

AWS Academy Cloud Foundations (Fundamentos de nuvem da
AWS Academy)

Módulo 10: Auto Scaling e monitoramento

Tópicos

- Elastic Load Balancing
- Amazon CloudWatch
- Amazon EC2 Auto Scaling

Atividades

- Atividade do Elastic Load Balancing
- Atividade do Amazon CloudWatch

Laboratório

- Escale e faça o balanceamento de carga da arquitetura



Teste de conhecimento

Depois de concluir este módulo, você deverá ser capaz de:

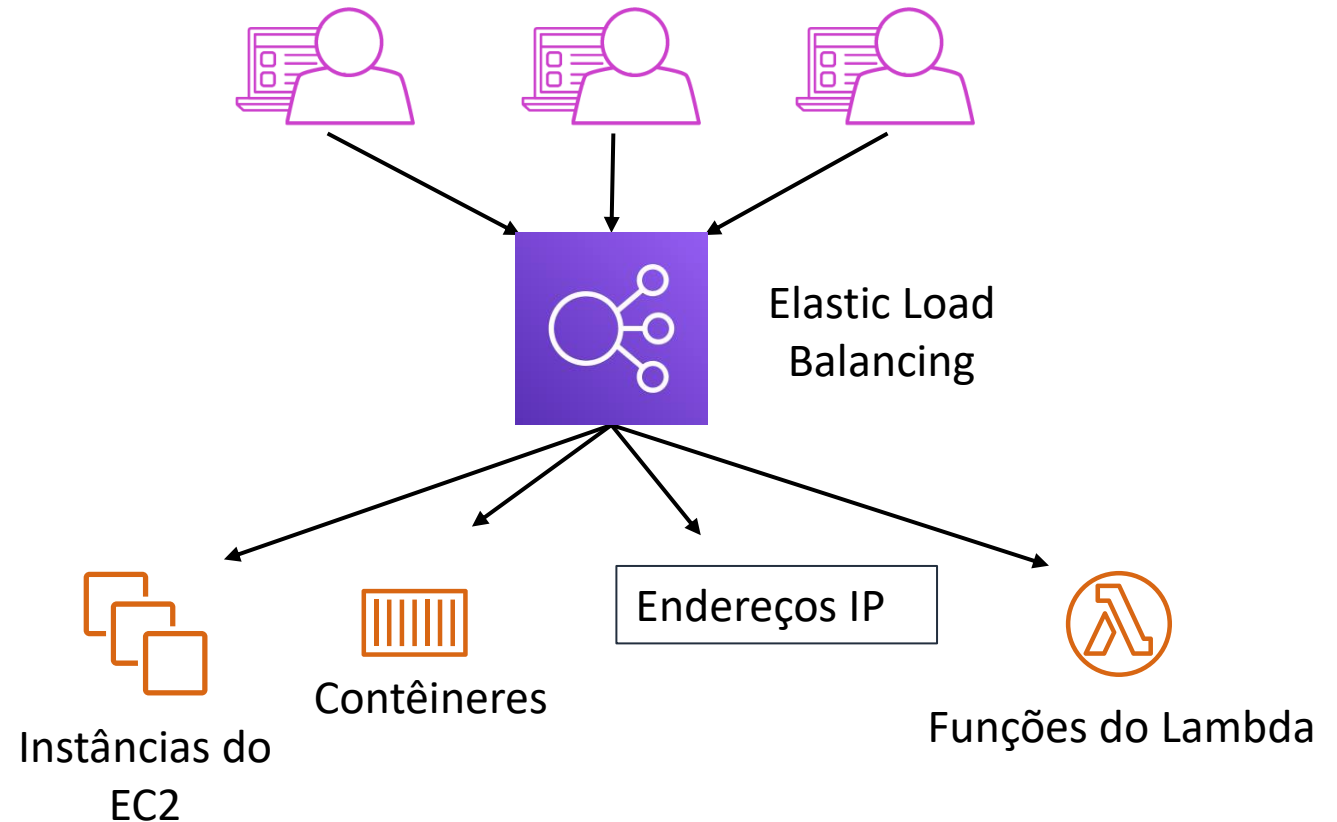
- Indicar como distribuir o tráfego entre instâncias do Amazon Elastic Compute Cloud (Amazon EC2) usando o Elastic Load Balancing
- Identificar como o Amazon CloudWatch permite que você monitore recursos e aplicativos da AWS em tempo real
- Explicar como o Amazon EC2 Auto Scaling é executado e libera servidores em resposta a alterações na carga de trabalho
- Executar tarefas de escalabilidade e balanceamento de carga para melhorar a arquitetura

Módulo 10: Auto Scaling e monitoramento

Seção 1: Elastic Load Balancing

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.



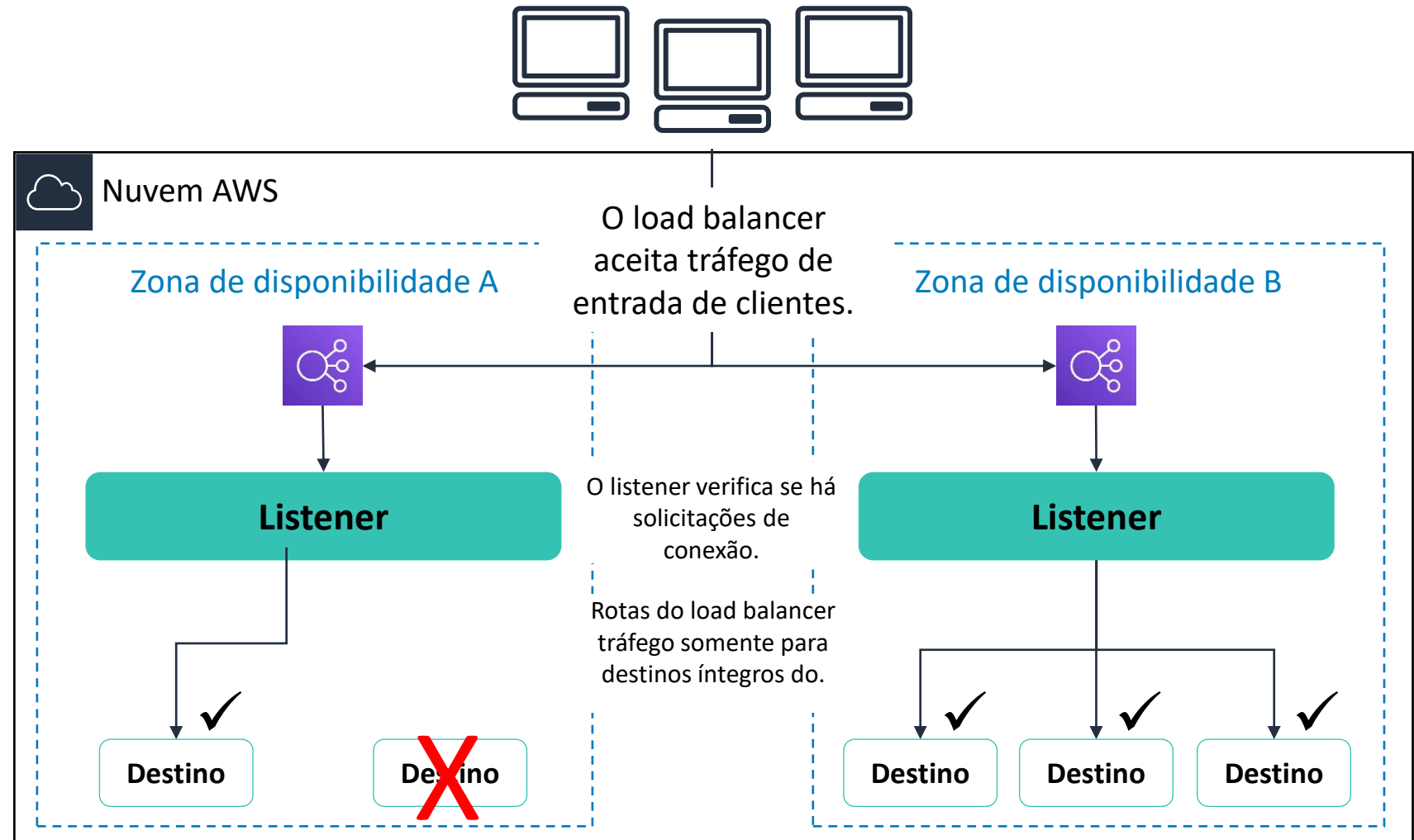
Tipos de load balancers

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**.

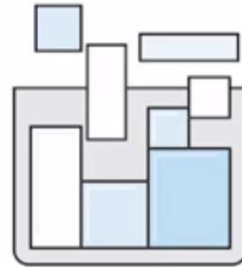
O load balancer executa verificações de integridde para monitorar a integridde dos destinos registrados.



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)

Atividade: Elastic Load Balancing

Você deve oferecer suporte ao tráfego para um aplicativo em contêineres.

Application Load Balancer

Você tem tráfego TCP extremamente variável e imprevisível.

Network Load Balancer

Você precisa de balanceamento de carga simples com vários protocolos.

Classic Load Balancer

Você precisa oferecer suporte a um endereço IP estático ou elástico ou a um destino IP fora de uma VPC.

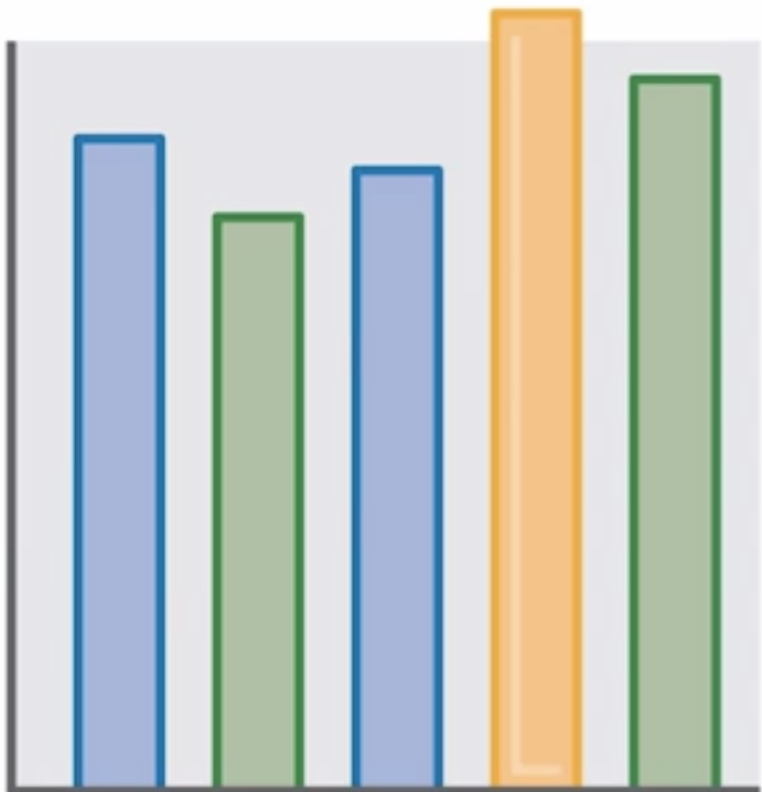
Network Load Balancer

Você precisa de um load balancer que possa lidar com milhões de solicitações por segundo, mantendo baixas latências.

Network Load Balancer

Você deve oferecer suporte a solicitações HTTPS.

Application Load Balancer



- **Métricas do Amazon CloudWatch** - usadas para verificar se o sistema está funcionando conforme o esperado e cria um alarme para iniciar uma ação se uma métrica sair de um intervalo aceitável.
- **Logs de acesso** - capture informações detalhadas sobre solicitações enviadas ao load balancer.
- **Logs do AWS CloudTrail** - Capture quem, o que, quando e onde das interações de API nos serviços da AWS.

Principais lições da Seção 1



- O Elastic Load Balancing distribui o tráfego de entrada de aplicativos ou rede entre vários destinos em uma ou mais zonas de disponibilidade.
- O Elastic Load Balancing oferece suporte a três tipos de load balancers:
 - Application Load Balancer
 - Network Load Balancer
 - Classic Load Balancer
- O ELB oferece verificações de integridade, segurança e monitoramento de instâncias.

Módulo 10: Auto Scaling e monitoramento

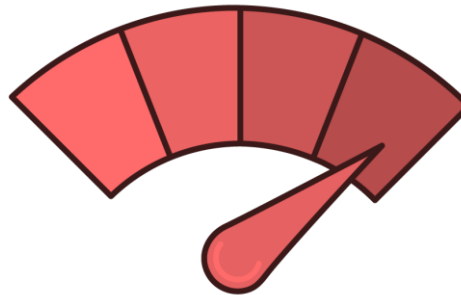
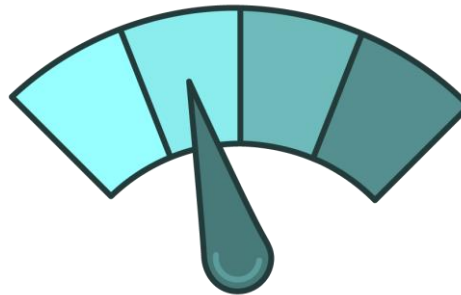
Seção 2: Amazon CloudWatch

Para usar a AWS com eficiência, você precisa de informações sobre seus recursos da AWS:

- Como você sabe quando deve **executar mais instâncias do Amazon EC2?**
- A **performance ou a disponibilidade do seu aplicativo** estão sendo afetadas por falta de capacidade suficiente?
- Quanto da infraestrutura está realmente **sendo usada?**



Amazon CloudWatch



- Monitora –
 - recursos da AWS
 - Aplicativos executados na AWS
- Coleta e rastreia –
 - Métricas padrão
 - Métricas personalizadas
- Alarmes –
 - Enviar notificações para um tópico do Amazon SNS
 - Executar ações do Amazon EC2 Auto Scaling ou do Amazon EC2
- Eventos –
 - Definir regras de acordo com as alterações no ambiente da AWS e rotear esses eventos para um ou mais fluxos ou funções de destino para processamento

Alarmes do CloudWatch

- Criar alarmes com base no -
 - Limite estático
 - Detecção de anomalias
 - Expressão matemática de métricas
- Especificar -
 - Namespace
 - Métrica
 - Estatística
 - Período
 - Condições
 - Configuração adicional
 - Ações

Statistic

Q Average X

Period

5 minutes ▼

Conditions

Threshold type

☒ Static
Use a value as a threshold

☐ Anomaly detection
Use a band as a threshold

Whenever CPUUtilization is...
Define the alarm condition

☒ Greater
> threshold

☐ Greater/Equal
≥ threshold

☐ Lower/Equal
≤ threshold

☐ Lower
< threshold

than...
Define the threshold value

100 ▼

Must be a number

► Additional configuration

Atividade: Amazon CloudWatch



Amazon EC2

Se a utilização média da CPU for $> 60\%$ por 5 minutos...

Correto!



Amazon RDS

Se o número de conexões simultâneas for > 10 por 1 minuto...

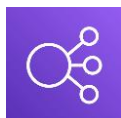
Correto!



Amazon S3

Se o tamanho máximo do bucket em bytes for de 3 por 1 dia...

Incorreto. *Around* não é uma opção de limite. Você deve especificar um limite de $>$, \geq , \leq ou $<$.



Elastic Load Balancing

Se o número de hosts íntegros for < 5 por 10 minutos...

Correto!



Amazon Elastic Block Store

Se o volume de operações de leitura for > 1.000 por 10 segundos...

Incorreto. Você deve especificar uma estatística (por exemplo, *volume médio*).

Principais lições da Seção 2

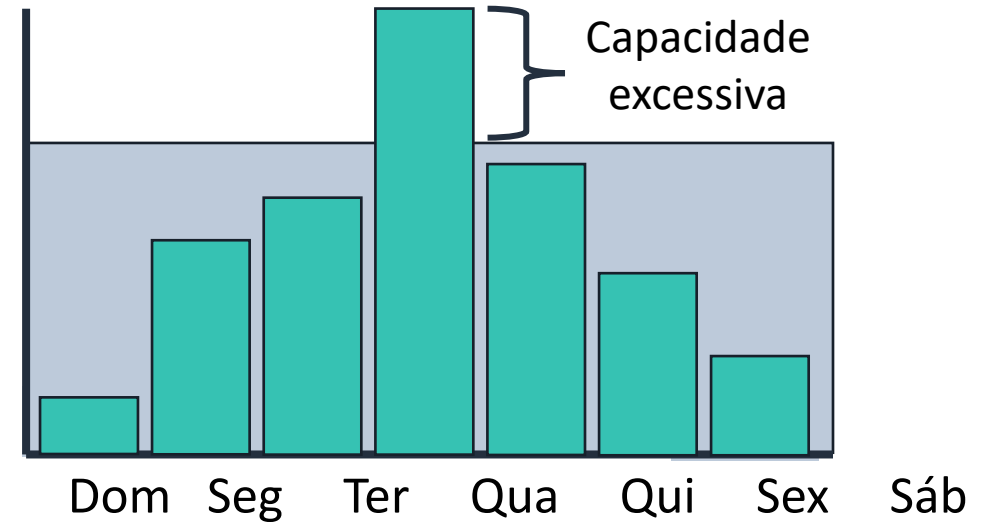
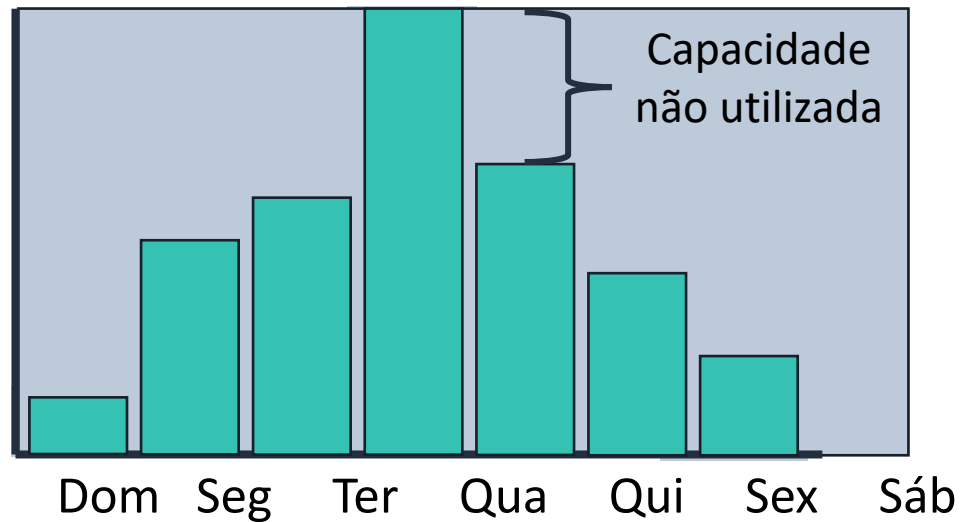


- O Amazon CloudWatch ajuda você a monitorar em tempo real seus recursos da AWS e os aplicativos que você executa na AWS.
- O CloudWatch permite que você:
 - Colete e rastreie métricas padrão e personalizadas.
 - Defina alarmes para enviar notificações automaticamente para tópicos do SNS ou executar ações do Amazon EC2 Auto Scaling ou do Amazon EC2.
 - Defina regras que correspondam a alterações em seu ambiente da AWS e roteie esses eventos para destinos de processamento.

Módulo 10: Auto Scaling e monitoramento

Seção 3: Amazon EC2 Auto Scaling

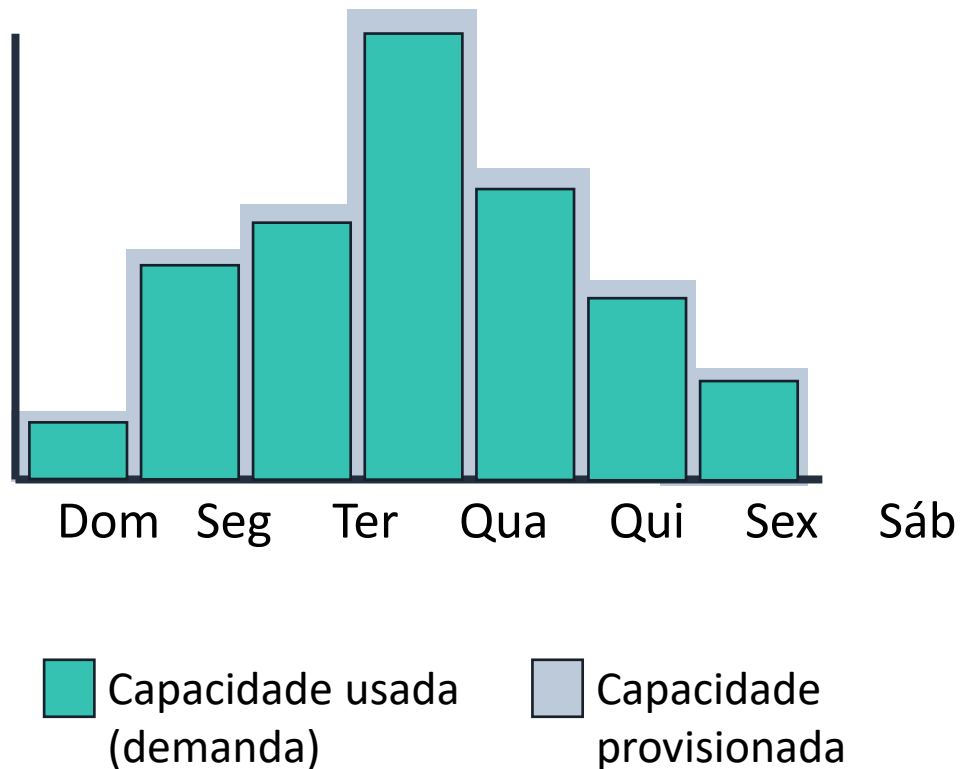
Por que a escalabilidade é importante?



Capacidade usada (demanda)

Capacidade provisionada

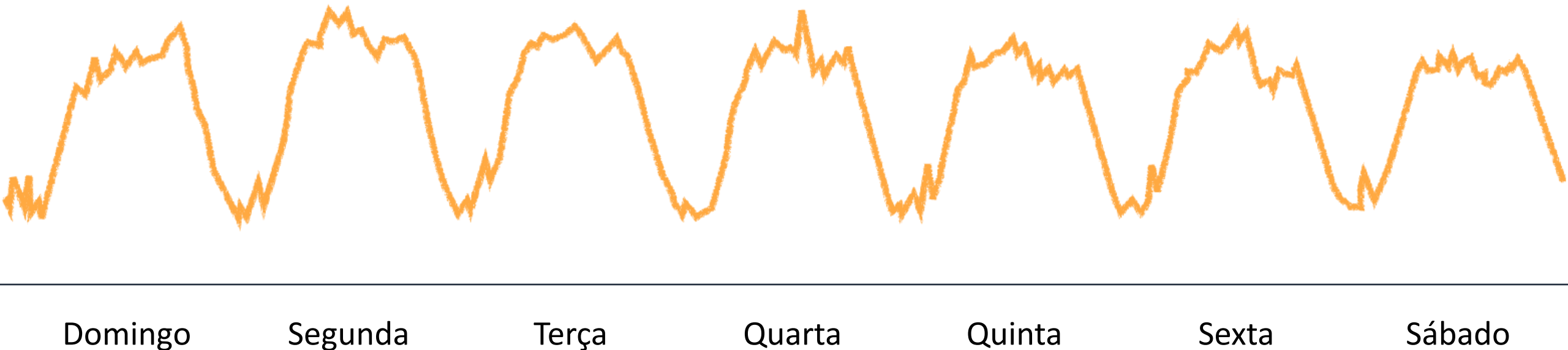
Amazon EC2 Auto Scaling



- 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

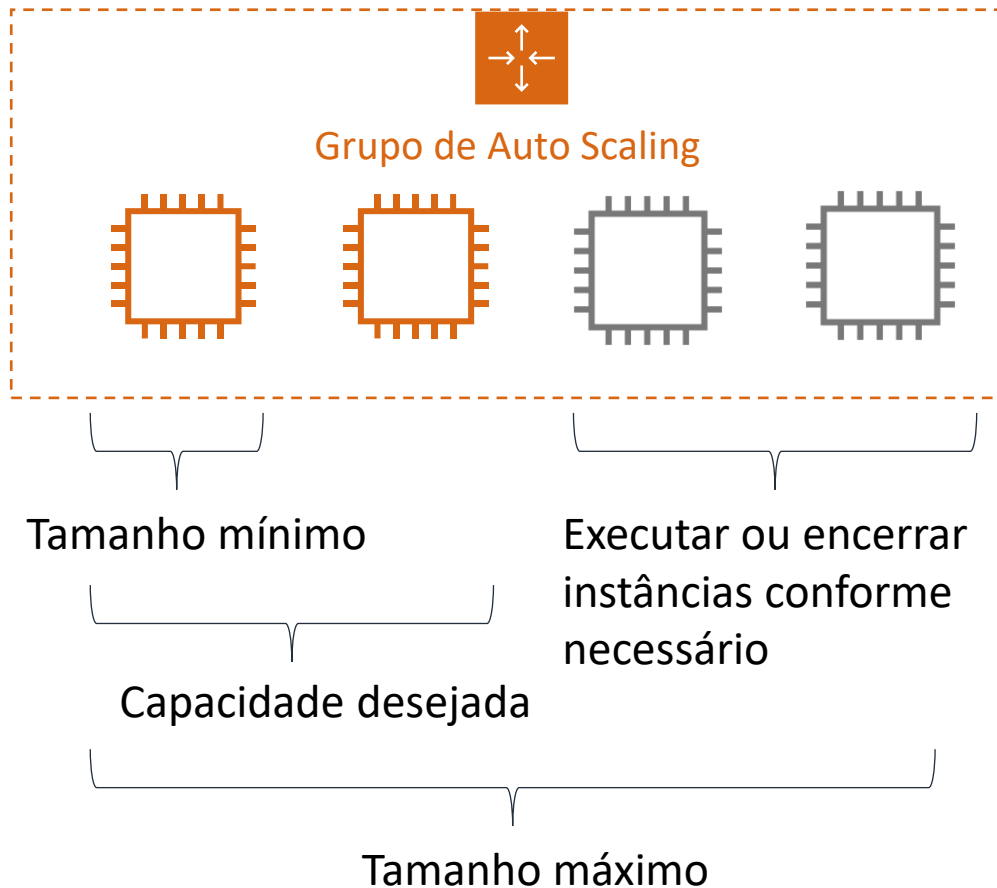
O desafio é adivinhar eficientemente a quantidade desconhecida de quanta capacidade de computação você precisa.

76%

24%

Novembro

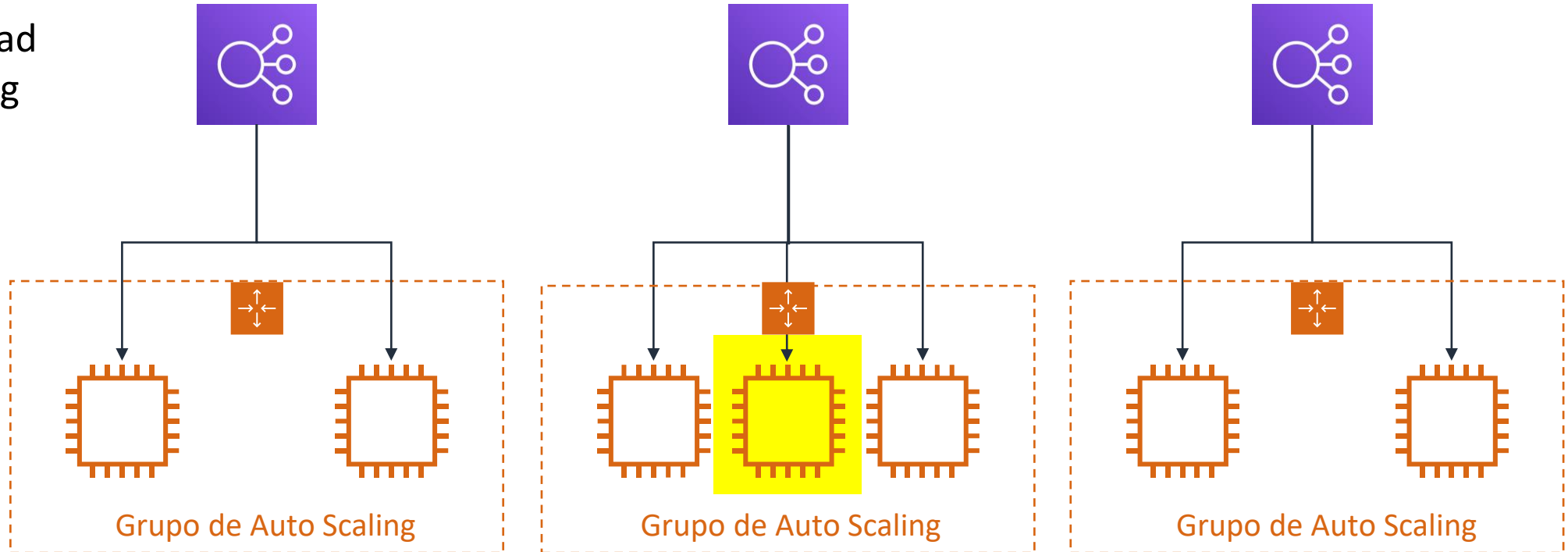
Grupos de Auto Scaling



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

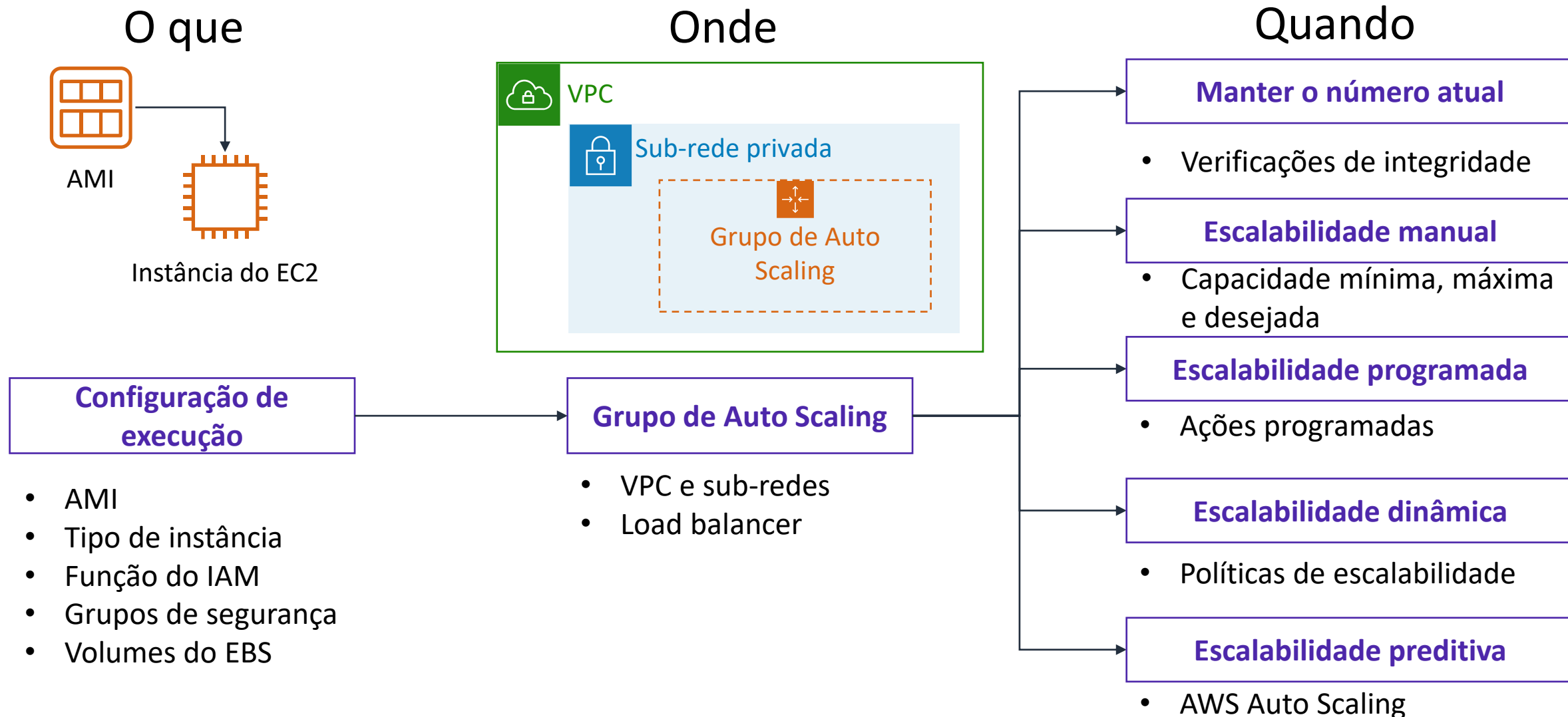


Configuração básica

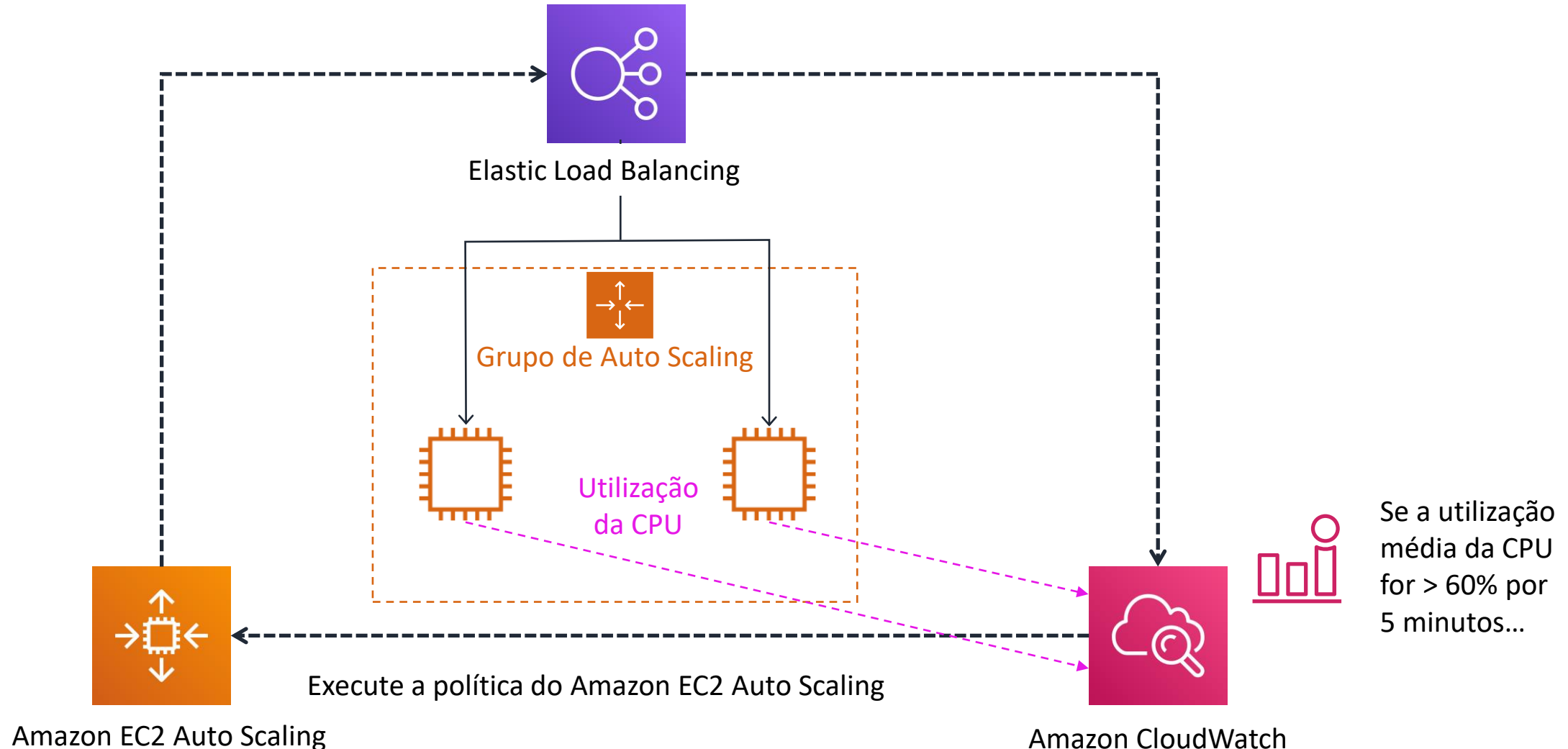
Aumento
(executar instâncias)

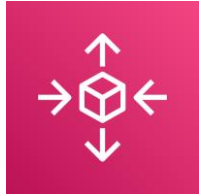
Redução
(encerrar instâncias)

Como funciona o Amazon EC2 Auto Scaling



Implementação da escalabilidade dinâmica





AWS Auto Scaling

- 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

Principais lições da Seção 3

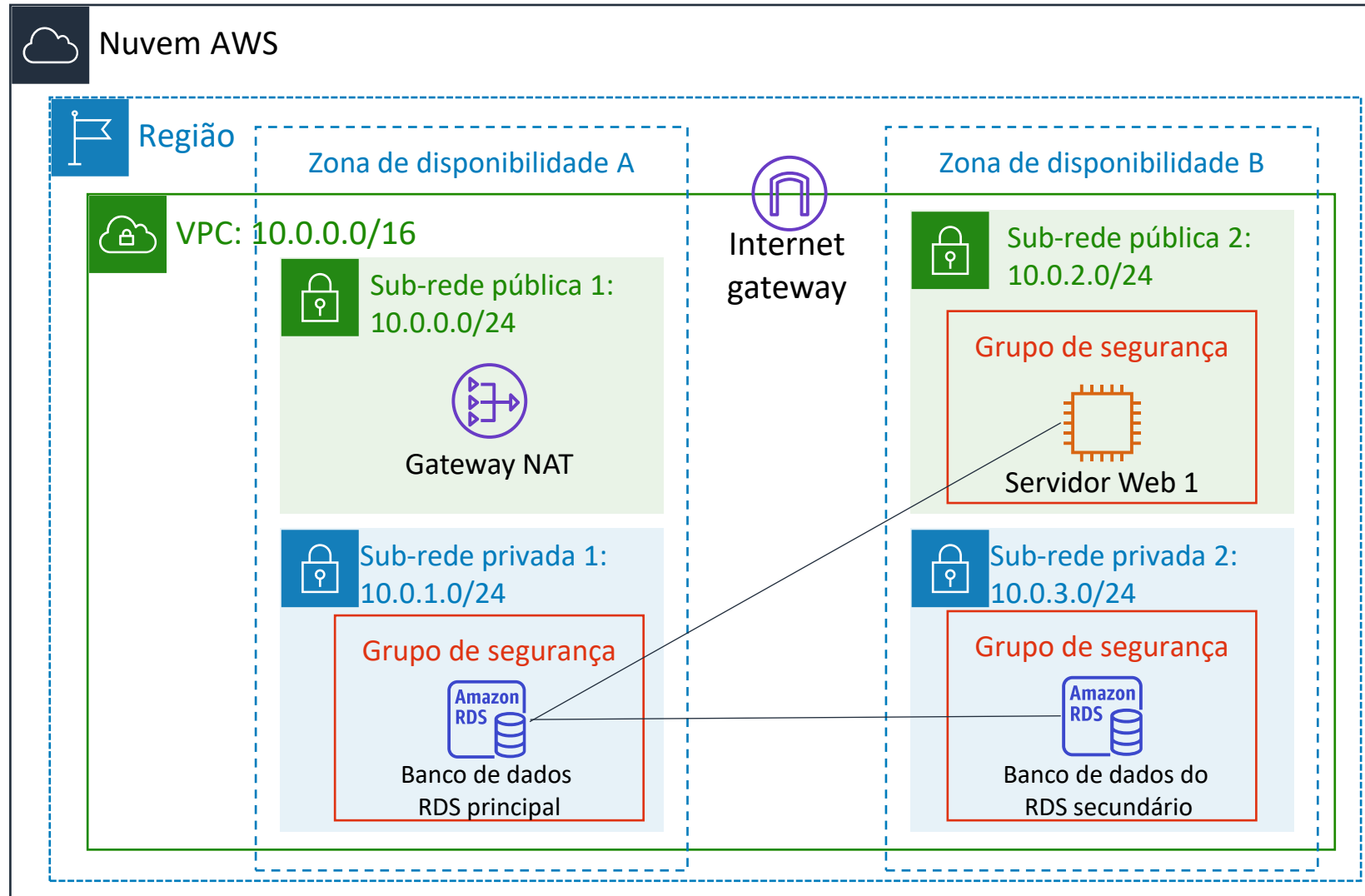


- A escalabilidade permite que você responda rapidamente a alterações nas necessidades de recursos.
- O Amazon EC2 Auto Scaling mantém a disponibilidade do aplicativo adicionando ou removendo automaticamente instâncias do EC2.
- Um grupo de Auto Scaling é um conjunto de instâncias do EC2.
- Uma configuração de execução é um modelo de configuração de instância.
- A escalabilidade dinâmica usa o Amazon EC2 Auto Scaling, o CloudWatch e o Elastic Load Balancing.
- O AWS Auto Scaling é um serviço separado do Amazon EC2 Auto Scaling.

Laboratório 6: Escale e faça o balanceamento de carga da arquitetura

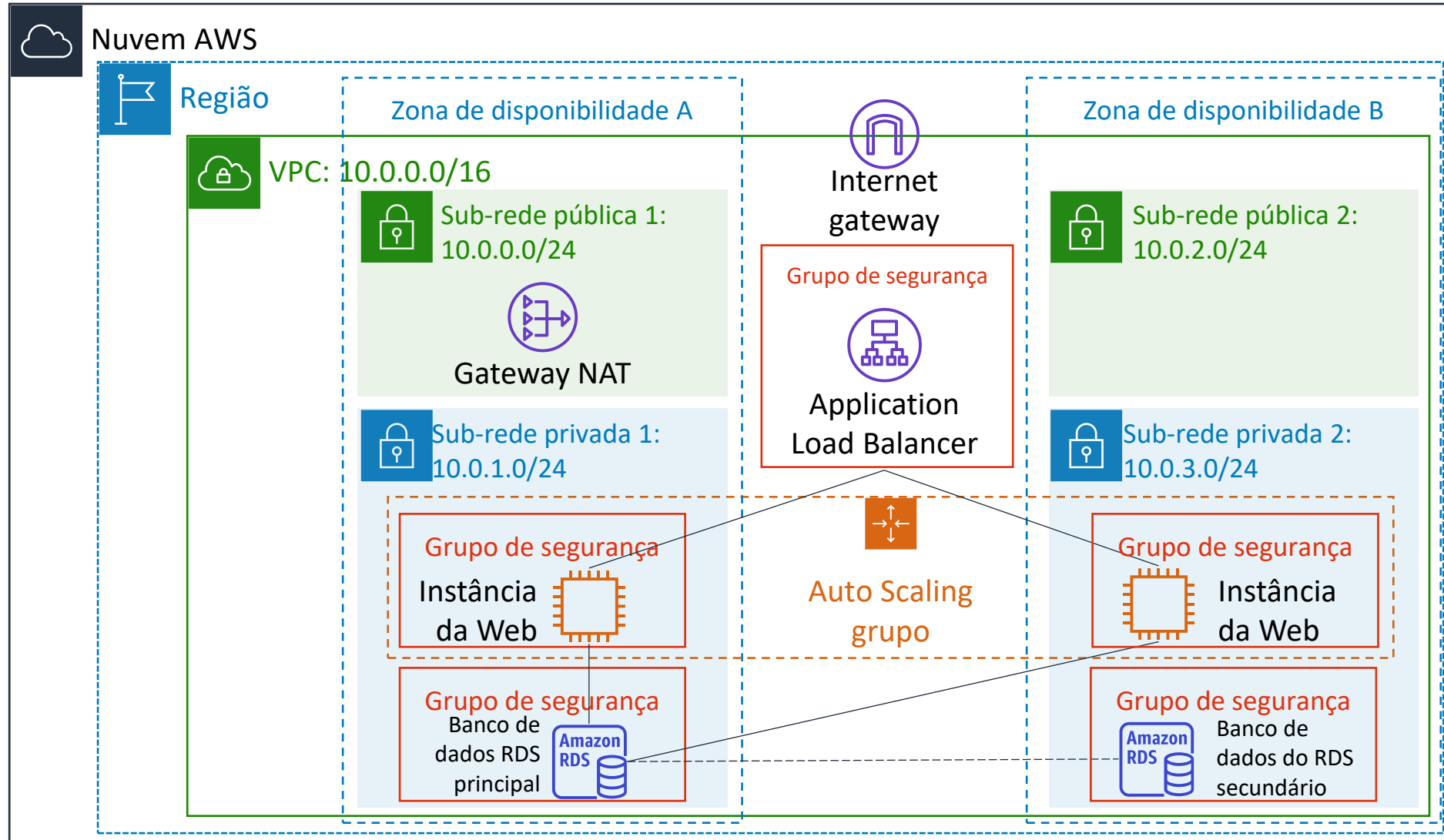


Laboratório 6: Cenário



- Criar uma nova Imagem de máquina da Amazon (AMI) por meio de instância em execução.
- Criar um Application Load Balancer.
- Criar uma configuração de execução e um grupo de Auto Scaling.
- Escalar automaticamente novas instâncias em uma sub-rede privada.
- Criar alarmes do Amazon CloudWatch e monitorar a performance da sua infraestrutura.

Laboratório 6: Produto final





Aproximadamente 30 minutos



Inicie o Laboratório 6: escale e balanceie sua arquitetura

Resumo do laboratório: Principais lições



Módulo 10: Auto Scaling e monitoramento

Conclusão do módulo

Resumindo, neste módulo você aprendeu a:

- Indicar como distribuir o tráfego entre instâncias do Amazon Elastic Compute Cloud (Amazon EC2) usando o Elastic Load Balancing.
- Identificar como o Amazon CloudWatch permite monitorar os recursos e os aplicativos da AWS em tempo real.
- Explicar como o Amazon EC2 Auto Scaling executa e libera servidores em resposta a alterações na carga de trabalho.
- Execute tarefas de escalabilidade e balanceamento de carga para melhorar uma arquitetura.

Conclua o teste de conhecimento



Exemplo de pergunta do exame

Qual serviço você usaria para enviar alertas baseados em alarmes do Amazon CloudWatch?

- A. Amazon Simple Notification Service
- B. AWS CloudTrail
- C. AWS Trusted Advisor
- D. Amazon Route 53

Obrigado

© 2019 Amazon Web Services, Inc. ou suas afiliadas. Todos os direitos reservados. Este trabalho não pode ser reproduzido ou redistribuído, no todo ou em parte, sem a permissão prévia por escrito da Amazon Web Services, Inc. É proibido copiar, emprestar ou vender para fins comerciais. Para correções ou comentários sobre o curso, envie um e-mail para: aws-course-feedback@amazon.com. Para todas as outras perguntas, entre em contato conosco em: <https://aws.amazon.com/contact-us/aws-training/>. Todas as marcas comerciais pertencem a seus proprietários.

