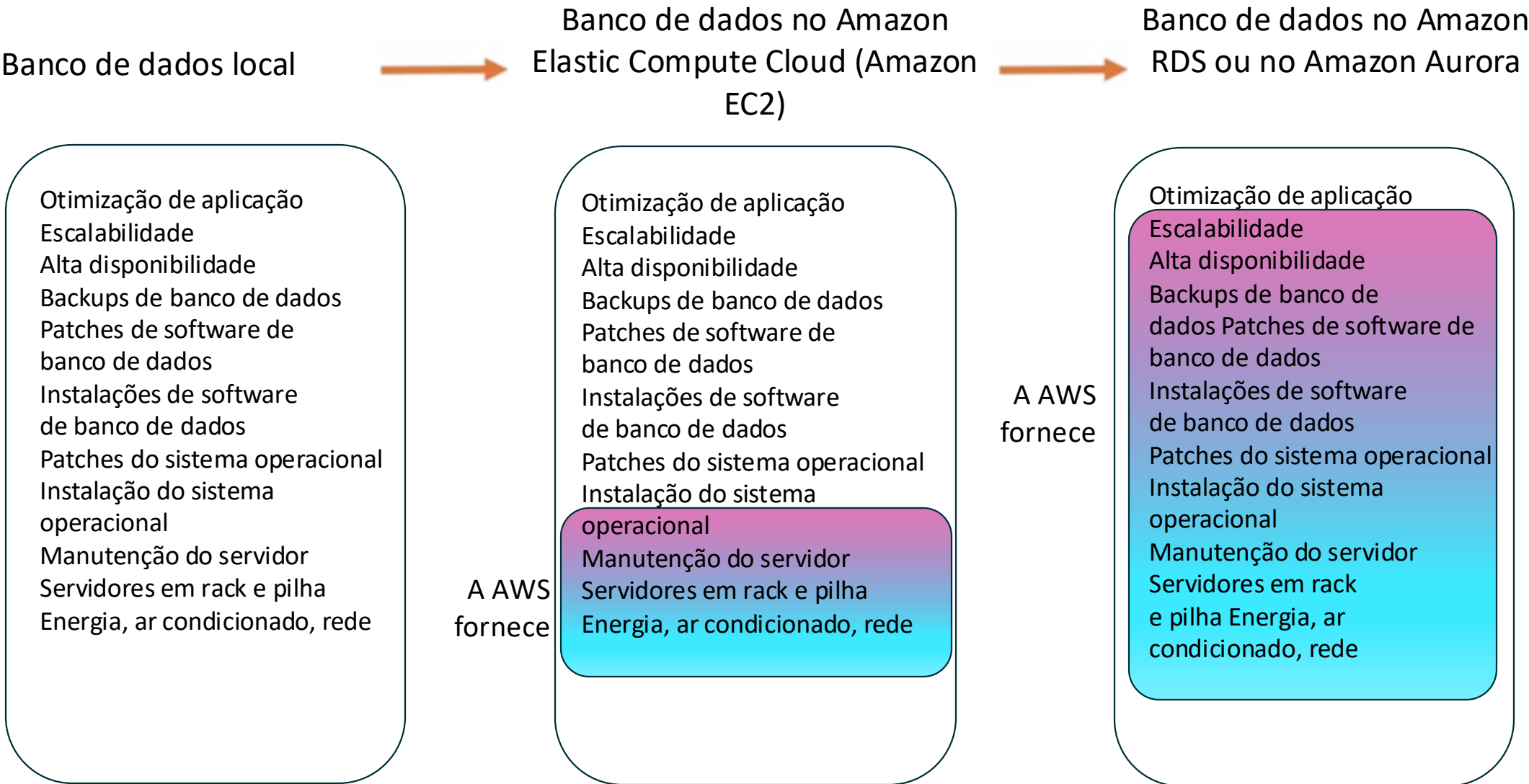


Gerenciados:

O scaling, a tolerância a falhas e a disponibilidade geralmente são incorporados ao serviço.



De bancos de dados on-premises para o Amazon RDS



Instâncias de banco de dados do Amazon RDS

Amazon RDS



Instância principal de
banco de dados do
Amazon RDS

Classe da instância de banco de dados

- CPU
- Memória
- Desempenho de rede

Armazenamento de instâncias de banco de dados

- Magnético
- De uso geral (unidades de estado sólido ou SSD)
- IOPS provisionadas

MySQL

Amazon Aurora

Microsoft SQL Server

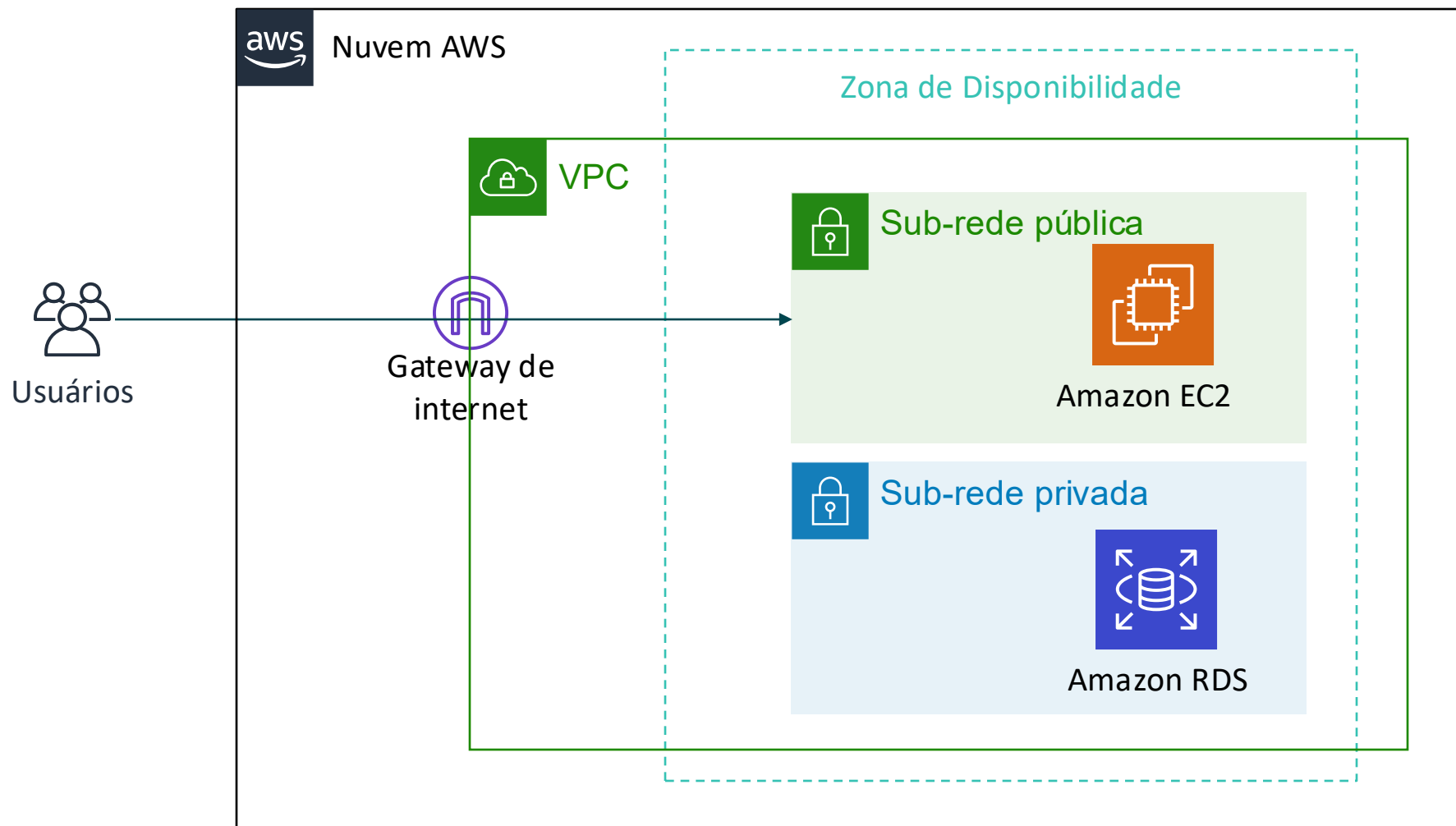
PostgreSQL

MariaDB

Oracle

Mecanismos de banco de dados

Amazon RDS em uma nuvem privada virtual (VPC)



Casos de uso

Aplicativos web e móveis	<ul style="list-style-type: none">✓ Alto vazão✓ Dimensionamento massivo de armazenamento✓ Alta disponibilidade
Aplicativos de comércio eletrônico	<ul style="list-style-type: none">✓ Banco de dados de baixo custo✓ Segurança de dados✓ Solução totalmente gerenciada
Jogos para dispositivos móveis e online	<ul style="list-style-type: none">✓ Aumente a capacidade rapidamente✓ Auto scaling✓ Monitoramento de banco de dados

Quando usar o Amazon RDS

Use o Amazon RDS quando sua aplicação exigir:

- Transações ou consultas complexas
- Uma taxa de consulta ou gravação média a alta - Até 30.000 IOPS (15.000 leituras+15.000 gravações)
- Não mais do que um único nó de operador ou fragmento
- Alta durabilidade

Não use o Amazon RDS quando sua aplicação exigir:

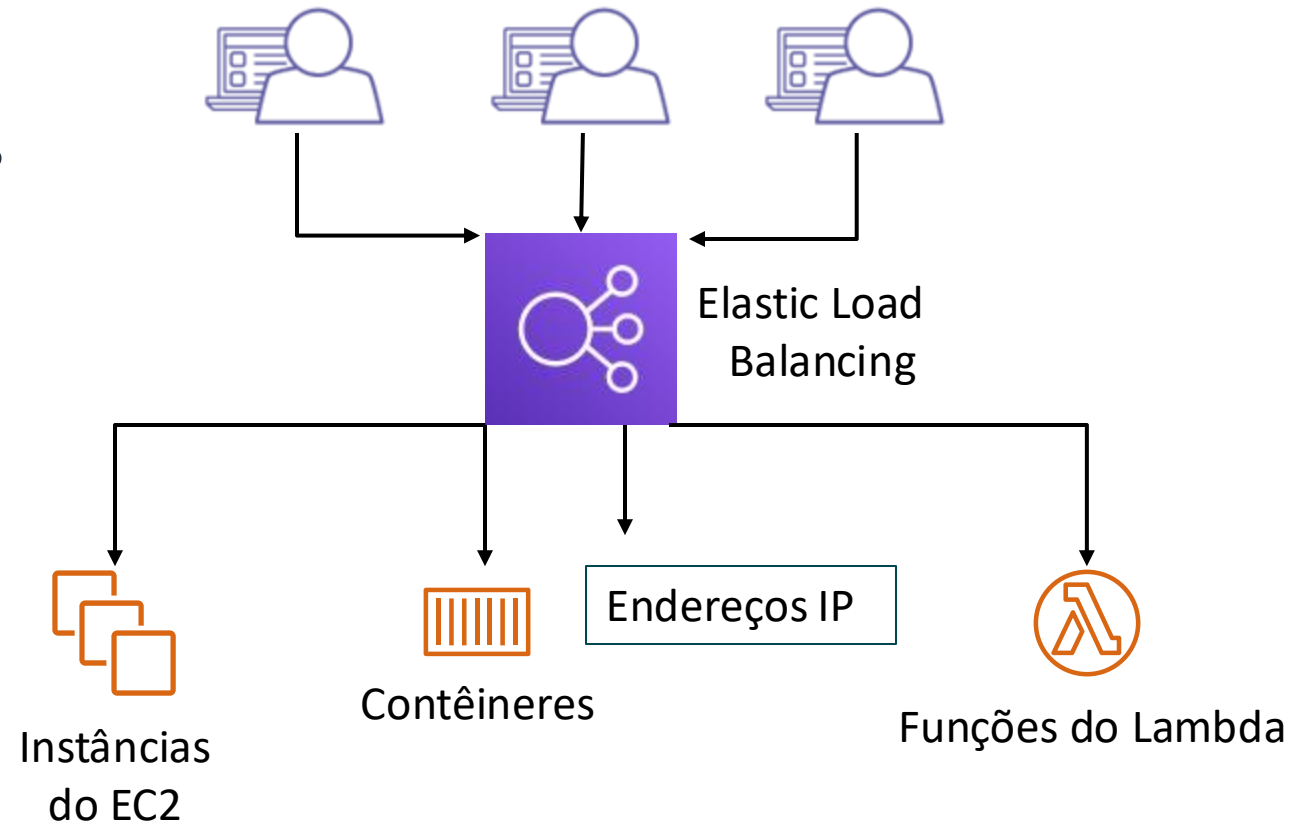
- Taxas massivas de leitura/gravação (por exemplo, 150.000 gravações/segundo)
- Fragmentação devido a altas demandas de throughput ou de volume de dados
- Solicitações e consultas GET ou PUT simples que um banco de dados NoSQL pode processar
- Personalização do sistema de gerenciamento de banco de dados relacional (RDBMS)

Seção 10: Elastic Load Balancing

Balanceamento de Carga

Elastic Load Balancing

- Distribui o tráfego de entrada do aplicativo ou da rede entre vários destinos em uma única zona de disponibilidade ou em várias zonas de disponibilidade.
- Escala seu load balancer à medida que o tráfego para seu aplicativo muda com o tempo.



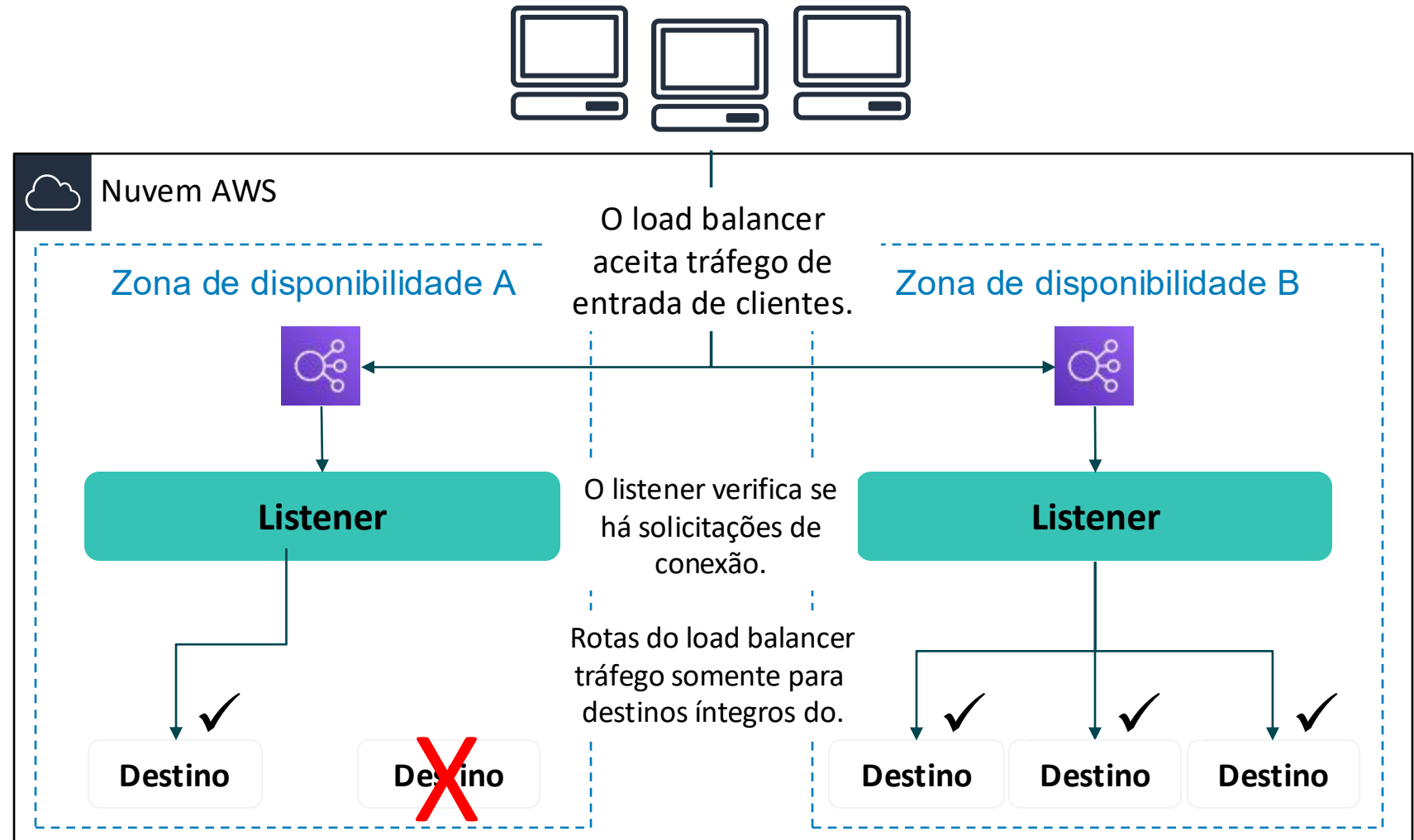
Tipos de load balancers

Application Load Balancer	Network Load Balancer	Classic Load Balancer (geração anterior)
<ul style="list-style-type: none">• Balanceamento de carga avançado de tráfego HTTP e HTTPS	<ul style="list-style-type: none">• Balanceamento de carga de tráfego TCP, UDP e TLS em que haja necessidade de uma performance excepcional	<ul style="list-style-type: none">• Balanceamento de carga de tráfego HTTP, HTTPS, TCP e SSL
<ul style="list-style-type: none">• Roteia o tráfego para os destinos com base no conteúdo da solicitação• Fornece roteamento avançado de solicitações direcionado para a entrega de arquiteturas de aplicativos modernas, incluindo microsserviços e contêineres	<ul style="list-style-type: none">• Roteia o tráfego para os destinos com base nos dados do protocolo IP• Pode processar milhões de solicitações por segundo e ainda manter latências ultrabaixas• É otimizado para lidar com padrões de tráfego súbitos e voláteis	<ul style="list-style-type: none">• Balanceamento de carga entre várias instâncias do EC2
<ul style="list-style-type: none">• Opera na camada de aplicativo (camada 7 do modelo OSI)	<ul style="list-style-type: none">• Opera na camada de transporte (camada 4 do modelo OSI)	<ul style="list-style-type: none">• Opera nas camadas de aplicativo e transporte

Como o Elastic Load Balancing funciona

- Com Application Load Balancers e Network Load Balancers, você **registra destinos em grupos de destino**, e roteia o tráfego para os grupos de destino.
- Com Classic Load Balancers, você **registra instâncias com o load balancer**.

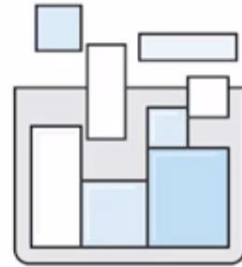
O load balancer executa verificações de integridde para monitorar a integridde dos destinos registrados.



Casos de uso de Elastic Load Balancing



Aplicativos altamente disponíveis e tolerantes a falhas



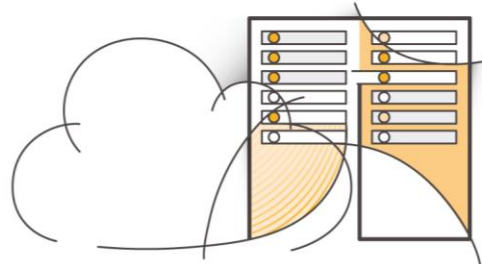
Aplicativos em contêineres



Elasticidade e escalabilidade



Virtual private cloud (VPC)



Ambientes híbridos

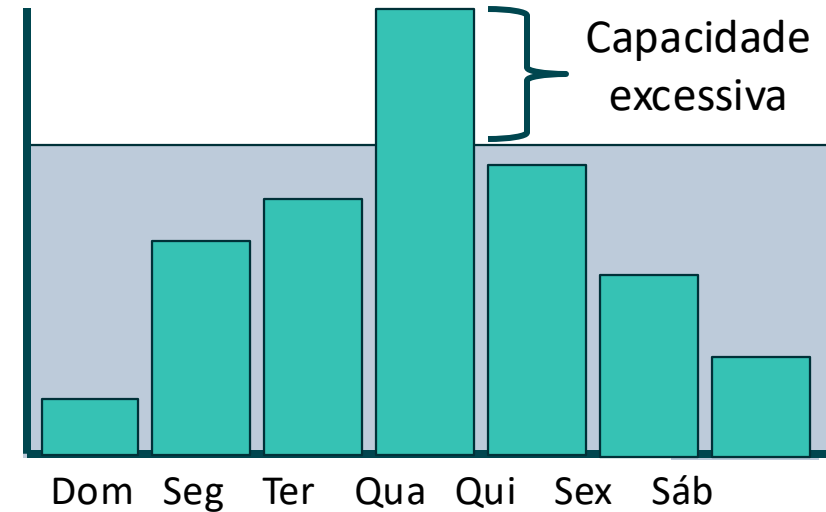
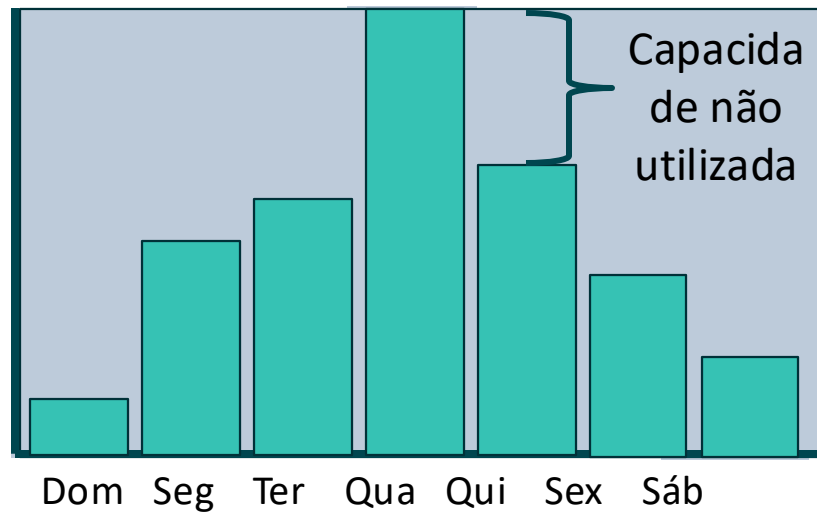


Invocar funções do Lambda por HTTP(S)

Seção 11: Amazon EC2 Auto Scaling

Balanceamento de Carga

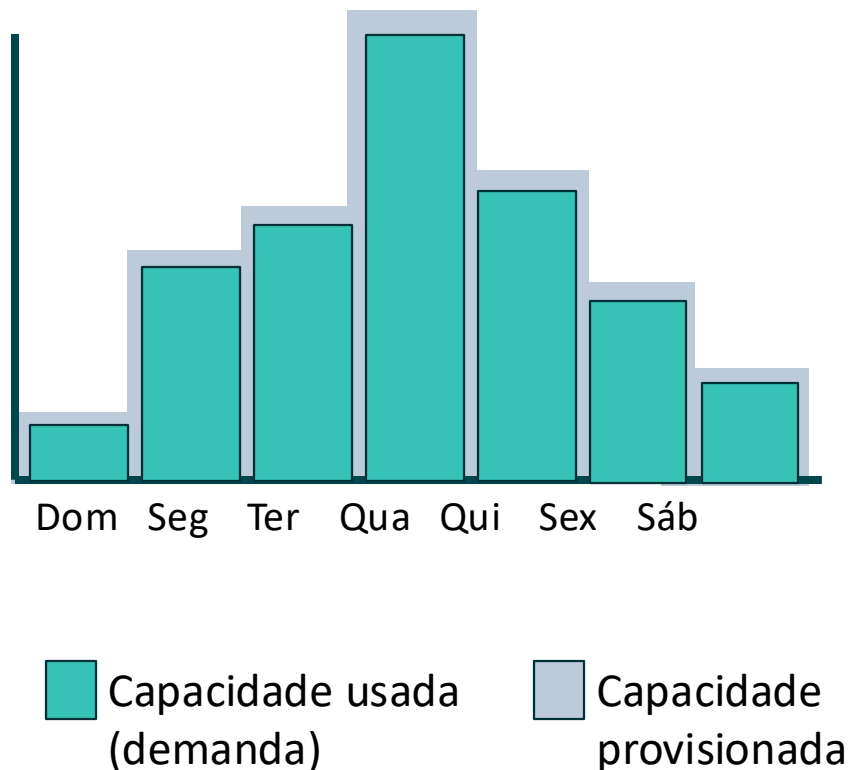
Por que a escalabilidade é importante?



Capacidade usada
(demanda)

Capacidade
provisionada

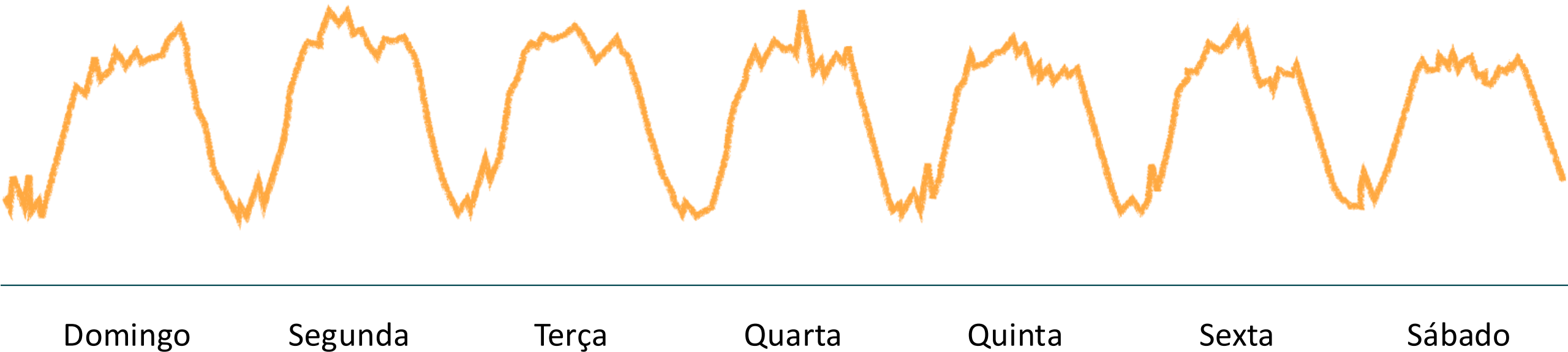
Amazon EC2 Auto Scaling



- Ajuda a manter a disponibilidade do aplicativo
- Permite adicionar ou remover automaticamente instâncias do EC2 de acordo com as condições que você define
- Detecta instâncias do EC2 danificadas e aplicativos não íntegros e substitui as instâncias sem sua intervenção
- Fornece várias opções de escalabilidade - manual, programada, dinâmica ou sob demanda e preditiva

Tráfego semanal típico para o site Amazon.com

Capacidade provisionada



Tráfego de novembro para o site Amazon.com

Capacidade provisionada

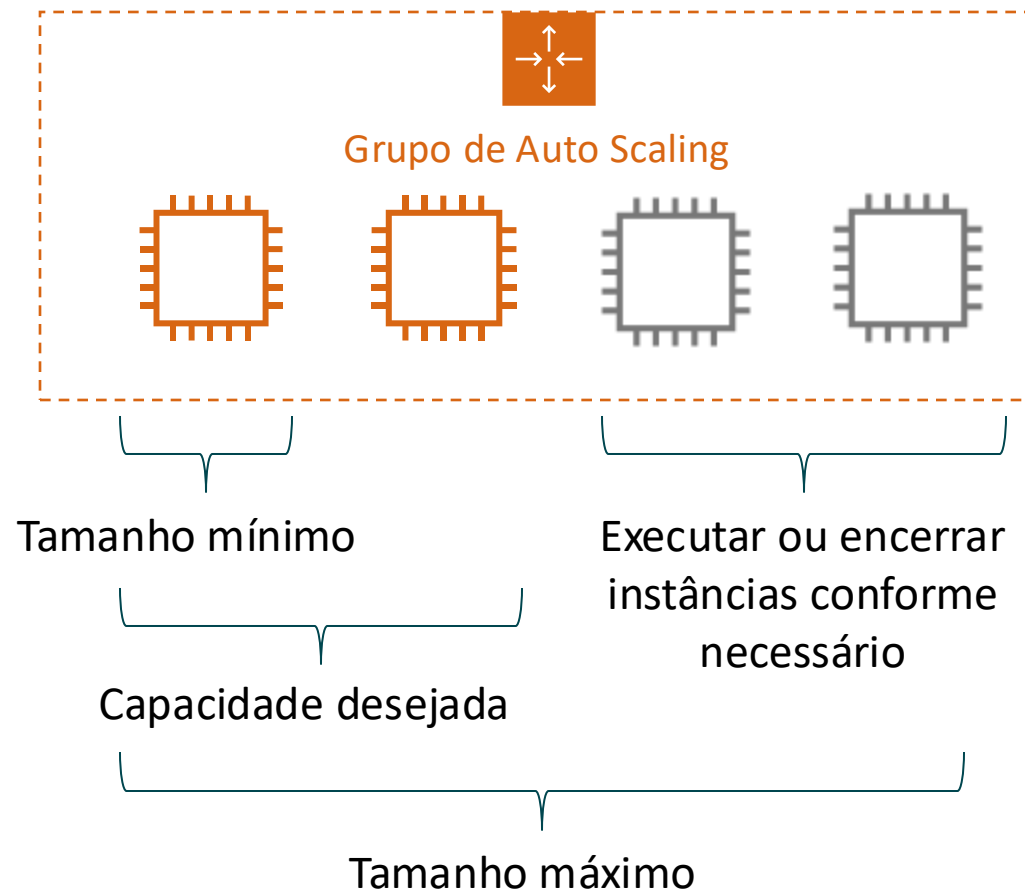
76 %

O desafio é adivinhar eficientemente a quantidade desconhecida de quanta capacidade de computação você precisa.

24 %

Novembro

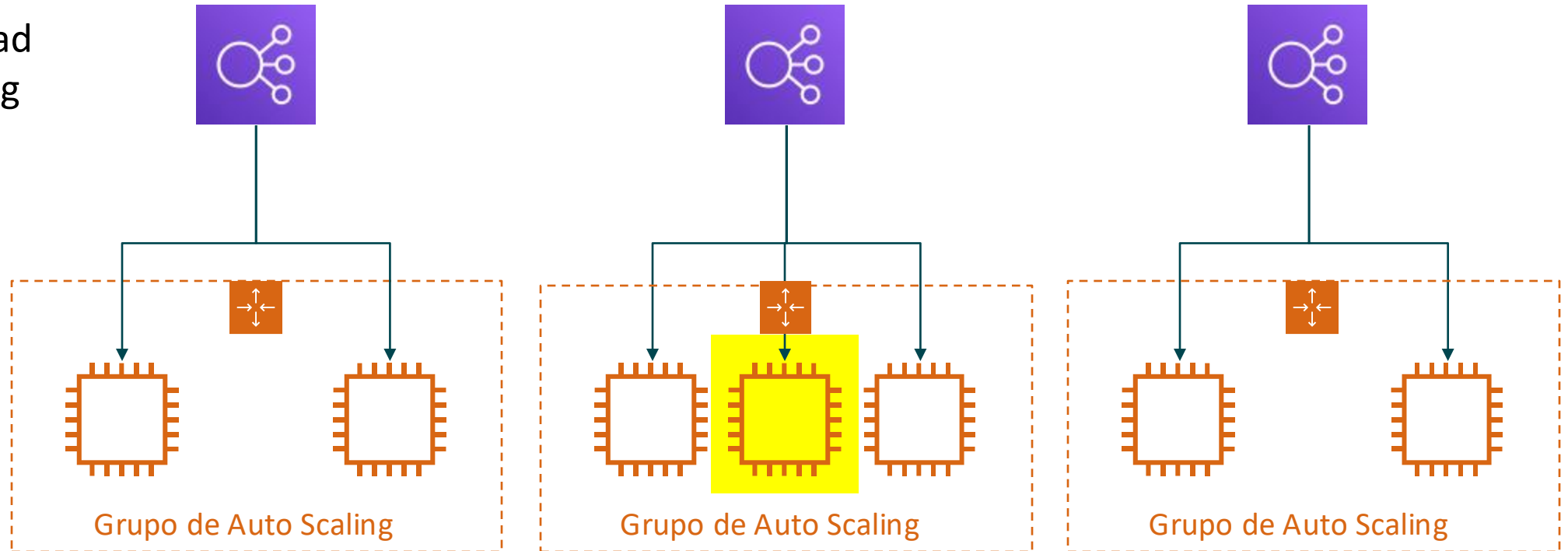
Grupos de Auto Scaling



Um **grupo de Auto Scaling** é um conjunto de instâncias do EC2 que são tratadas como um agrupamento lógico para fins de escalabilidade automática e gerenciamento.

Expansão versus redução

Elastic Load
Balancing

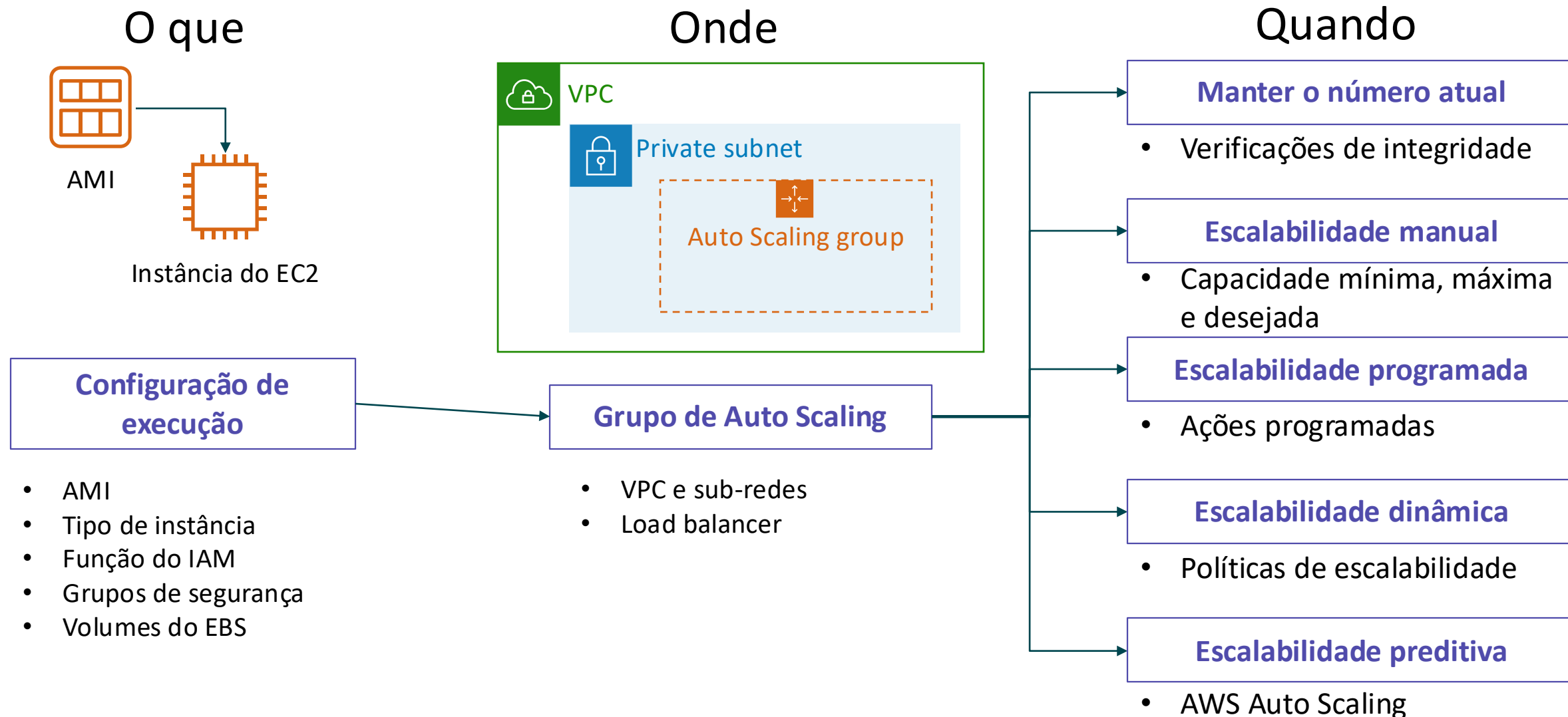


Configuração básica

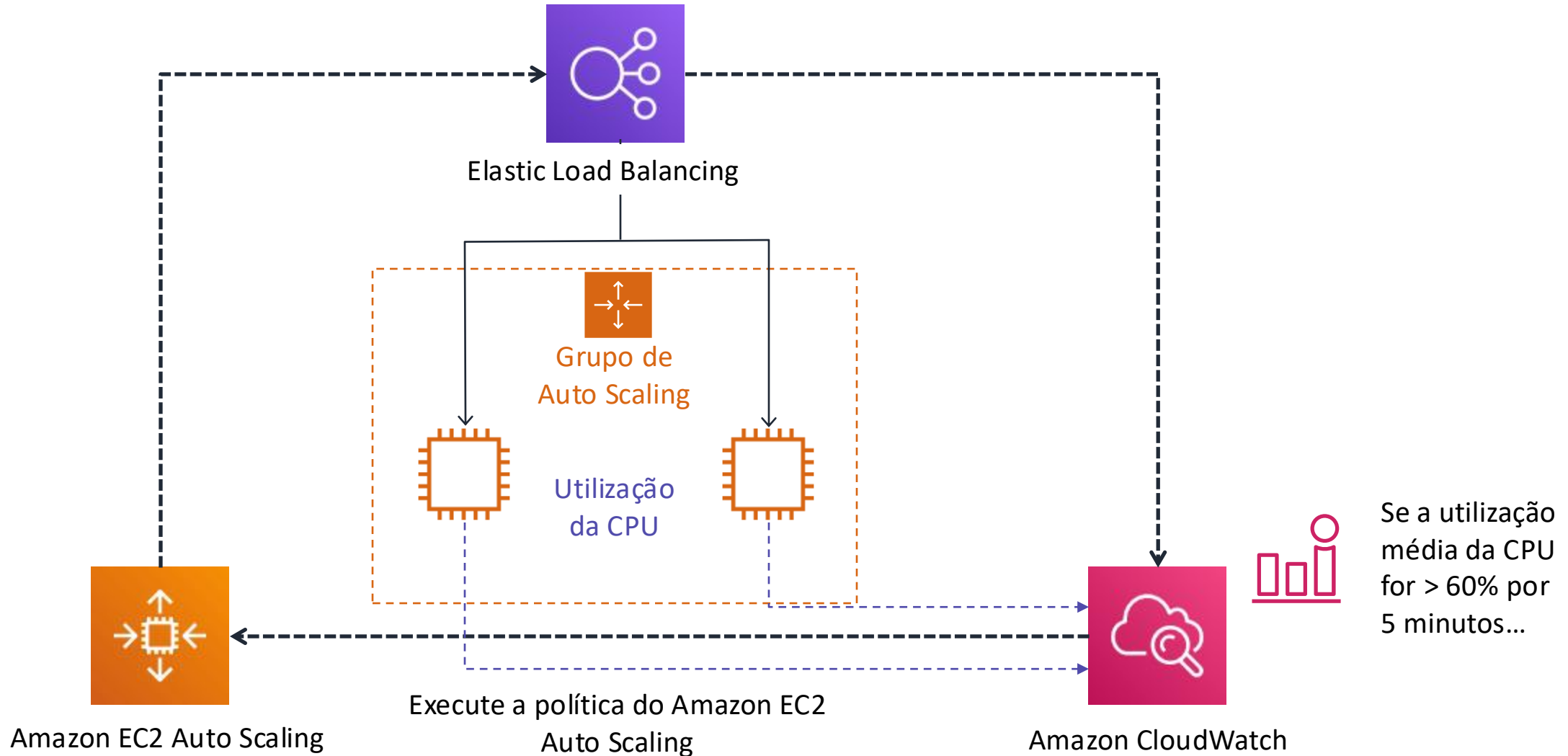
Aumento
(executar instâncias)

Redução
(encerrar instâncias)

Como funciona o Amazon EC2 Auto Scaling



Implementação da escalabilidade dinâmica



AWS Auto Scaling



AWS Auto Scaling

- Monitora os aplicativos e ajusta automaticamente a capacidade para manter uma performance constante e previsível pelo menor custo possível
- Fornece uma interface de usuário simples e eficiente que permite criar planos de escalabilidade para recursos do, incluindo -
 - Instâncias do Amazon EC2 e frotas spot
 - Tarefas do Amazon Elastic Container Service (Amazon ECS)
 - Índices e tabelas do Amazon DynamoDB
 - Réplicas do Amazon Aurora



Obrigado

Questões?

Contato AWS <https://support.aws.amazon.com/#/contacts/aws-training>.