Serviços de Computação em Nuvem: Escalabilidade com Auto-Scaling

Resumo

Referências Bibliográficas

- Criação de Instância EC2 Amazon Linux https://www.awsacademy.com/LMS_Login;
- Endereços IP elásticos - Amazon Elastic Compute Cloud
- Amazon EBS - Armazenamento em blocos de alta performance – Amazon EBS; Amazon Elastic Block Store (Amazon EBS) - Amazon Elastic Compute Cloud
- Snapshots - Snapshots do Amazon EBS - Amazon Elastic Compute Cloud

Auto Scaling - Criando instância Servidor Web APP para o Auto Scalling

Benefícios do Multi-AZ

Multi-AZ é importante para implementações de serviços na AWS pois com múltiplas Zonas de Disponibilidade (AZ) temos uma série de benefícios. Entre eles são:

- Resiliência;
- Failover automático;
- Maior disponibilidade;
- Durabilidade aperfeiçoada;
- Proteção do desempenho do seu banco de dados;

O que é Resiliência e Failover automático?

Resiliência é a capacidade de uma carga de trabalho se recuperar de interrupções de infraestrutura ou serviço, adquirir dinamicamente recursos computacionais para atender à demanda e mitigar interrupções, como configurações incorretas ou problemas temporários de rede.

Se um volume de armazenamento na instância primária falhar em uma implantação Multi-AZ do Amazon RDS, ele iniciará automaticamente um failover (instância que estava em standby) para a espera atualizada (ou para uma réplica, no caso do Amazon Aurora).

O que é maior disponibilidade?

Você se beneficia da disponibilidade otimizada do servidor ao executar implantações Multi-AZ. Se ocorrer uma falha em uma Zona de disponibilidade ou em uma instância de DB, por exemplo, o impacto na disponibilidade será limitado ao tempo necessário para a conclusão do failover automático, que geralmente é inferior a um minuto para o Amazon Aurora (chegando a 30 segundos com o uso do MariaDB Connector/J) e de um a dois minutos para outros mecanismos de banco de dados (consulte as perguntas frequentes do RDS para obter detalhes).

O que é Proteção do desempenho do seu banco de dados?

Diferentemente das implantações Single-AZ, a atividade de E/S não é suspensa em sua instância principal durante o backup nas implantações Multi-AZ de mecanismos MySQL, MariaDB, Oracle e PostgreSQL, pois o backup é feito por meio da instância de espera.

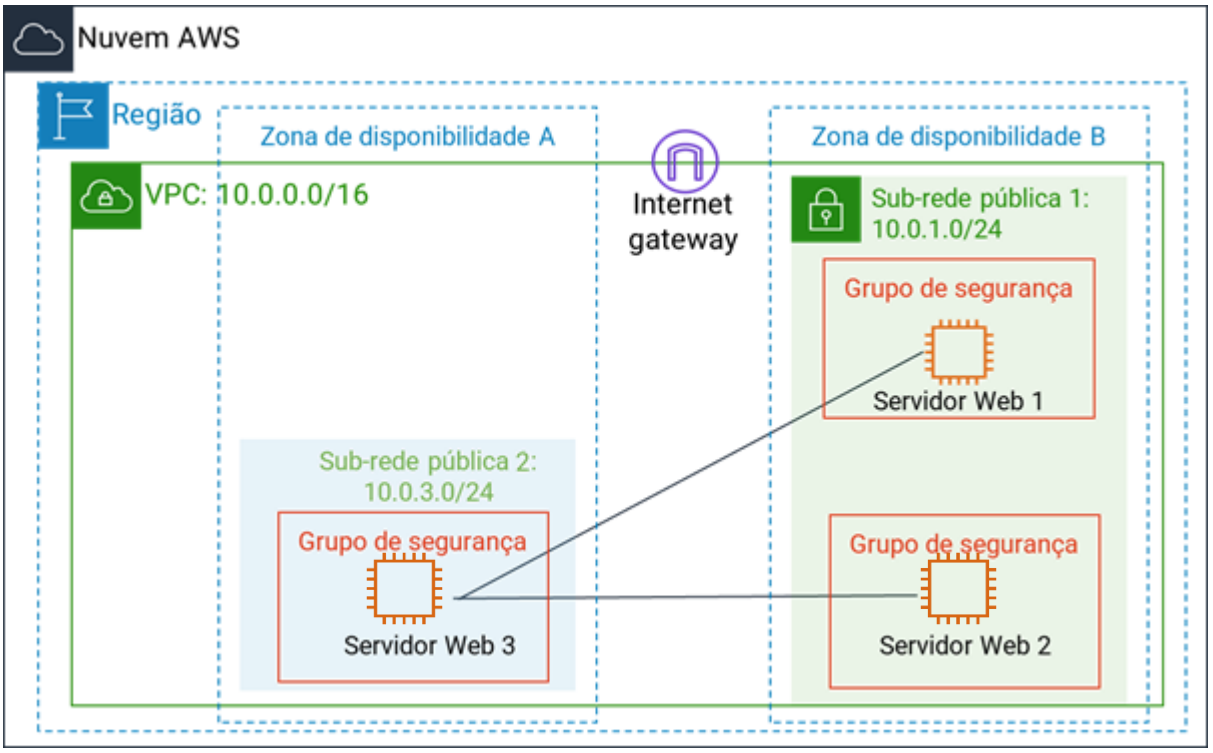
Mais sobre implantações Multi-AZ

Implantações Multi-AZ podem ter uma ou duas instâncias de banco de dados, por exemplo, em espera.

Uma implantação de instância de banco de dados Multi-AZ tem uma instância de banco de dados em espera (standby) que fornece suporte para failover, mas não serve tráfego de leitura.

Quando a implantação tem duas instâncias de banco de dados em espera, ela é chamada de implantação de cluster de banco de dados Multi-AZ. Uma implantação de cluster de banco de dados Multi-AZ tem instâncias de banco de dados em espera que fornecem suporte para failover e podem servir tráfego de leitura.

Exemplo de topologia MultiAZ:



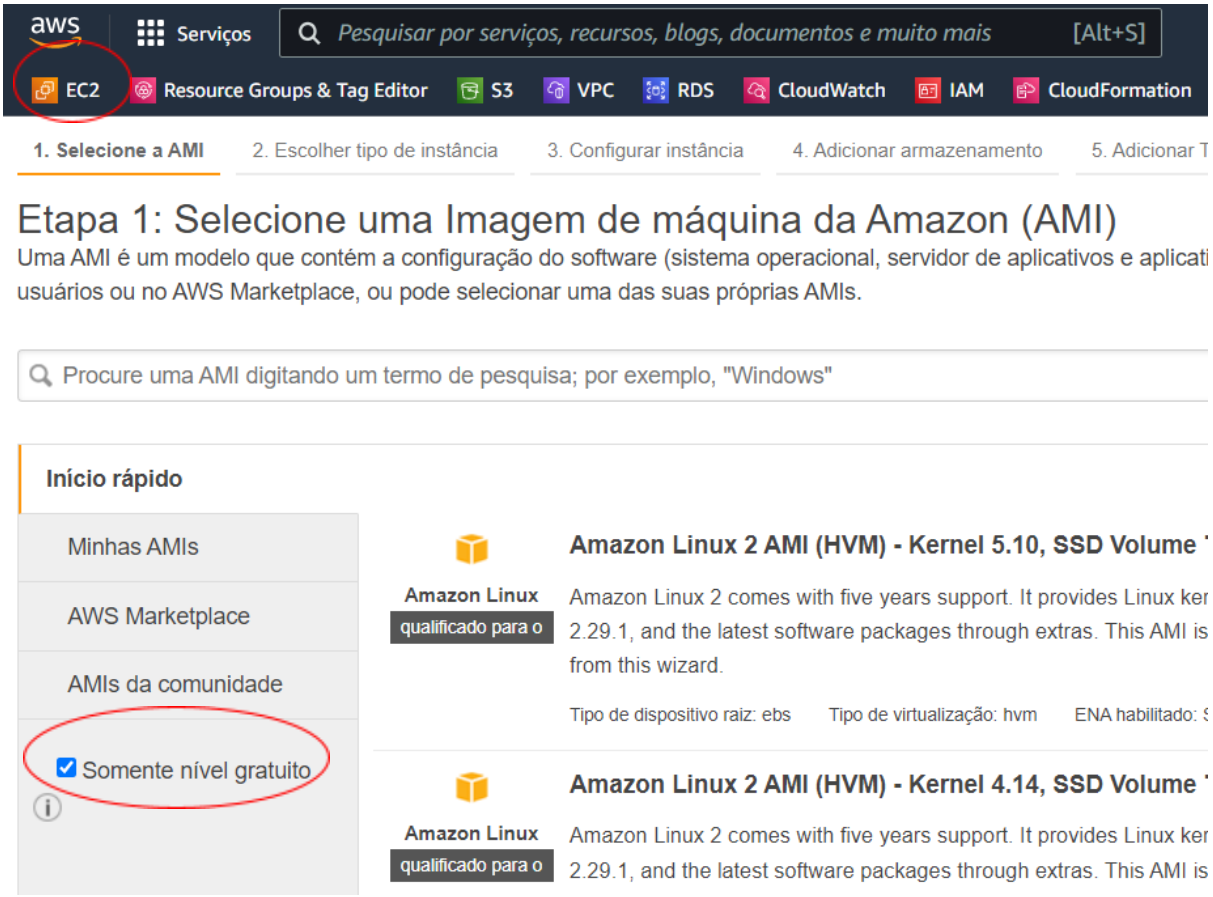
Aqui temos um exemplo de uma topologia com multi-AZ com uma alta disponibilidade e resiliência.

Nesta topologia temos 3 instâncias web, estando uma em uma Zona de Disponibilidade A e duas em uma Zona de Disponibilidade B. Caso uma das zonas caia, perca conectividade ou tenha algum problema parecido.

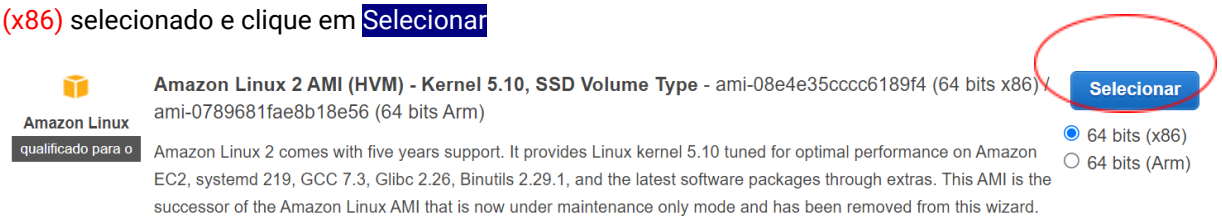
Laboratório 10: Criar Instância com WebServer APP para o Auto Scalling

---

1. No Console de Gerenciamento da AWS no menu **Serviços**, clique em **EC2**.
2. No **Painel EC2,** clique em Executar instância.
3. **Etapa 1 -** você irá se deparar com diversas AMIs encontradas na página de AMIs
4. Caso queira ver apenas as AMIs gratuitas, clique em **Somente nível gratuito**



Seleciona a AMI Amazon Linux 2 AMI (HVM) que é qualificada para o nível gratuito, mantenha 64 bits (x86) selecionado e clique em Selecionar



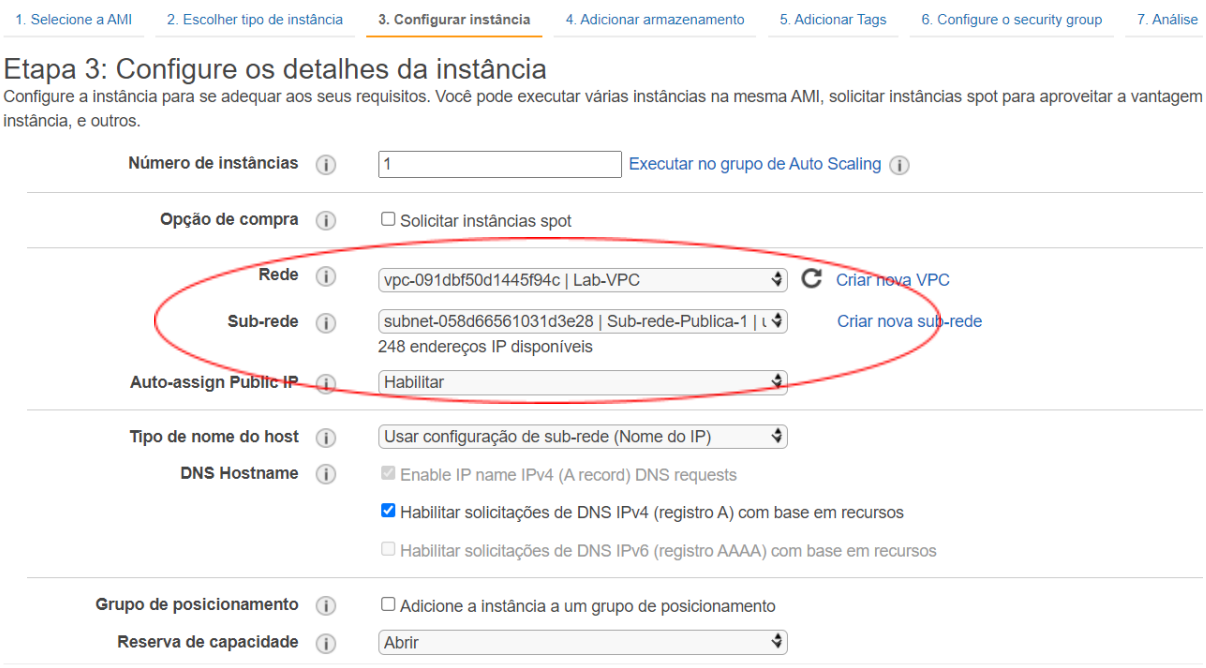
Etapa 2 – escolha o tipo da instância, **t2.micro** é neste momento que estamos escolhendo um processador e memória mais potentes conforme solicitação da empresa.

Clique em "Próximo: Configure os detalhes da instância";

Segue relação e detalhamento de cada família de máquinas: https://aws.amazon.com/pt/ec2/instance-types/

Etapa 3 – nesta etapa podemos fazer configurações detalhadas de nossa instancia para este caso, usaremos **LAB-VPC** criada no **Laboratório 1**.

nossa instância **Amazon EC2** ficará na a **Sub-rede-Publica-1**, onde será atribuído automaticamente um endereço IP público à nossa máquina virtual incluindo DNS.

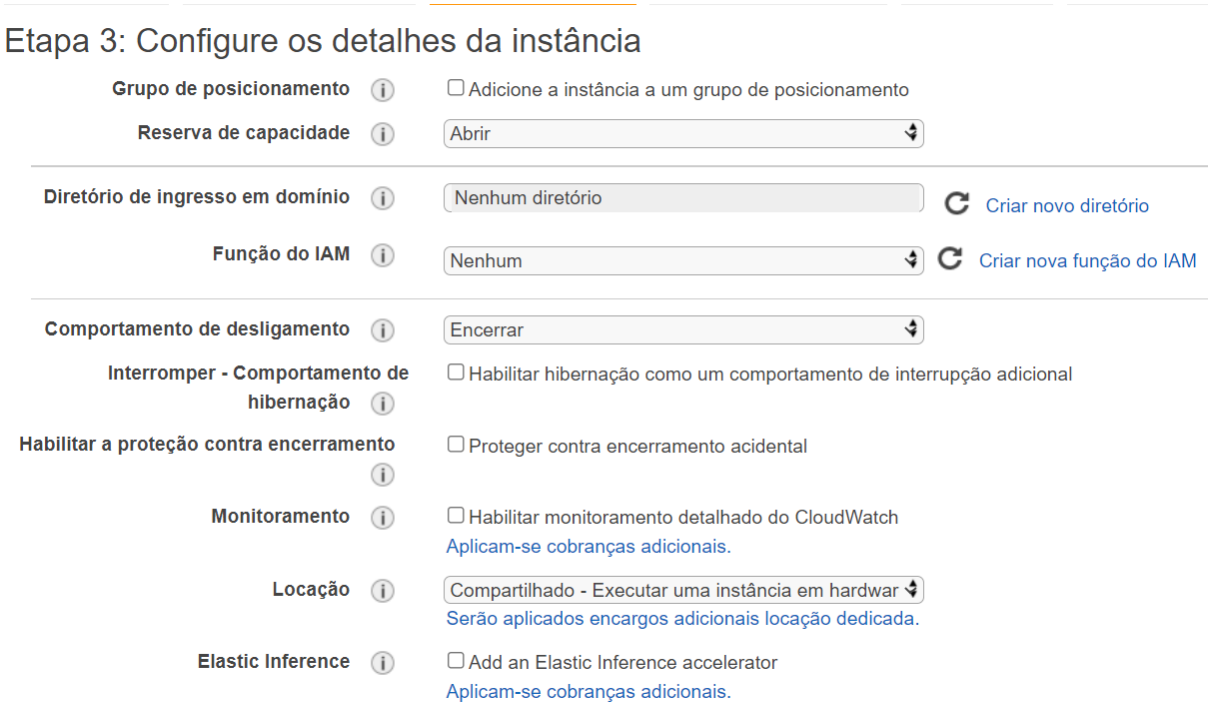


a) Estamos usando **Lab-VPC** padrão da AWS "vpc-05c689....";

b) Alterar **Auto-assign Public IP** para: Habilitar

c) Deixar habilitado a opção de DNS - "Habilitar solicitações de DNS IPv4 (Registro A) com base em recursos".

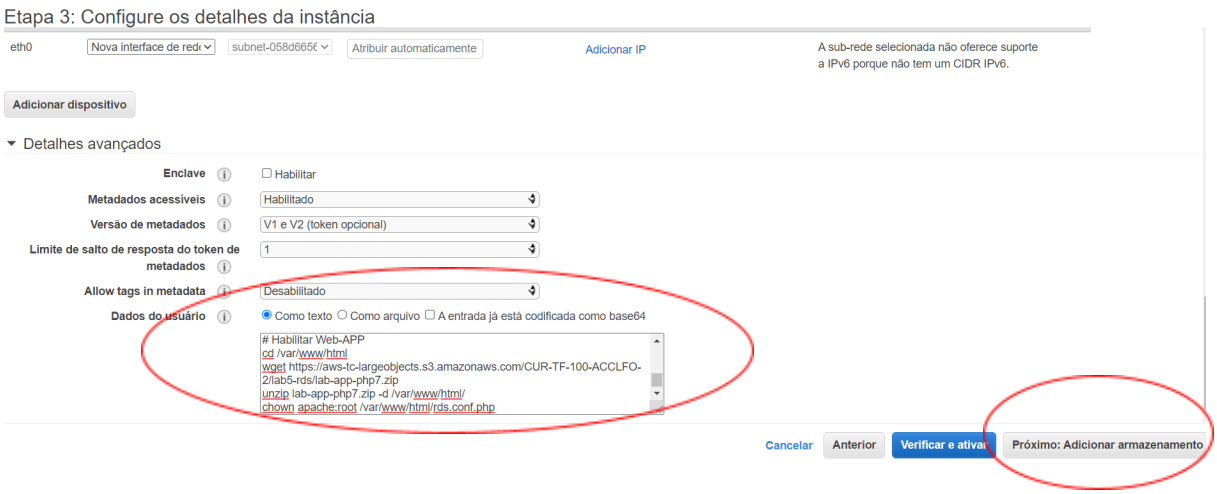
Role para baixo para continuar nesta etapa 3, não vamos alterar nada neste campos.



Role para baixo para ver a seção **Detalhes avançados**.

a) Expanda **detalhes avançados**. Aparecerá no último campo **Dados do Usuário**.

b) Copie os comandos a seguir e cole-os no campo **Dados do Usuário**. Mantendo a opção "Como texto" selecionada.



Clique em "Próximo: Configure os detalhes da instância";

```
#!/bin/bash -ex
# Updated Amazon Linux 2
yum -y update
# Instalar Apache2 - PHP - Banco de Dados MySQL
yum -y install httpd php mysql php-mysql
# Instalar LAMP para MariaDB 10.2 e PHP 7.2
amazon-linux-extras install -y lamp-mariadb10.2-php7.2 php7.2
# Instalar Apache2 - MariaDB
yum install -y httpd mariadb-server
# Habilitar Apache2
/usr/bin/systemctl enable httpd
# Iniciar Apache2
/usr/bin/systemctl start httpd
# Habilitar Web-APP
cd /var/www/html
wget
https://aws-tc-largeobjects.s3.amazonaws.com/CUR-TF-100-ACCLFO-2/lab5-rds/lab-app
-php7.zip
unzip lab-app-php7.zip -d /var/www/html/
chown apache:root /var/www/html/rds.conf.php
```

Bootstrapping, código fornecido que é executado quando um computador é inicializado.
Neste caso para o servidor se tornar um servidor Web com APP para Sobrecarregar a máquina virtual e fazer com que o Auto Scaling comece a funcionar.

Etapa 4 – nesta etapa de armazenamento vamos manter o padrão de 8GB.

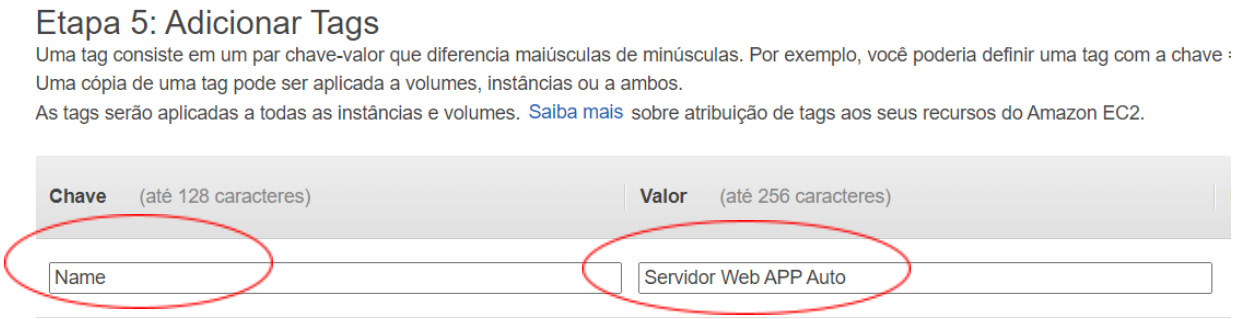
Esta opção "Excluir no encerramento", precisa estar avida para quando excluirmos a instâncias o volume não fique consumindo nossos créditos. Se tiver em produção deixe desmarcado para evitar exclusões desnecessárias.

Clique em "Próximo: Adicionar Tags";

Etapa 5 – nesta etapa de **Tags**, vamos colocar:

Chave: Name e Valor: Servidor Web APP Auto

Esse nome será usado nas Instâncias, Volumes e Interfaces de Rede.



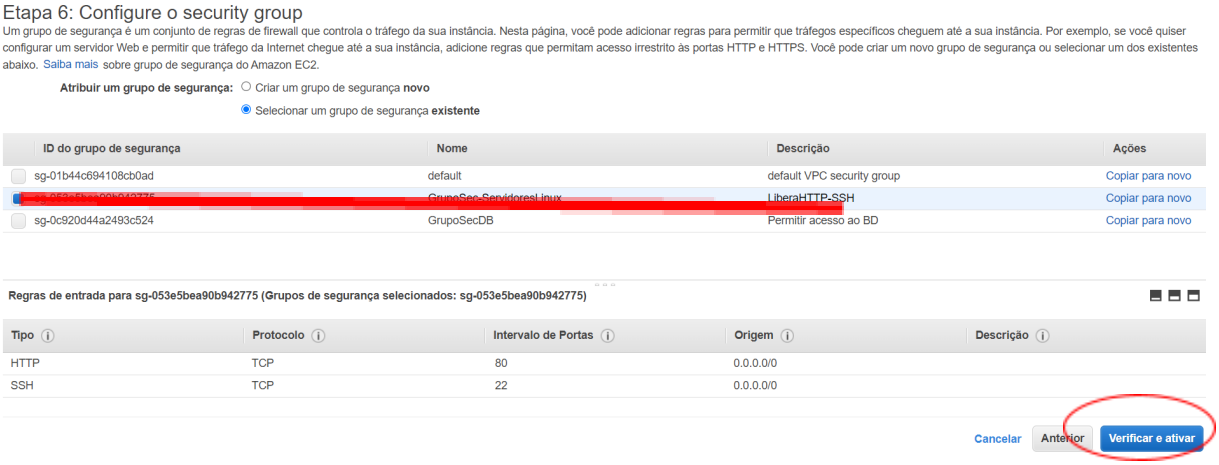
Clique em "Próximo: Configure o security group";

Etapa 6 – nesta etapa são os grupos de segurança, já existe um grupo criando anteriormente.

Que liberar a porta de acesso SSH porta 22 e acesso Web HTTP porta 80 para qualquer IP vindo da internet.

Marque "**Selecionar um grupo de segurança existente**"

a) Seleciona o grupo com nome: GrupoSec-ServidoresLinux



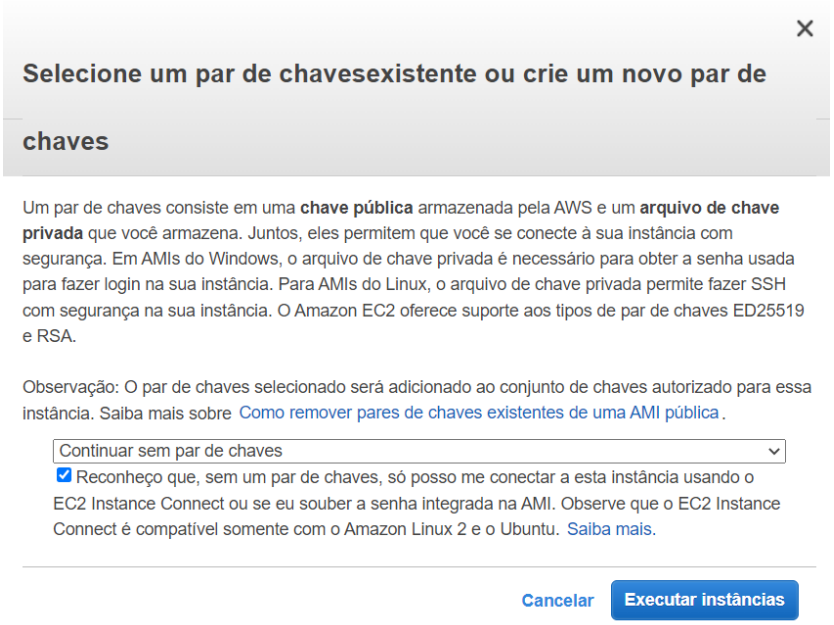
Veja que as portas SSH 22/TCP e HTTP 80/TCP já estão cadastradas, clique em Verificar e Ativar

Etapa 7 – Esta é a última etapa de revisão, temos dois alertas, mas podemos ignorá-los.

Clique em Executar

Essa janela será aberta, é o momento para gerar o par de chaves:

Não vamos usar ele agora



No menu suspenso, selecione **Continuar sem par de chaves**, marque a caixa a com a instrução "**Reconheço que, sem um par de chaves, só posso me conectar usando o EC2 Instance Connect…** "e clique em Executar Instância.

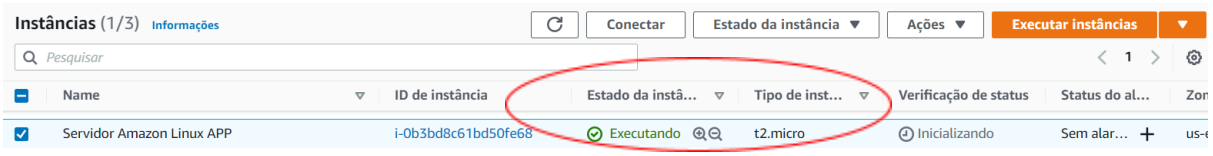
Você será direcionado para essa tela:

Clique em Exibir Instâncias

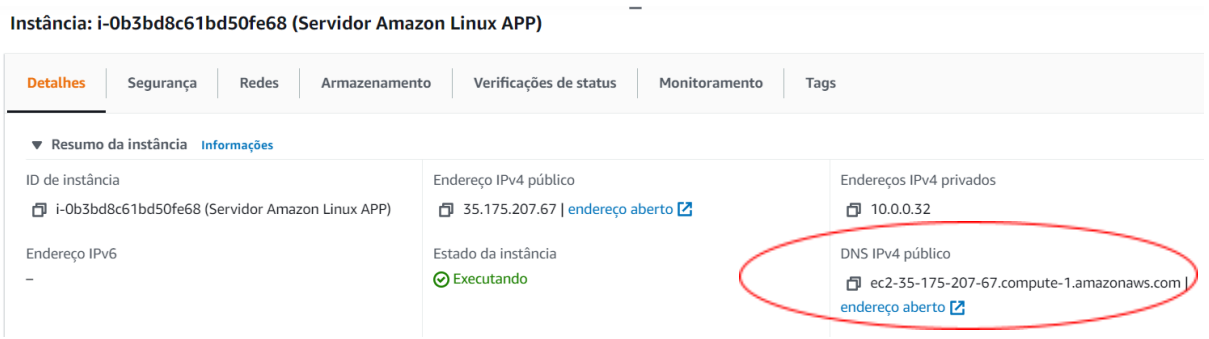
Etapa 8 – Acesso APP do Servidor Web.

Temos as colunas:

- Name “Servidor WEB Descomplica”;
- Estado da Instância (**Executando**);
- Tipo da Instância: ex.: **t2.micro**;
- Verificação de Status (Inicializando, 2/2 verificações aprovadas).



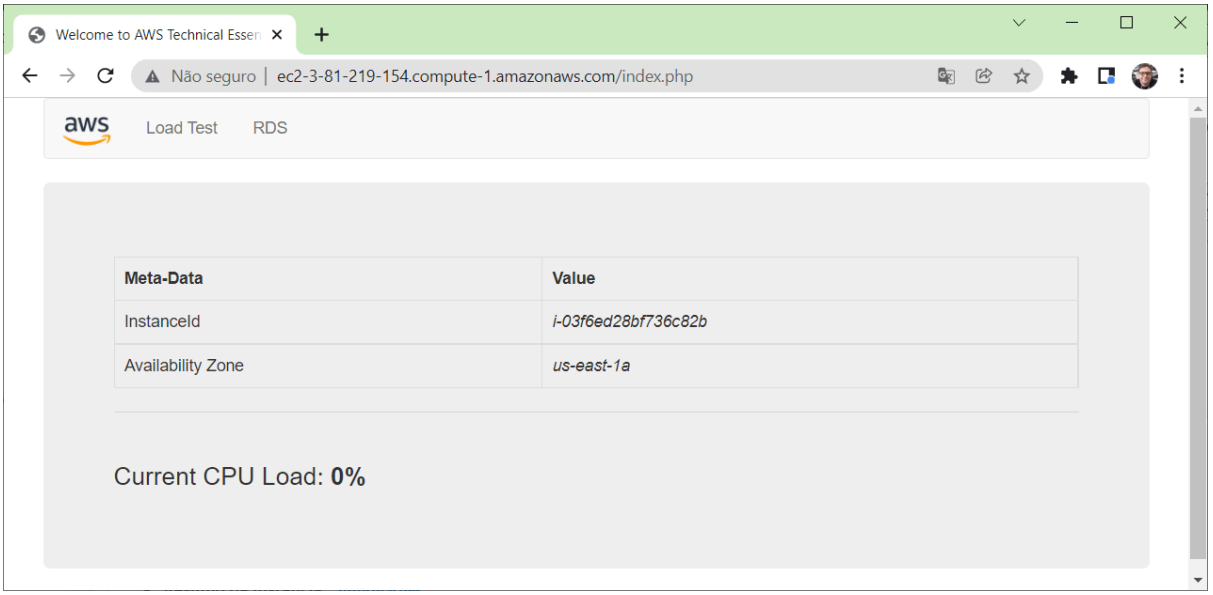
Assim que iniciar, faça o teste de acesso WEB.



Nesta mesma aba que já está marcada colete a informação: “Nome IPv4 público”: ex.: ec2-35-175-207-67.compute-1.amazonaws.com

Complete com: ec2-35-175-207-67.compute-1.amazonaws.com/index.php

Será exibida essa tela:



Laboratório 10 concluído, vamos para o próximo laboratório sequencial.

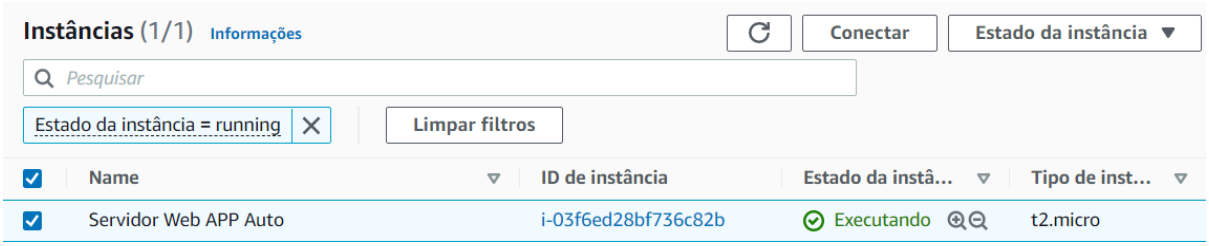
Laboratório 11 - Criando Instância modelo AMI para o Auto Scaling

Nesta tarefa, você criará uma AMI pelo **Servidor Web APP Auto** existente.

Isso salvará o conteúdo do disco de inicialização para que novas instâncias possam ser executadas com conteúdo idêntico.

5. No **Console de Gerenciamento da AWS**, no menu **Serviços**, clique em **EC2**.
6. No painel de navegação esquerdo, clique em **Instâncias**

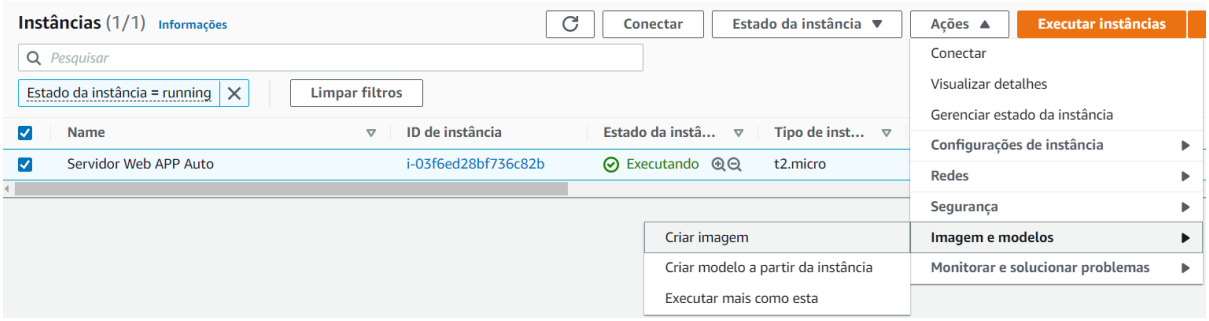
   1° - Confirme se a instância está em execução.
7. Aguarde até que as etapas de **Verificação de status** do **Servidor Web APP Auto**, estão 2/2 verificações aprovadas.



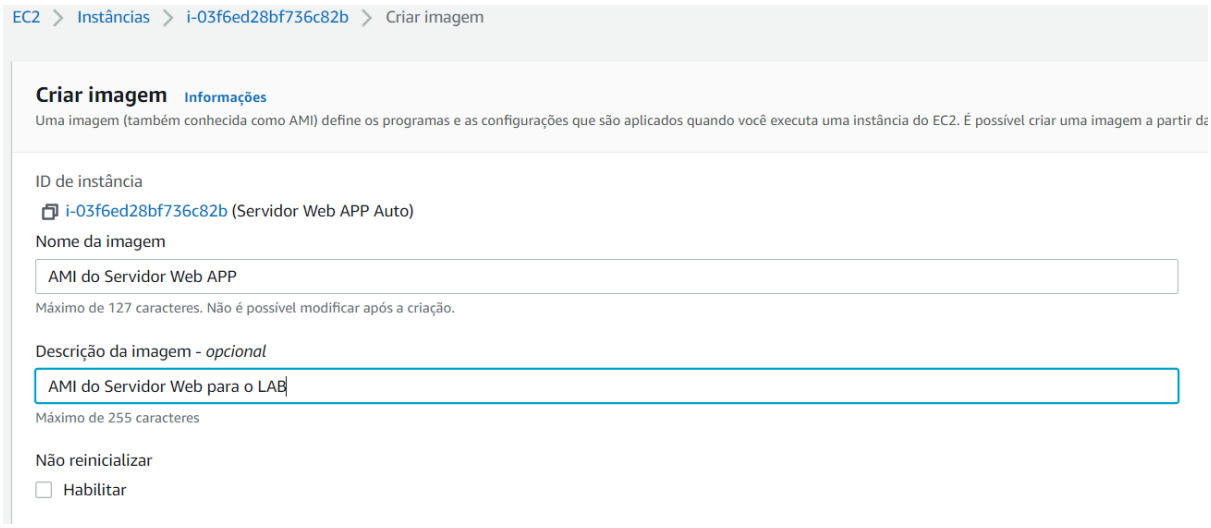
8. Clique em Refresh  :ar) até chegar ao estado desejado.

   Agora, você criará uma AMI com base nessa instância.

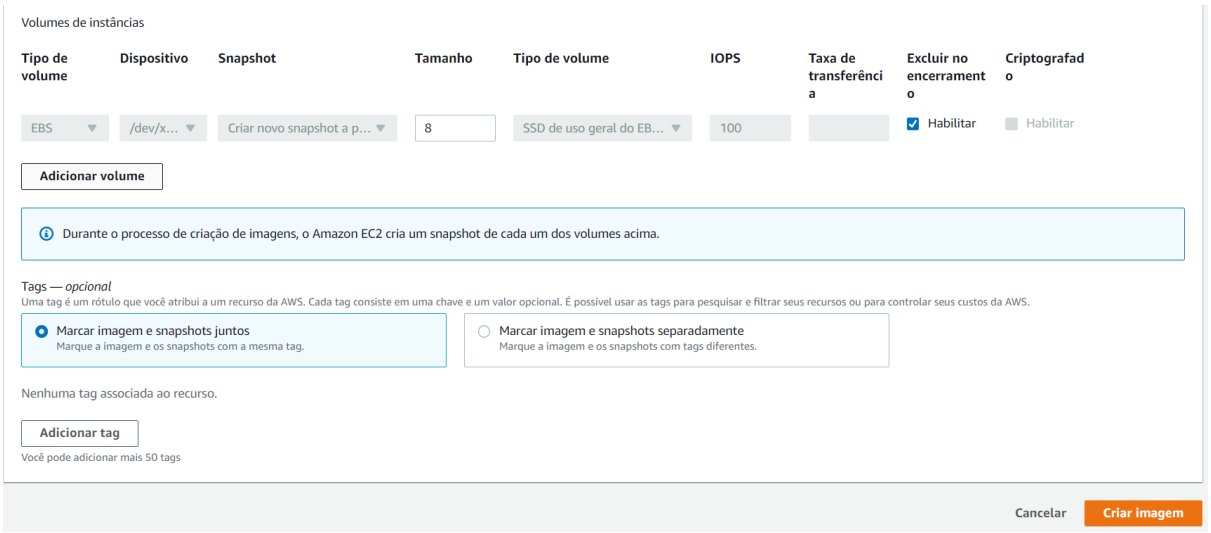
9. Selecione o **Servidor Web APP Auto**.
10. No menu **Ações**), clique em **Imagem e Modelos** > **Criar Imagem**



Configure:



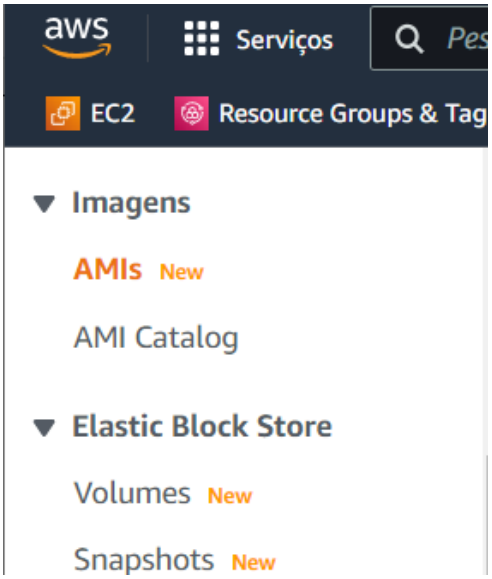
- **Nome da imagem**: AMI do Servidor Web APP
- Descrição da imagem: AMI do Servidor Web para o LAB
- Deixe desmarcado: Não reiniciar

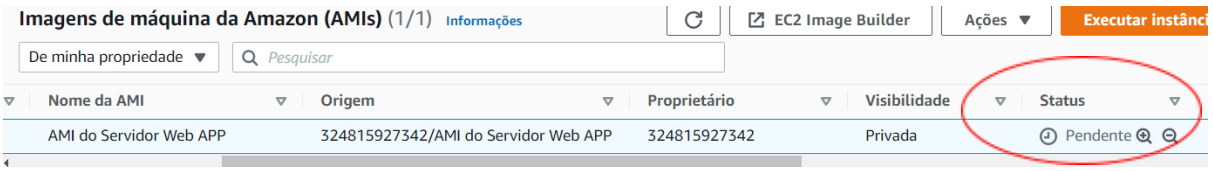


11. Clique em **Criar imagem**

No **Console de Gerenciamento da AWS**, no menu **Serviços**, clique em **EC2**.

No painel de navegação esquerdo, clique em **AMIs**

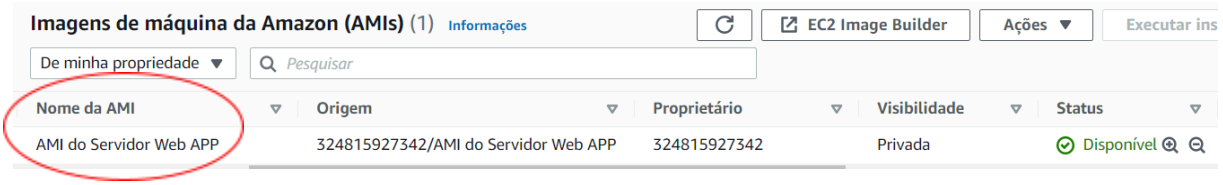




Aguarde o Status mudar de Pendente para Disponível

Demora até 5 minutos para criar.

Máquina modelo criada com sucesso:

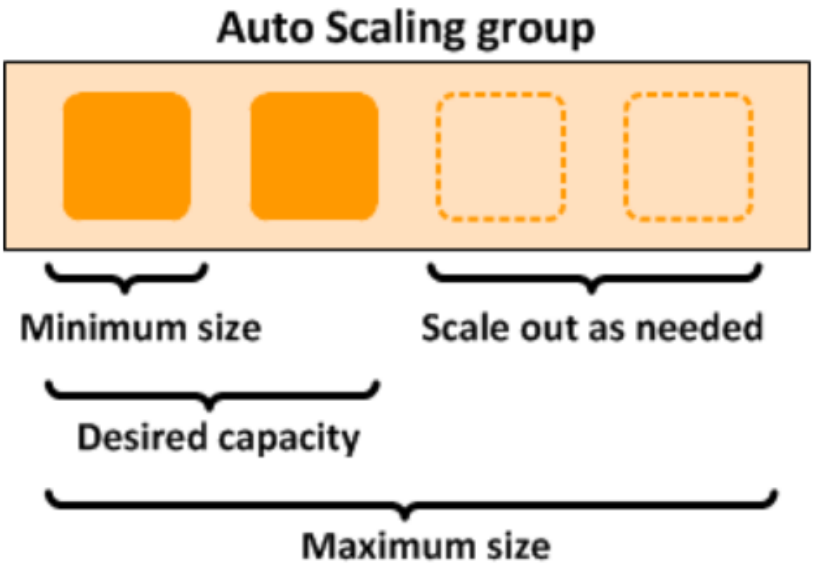


Você usará essa AMI ao iniciar o grupo de Auto Scaling posteriormente, no laboratório 12.

Auto Scaling Group

É um conjunto de instâncias do Amazon EC2 que são tratadas como um agrupamento lógico para fins de escalabilidade e gerenciamento automáticos. O tamanho de um grupo de Auto Scaling depende do número de instâncias definidas como a capacidade desejada.
Você pode ajustar seu tamanho para atender à demanda manualmente ou usando a escalabilidade automática.

A seguir, você verá uma imagem ilustrativa contendo as informações necessárias ao criar um Auto Scaling Group e poder entender um pouco melhor como ele funciona na prática logo em seguida.



Fonte: https://docs.aws.amazon.com/pt_br/autoscaling/ec2/userguide/what-is-amazon-ec2-auto-scaling.html

- Minimum size (tamanho mínimo): O menor número de instâncias ativas que seu Auto Scaling pode ter.

- Desired capacity (capacidade desejada): O número de instâncias recomendado.

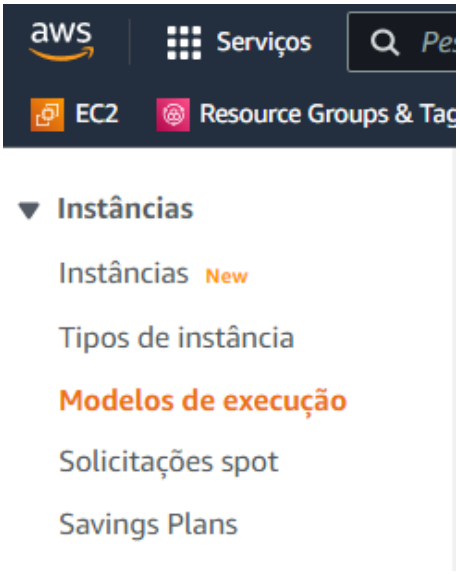
- Scale out as needed (dimensionar conforme necessário): Espaço disponível para expansão do número de instâncias.

- Maximum size (tamanho máximo): O número máximo de instâncias que podem ser executadas ao mesmo tempo.

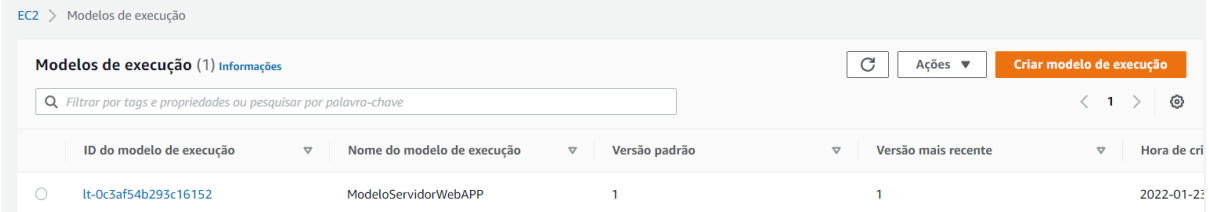
Laboratório 12 – Auto Scaling Group

Etapa 1 – Criar Modelo de Execução

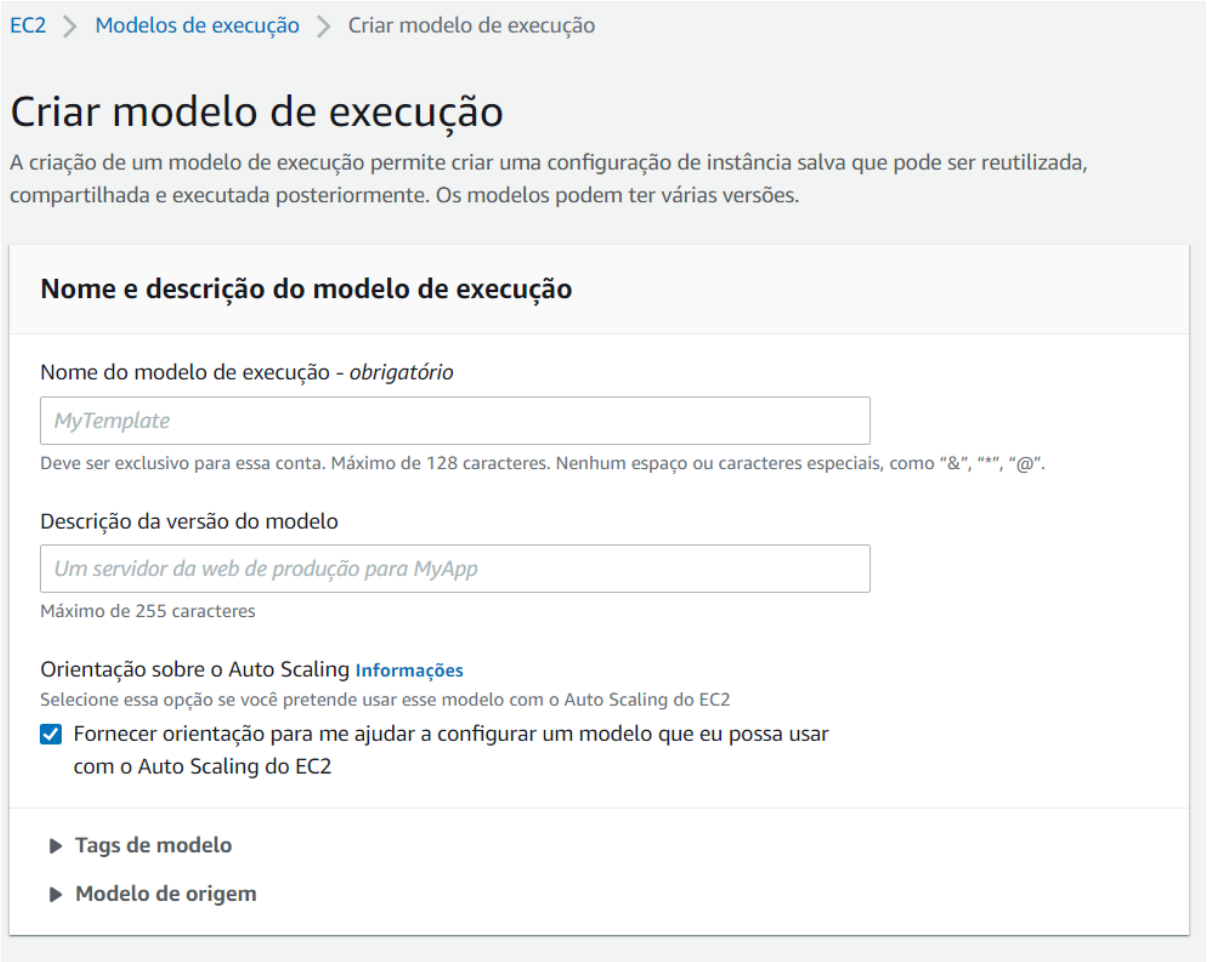
Entre no console AWS, EC2 no menu à esquerda clique em Modelos de execução



Dentro de **Modelos de execução**



Clique em **Criar modelo de execução**



Preencha o nome: **ModeloServidorWebAPP**

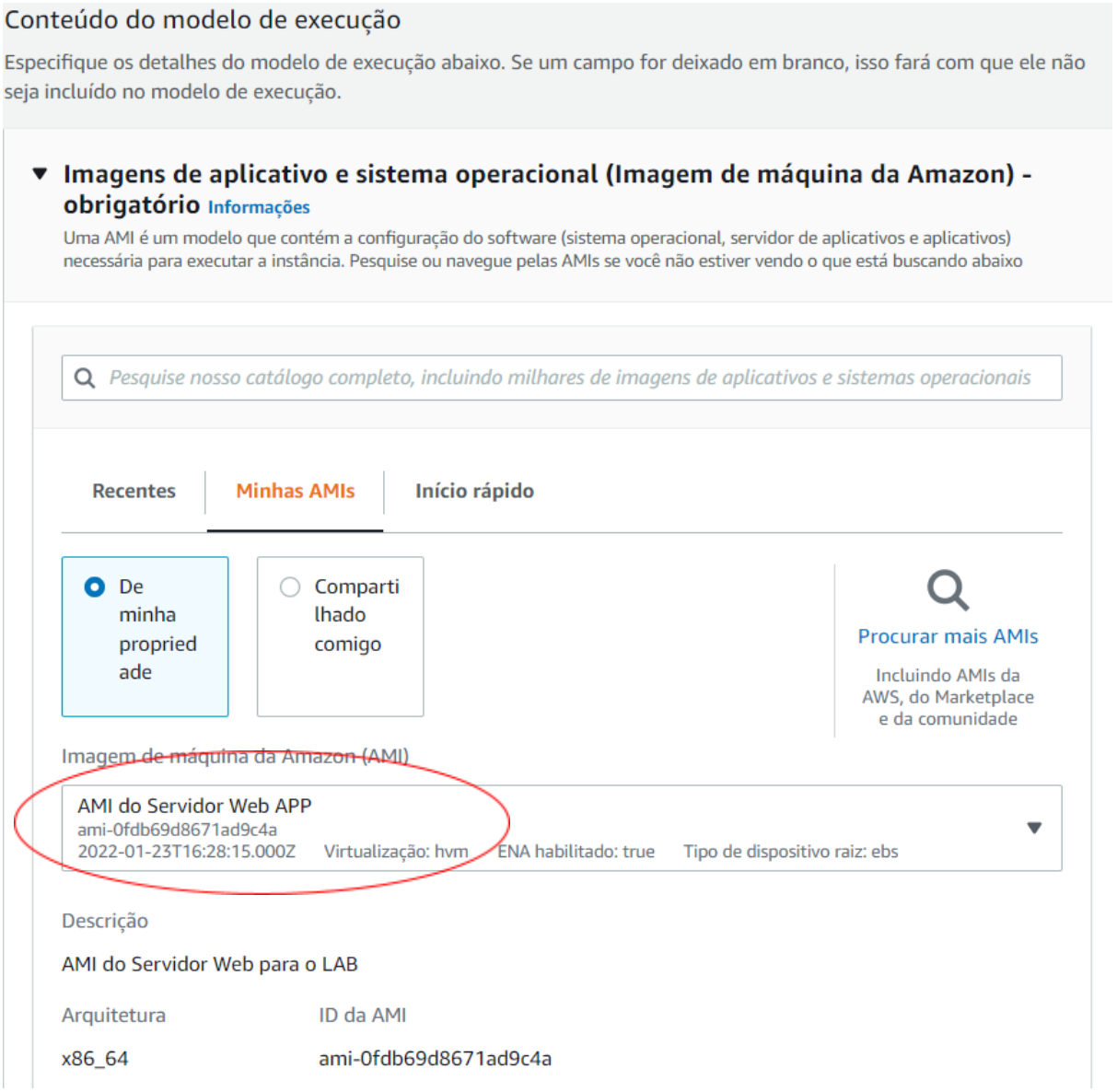
Descrição o nome: Modelo para ser usado no Grupo de Auto Scaling

Marque: "Fornecer orientação para me ajudar a configurar um modelo..."

Não vamos preencher **Tags de modelo**

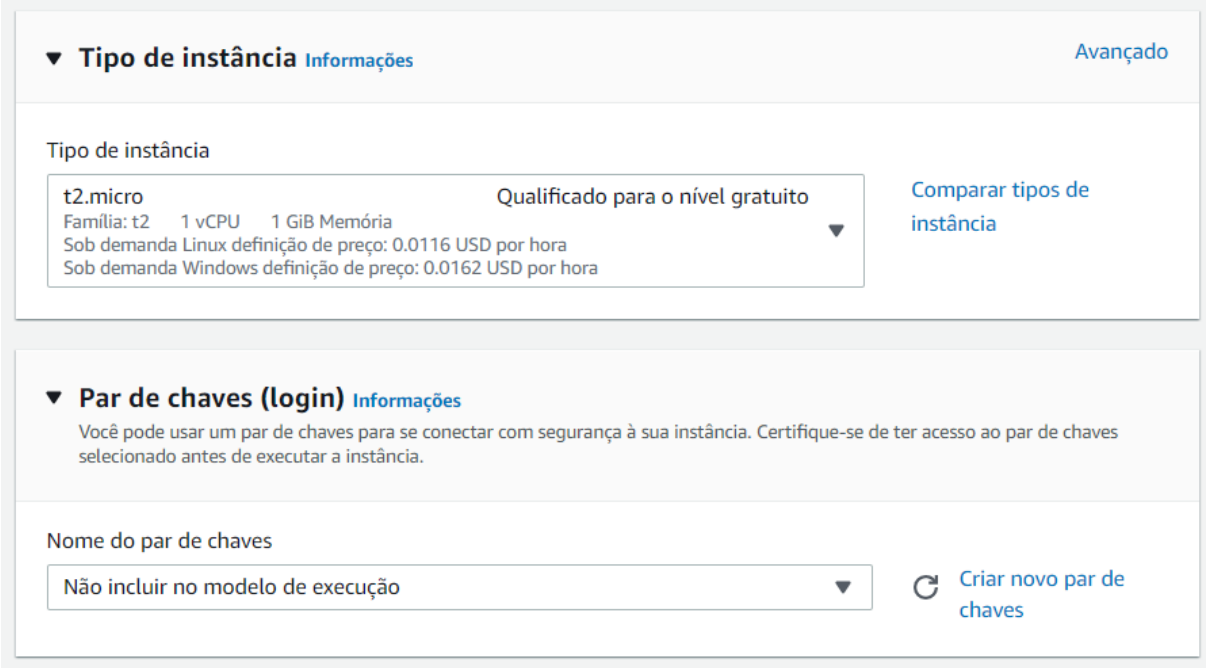
Não vamos mudar **Modelo de origem**

Dentro de conteúdo do modelo, marque "**Minhas AMIs**"



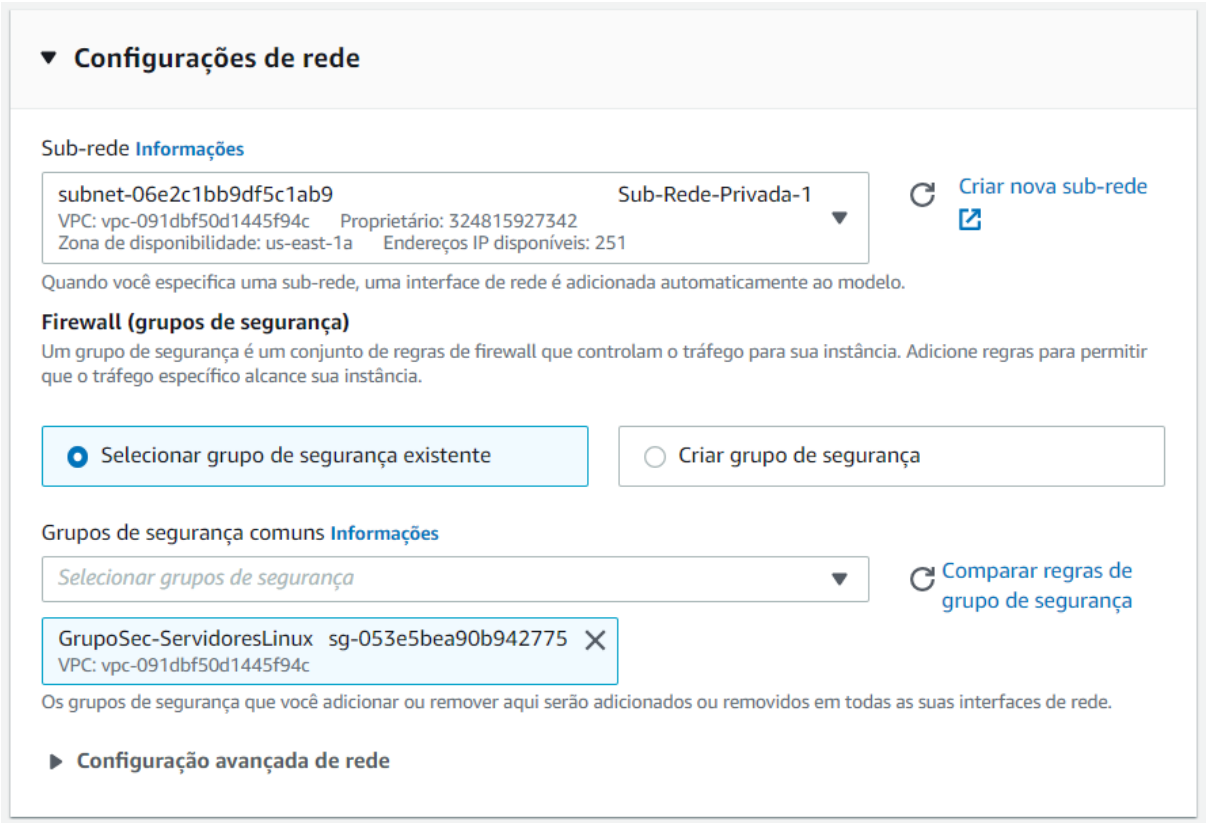
Selecione a **AMI do Servidor Web APP**

Dentro de "**Tipo de instância**", selecione: t2.micro



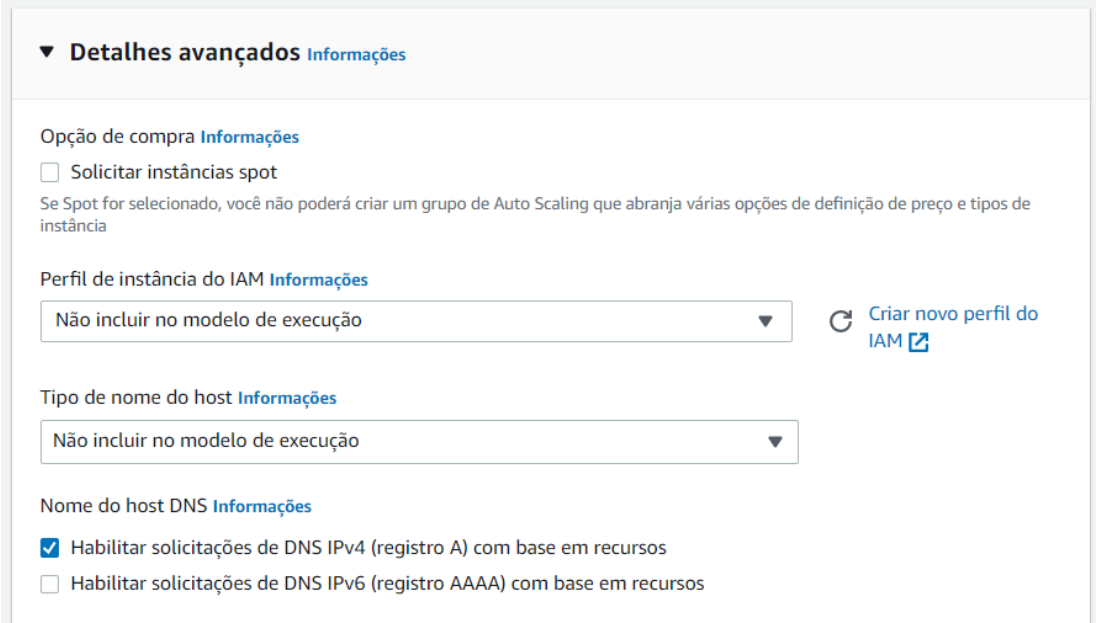
Dentro de "**Par de chaves**", não selecione nada, vamos aumentar nossa segurança.

Dentro de "**Configurações de rede**", selecione: Sub-Rede-Publica-1

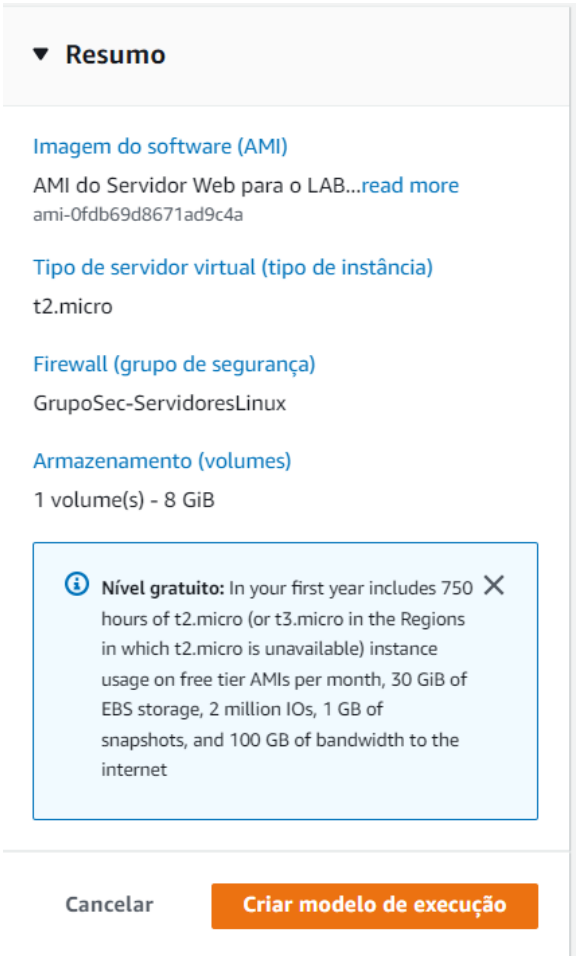


Dentro de "**Firewall**", "**Selecionar grupo de segurança existente**", selecione: GrupoSec-ServidoresLinux.

Dentro de “**Detalhes avançados**”, role até “**Nome do host DNS**”, selecione: Habilitar solicitações de DNS IPv4 (registro A).



No menu do lado direito “**Resumo**”

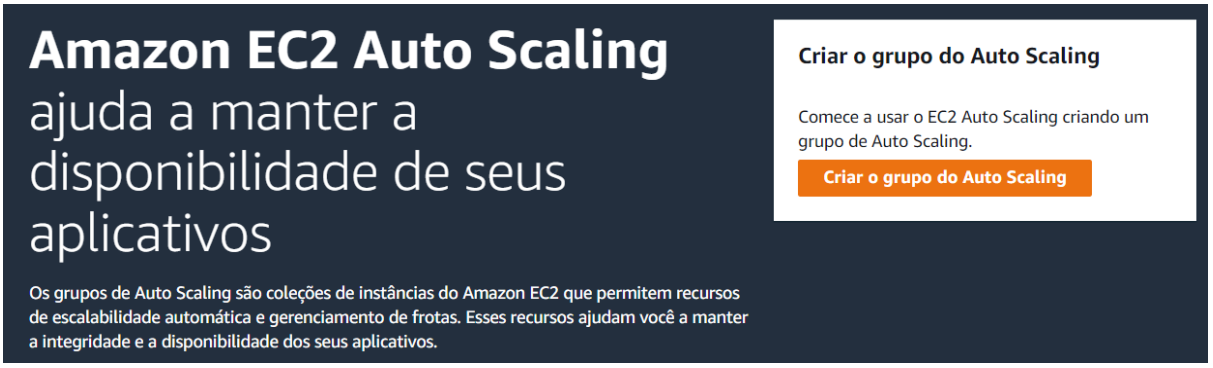


Clique em "**Criar um modelo de execução**

Etapa 2 – Criar Grupo do Auto Scaling

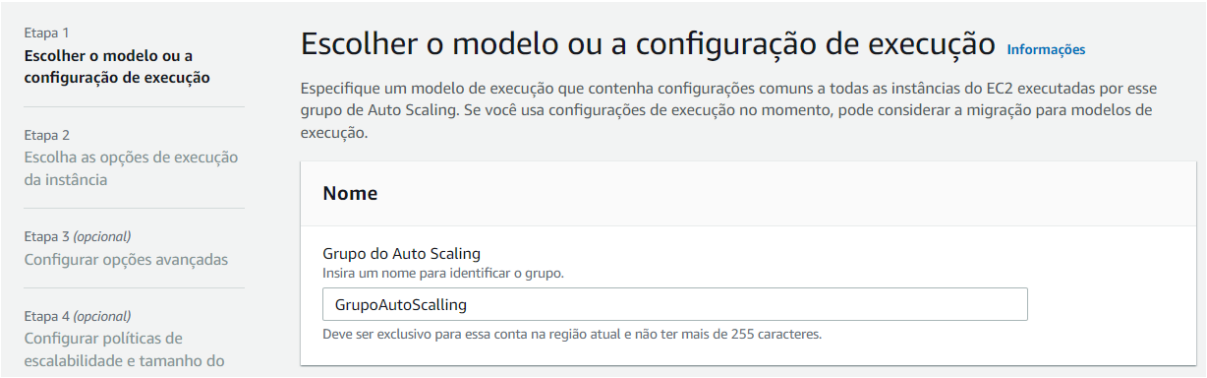
Entre no console AWS, EC2 no menu à esquerda clique em Grupos do Auto Scaling

Você será redirecionado para essa página:



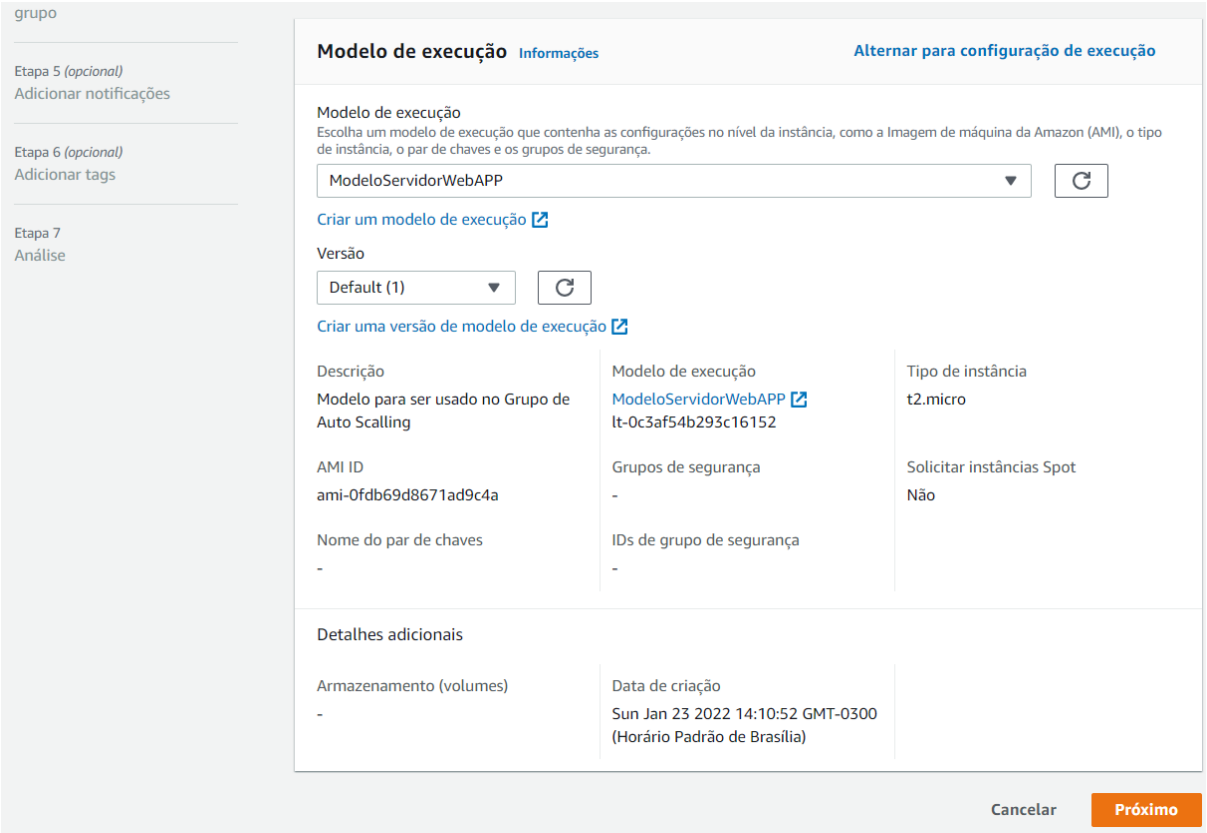
Clique em "Criar o grupo de Auto Scaling":

Etapa 1: Escolha o nome e o modelo de execução:



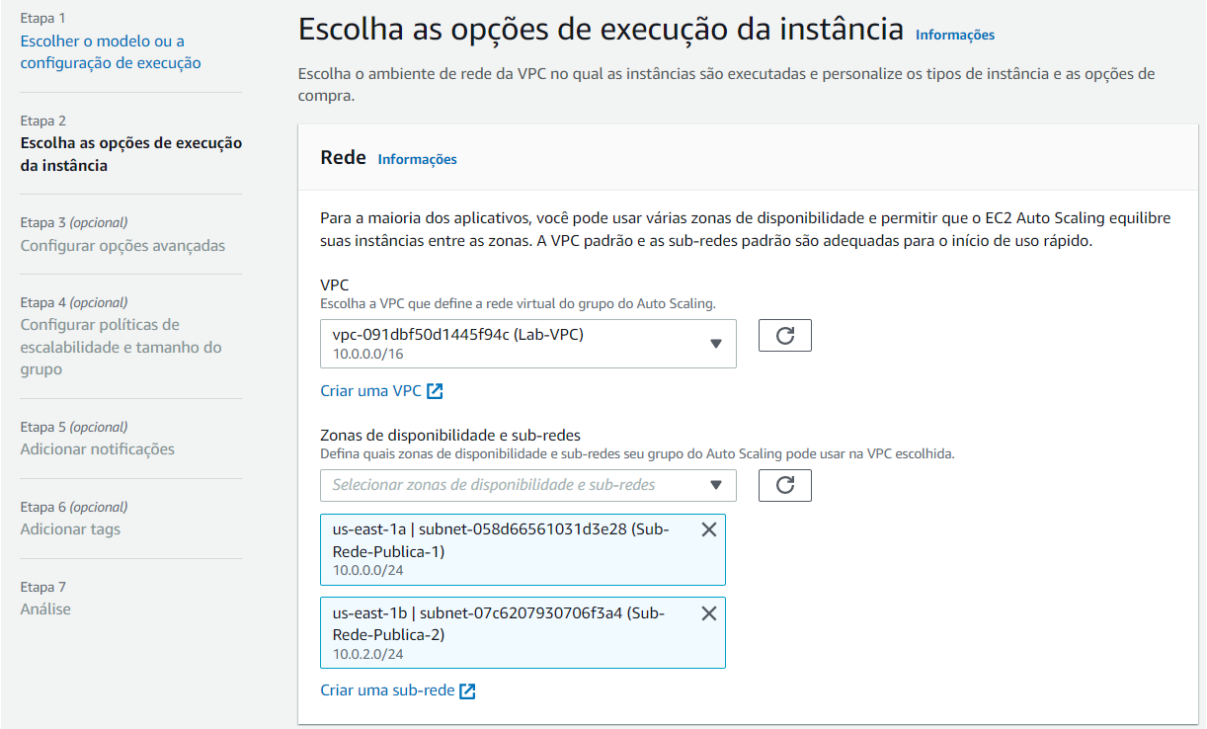
- Informe o nome: **GrupoAutoScaling**

- Escolha em "**Modelo de execução**": **ModeloServidorWebAPP**

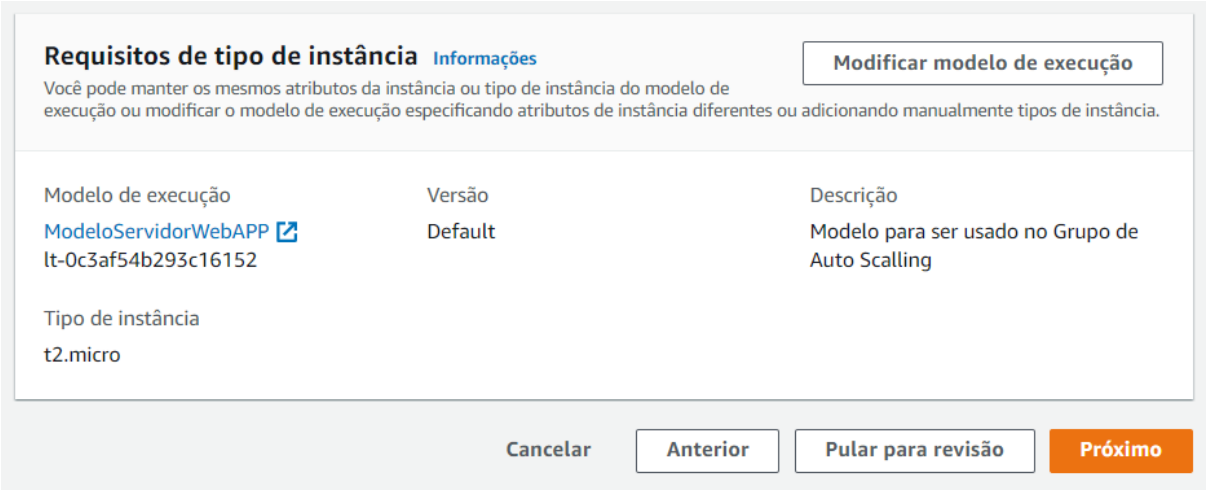


Manter versão "Default(1) e Clique em **Próximo**.

Etapa 2: Selecione sua VPC e AZ's:

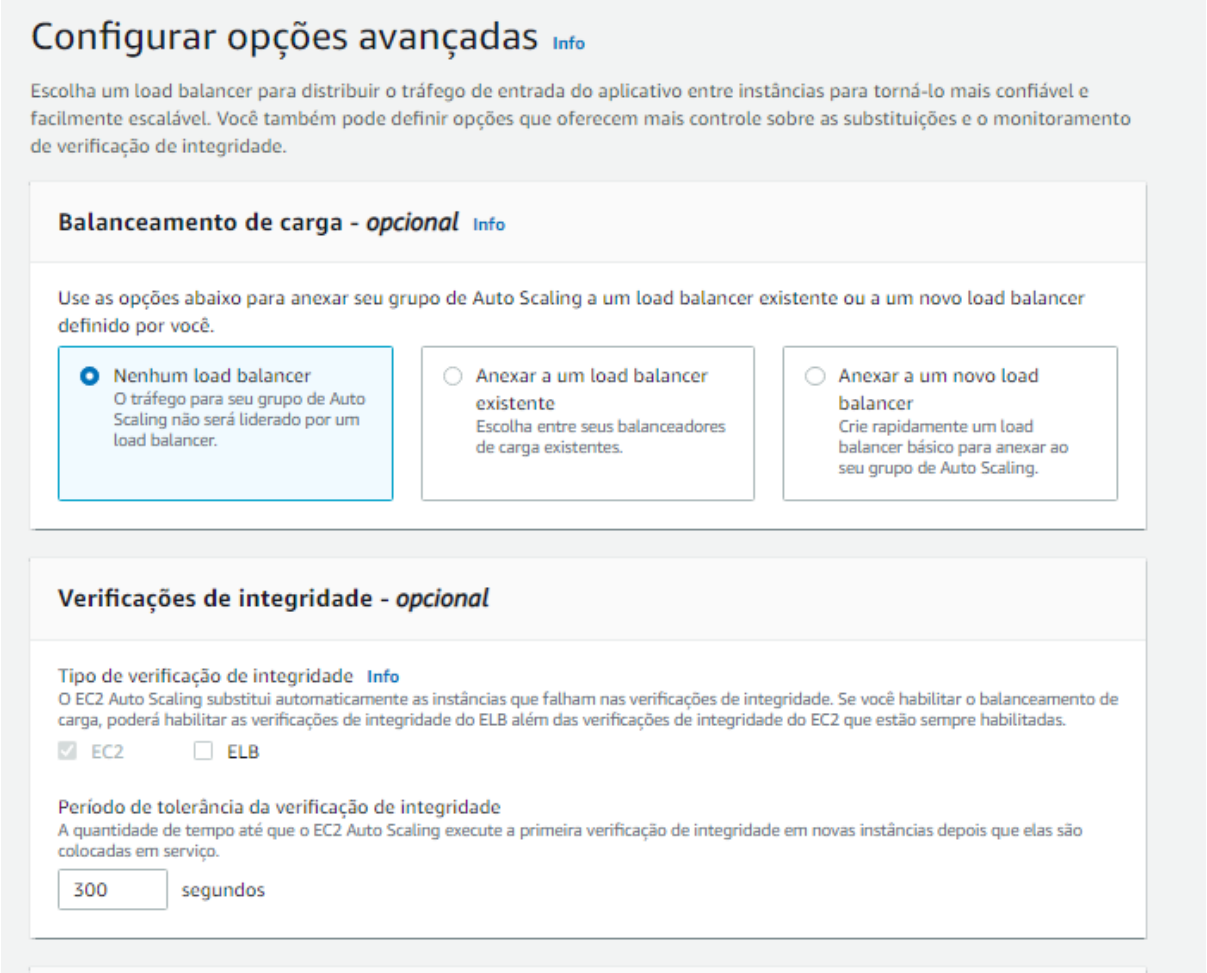


Usar a VPC **Lab-VPC** e as AZs **Sub-Rede-Publica-1** e **Sub-Rede-Publica-2**.

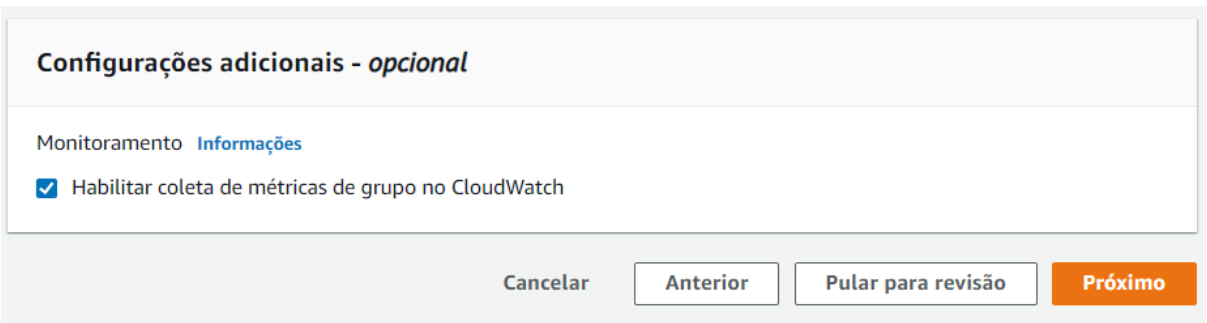


Clique em **Próximo**.

Etapa 3:

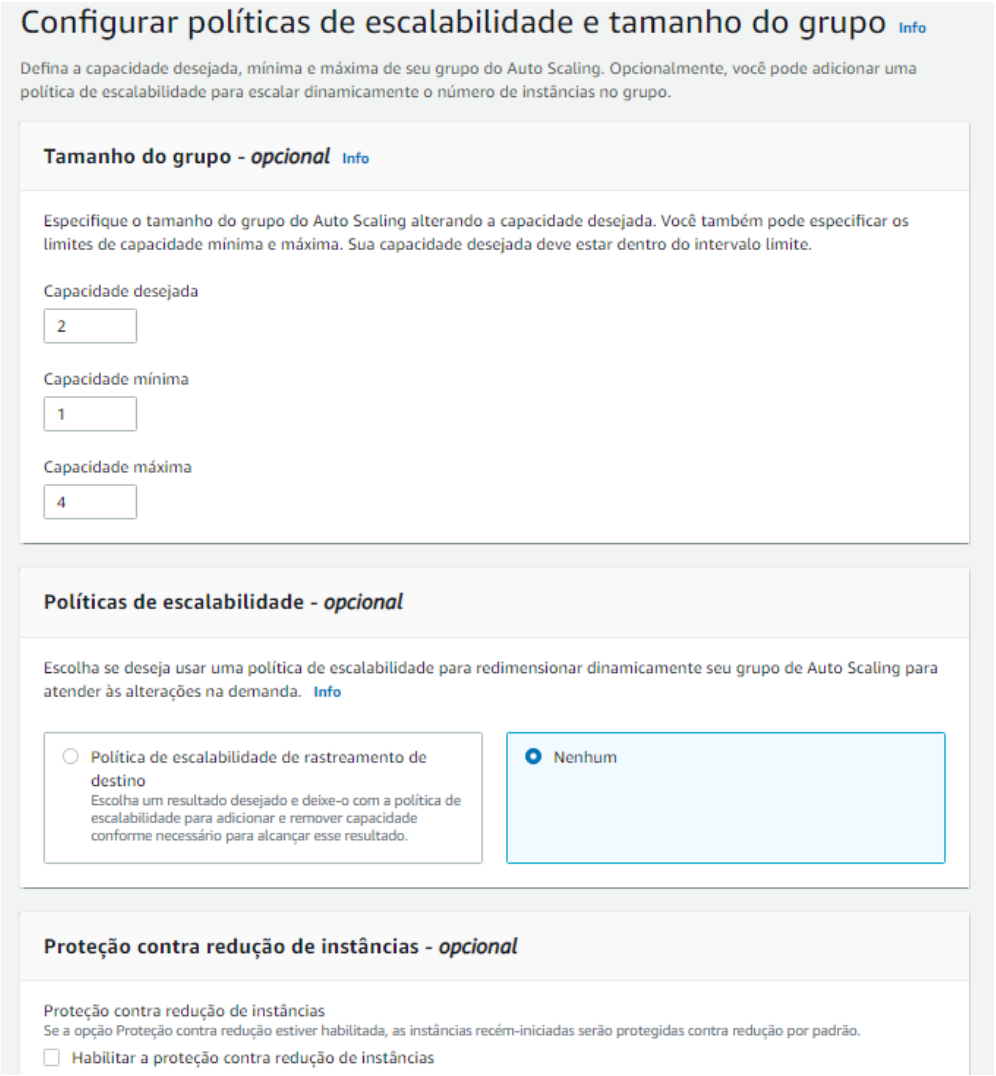


- Aqui, nas configurações avançadas, você pode anexar diretamente com um load balancer (balanceador de carga) existente.
- Porém, como esse conteúdo não contempla o ELB, não estaremos fazendo por enquanto.

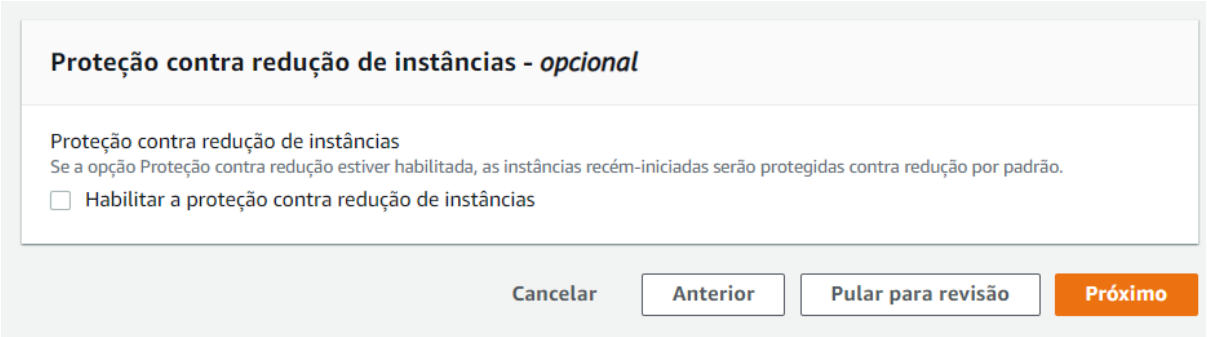


Habilitar coleta de métricas de grupo no CloudWatch e clique em **Próximo**.

Etapa 4:

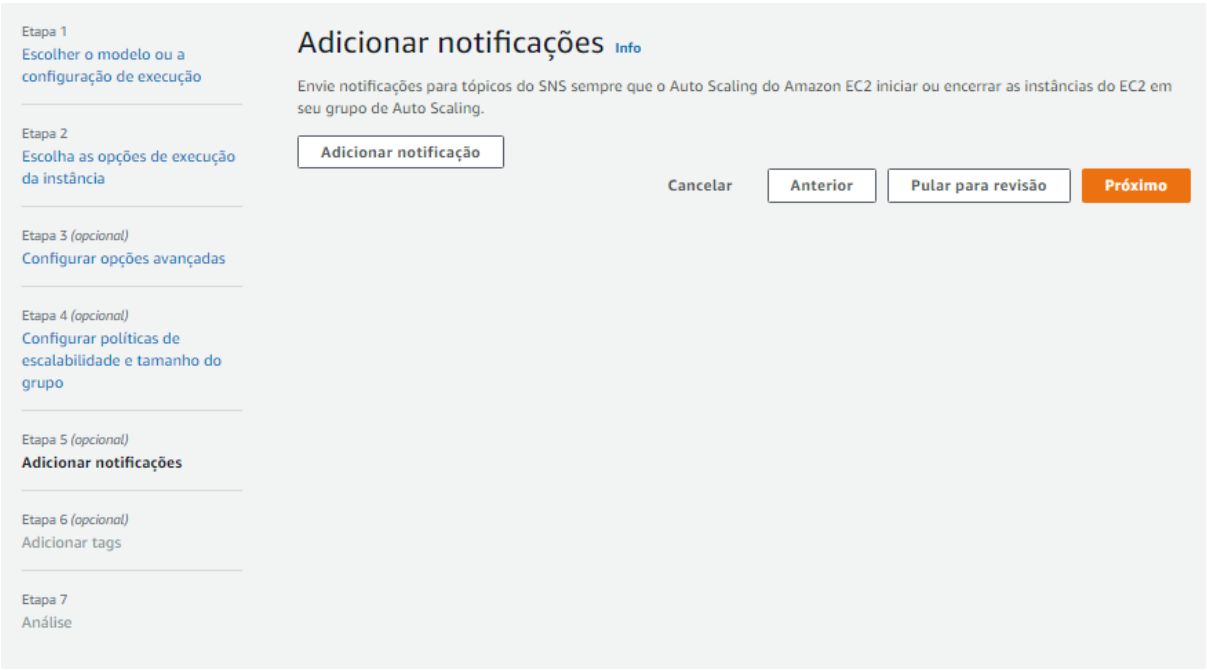


- Aqui se encontra a tela de configuração que foi representada pela imagem ilustrativa anteriormente, com as mesmas informações:
  - Capacidade mínima: 1
  - Capacidade desejada: 2
  - Capacidade máxima: 4
- Política de escalabilidade – opcional, manter marcado “Nenhum”.



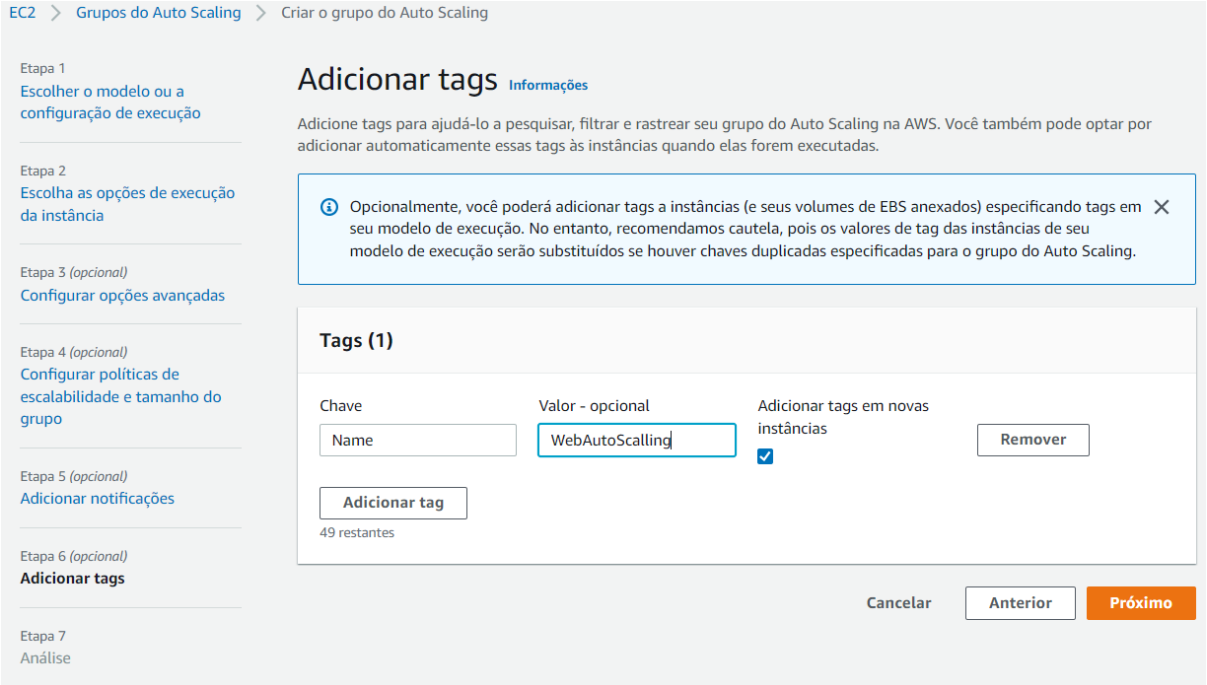
Desmarcar "Habilitar a proteção contra redução de instâncias" e clique em **Próximo**.

Etapa 5:



- Você também pode enviar notificações para tópicos do SNS (Simple Notification Service) mas neste caso não vamos usar.

Etapa 6 - Tags:

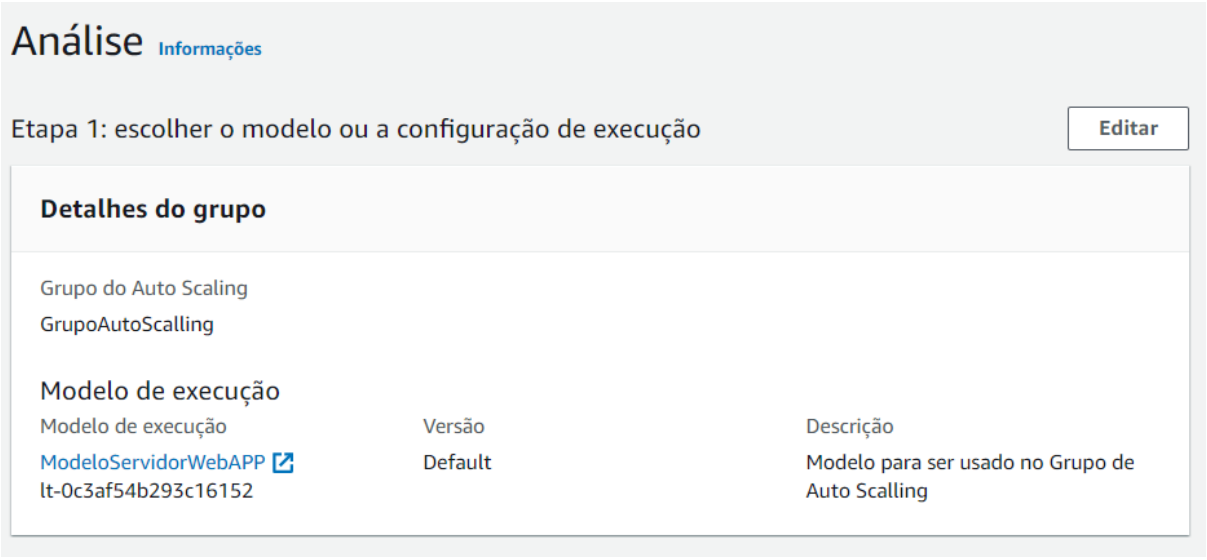


Adicionar Tag: **WebAutoScaling**

Adicionar tags às suas configurações na AWS são de extrema ajuda em casos de grandes empresas, podendo te poupar muito tempo na procura por erros e mais.

Etapa 7 - Revisão:

Aqui, você irá revisar suas configurações, mostrando as 6 etapas anteriores.



Etapa 2: Escolher as opções de execução de instâncias — Editar

Rede

Rede

VPC
vpc-091dbf50d1445f94c

Zona de disponibilidade	Sub-rede	
us-east-1a	subnet-058d66561031d3e28	10.0.0.0/24
us-east-1b	subnet-07c6207930706f3a4	10.0.2.0/24

Requisitos de tipo de instância

Esse grupo do Auto Scaling aderirá ao modelo de execução.

Etapa 3: Configurar opções avançadas — Editar

Balanceamento de carga

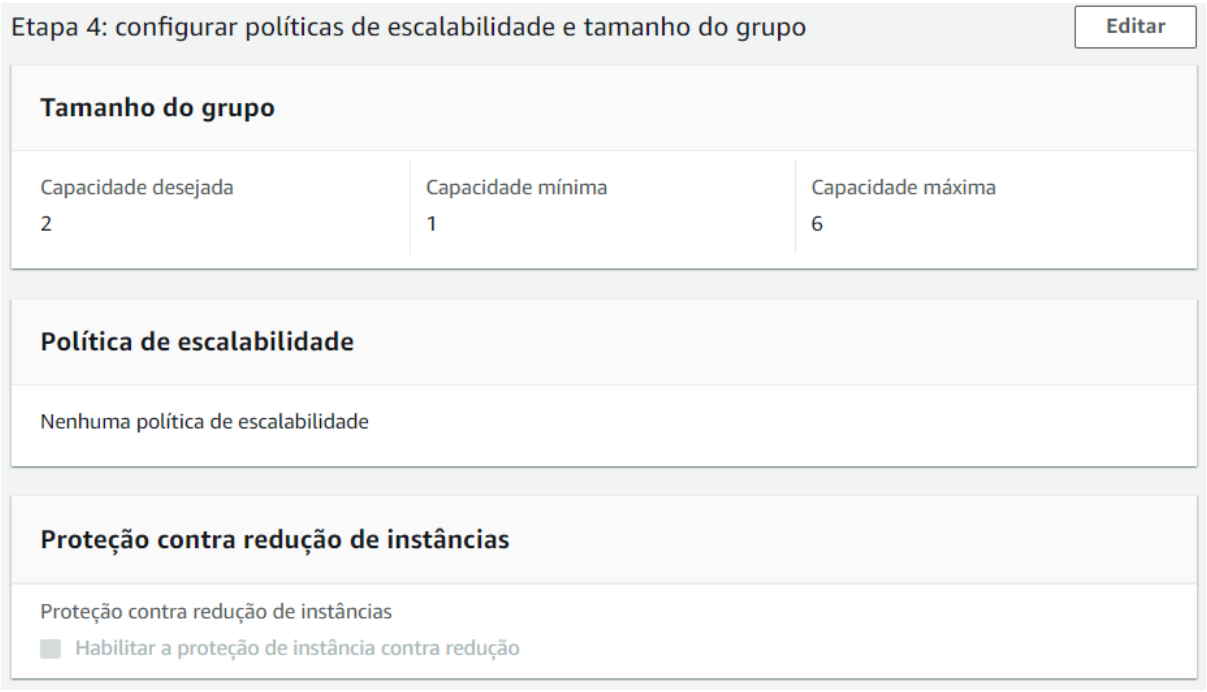
Verificações de integridade

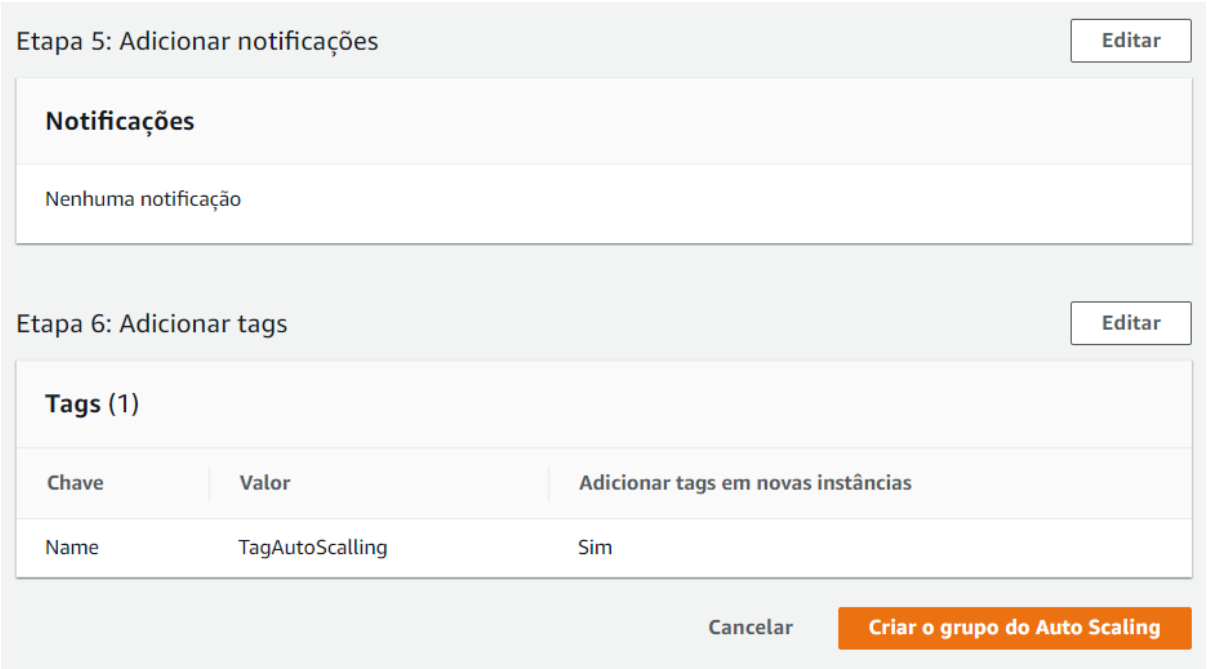
Tipo de verificação de integridade
EC2

Período de tolerância da verificação de integridade
300 segundos

Configurações adicionais

Monitoramento
Habilitado





Clique em "**Criar o grupo do Auto Scaling**".

Veja em **EC2** e **Instâncias** que as máquinas começaram a ser criadas:

Laboratório 13 - Aplicando Auto-Scaling e ambiente com muitos acessos

Este laboratório orienta você a usar os serviços Elastic Load Balancing (ELB) e Auto Scaling para balancear cargas e dimensionar automaticamente a infraestrutura.

O **Auto Scaling** ajuda a manter a disponibilidade da aplicação e permite aumentar ou reduzir a capacidade do Amazon EC2 de forma automática, de acordo com condições definidas por você. Você pode usar o Auto Scaling para ajudar a garantir que está executando o número desejado de instâncias do Amazon EC2. O Auto Scaling também pode aumentar automaticamente o número de instâncias do Amazon EC2 durante picos de demanda para manter a performance, e diminuir a capacidade durante períodos ociosos para reduzir os custos. O Auto Scaling é ideal para aplicações com padrões de demanda estáveis ou que passam por variações de utilização horárias, diárias ou semanais.

Objetivos

Depois de concluir este laboratório, você será capaz de:

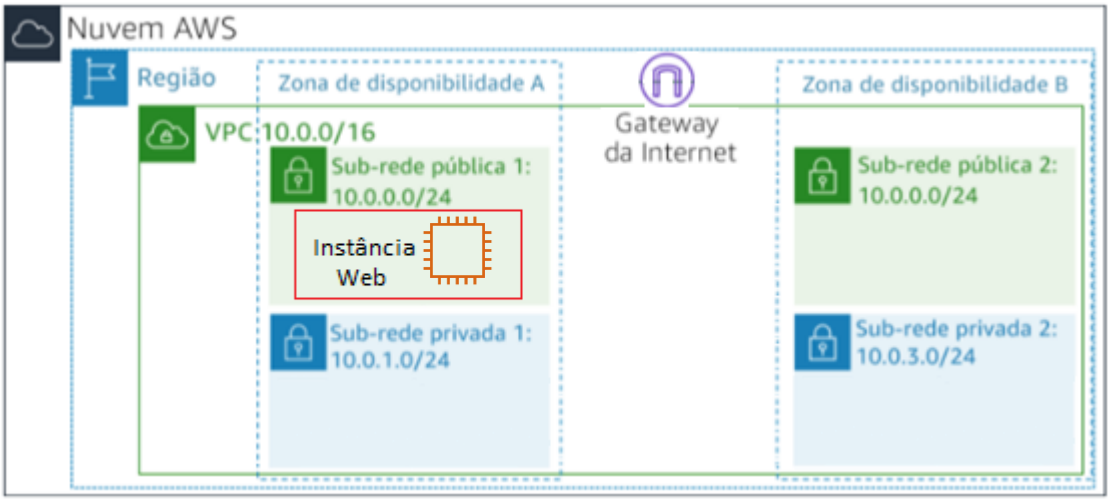
- Criar uma imagem de máquina da Amazon (AMI) por meio de uma instância em execução (Servidor Web APP Auto);
- Criar uma configuração de execução e um grupo de Auto Scaling;
- Ajustar a escala automaticamente de novas instâncias em uma sub-rede privada;
- Criar alarmes do Amazon CloudWatch e monitorar a performance da sua infraestrutura.

Duração

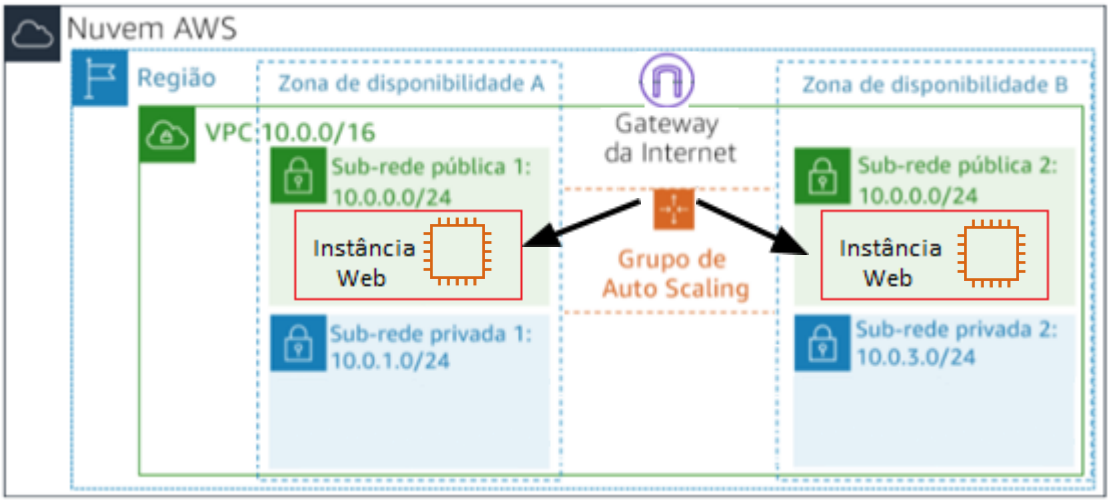
Este laboratório tem duração aproximada de **10 minutos**.

Cenário

Você começa com a seguinte infraestrutura:



O estado final da infraestrutura é:



Acessar o Console de Gerenciamento da AWS

1. Na parte superior destas instruções, clique em **Start Lab** (Iniciar laboratório) para iniciar o laboratório.

   Um painel Start Lab (Iniciar laboratório) é aberto exibindo o status do laboratório.

2. Aguarde até ver a mensagem "**Lab status: in creation** (Status do laboratório: em criação)" e clique no **X** para fechar o painel Start Lab (Iniciar laboratório).

   Observação: pode levar aproximadamente 20 minutos ou mais para que o status do laboratório mude para "ready" (pronto).

3. Na parte superior destas instruções, clique em AWS

   Isso abrirá o Console de Gerenciamento da AWS em uma nova guia do navegador. O sistema fará o seu login automaticamente.

   Dica: se uma nova guia do navegador não for aberta normalmente, um banner ou um ícone na parte superior do navegador indicará que o navegador está impedindo que o site abra janelas pop-up. Clique no banner ou ícone e escolha "Allow pop ups" (Permitir pop-ups).

4. Organize a guia do Console de Gerenciamento da AWS para que ela seja exibida com estas instruções. Em um cenário ideal, você poderá ver as duas guias do navegador ao mesmo tempo, para facilitar o acompanhamento das etapas do laboratório.

Já passamos pelos Laboratórios

Laboratório 10: Criar uma Instância EC2 Amazon Linux com APP

Laboratório 11: Criar um Modelo AMI (Amazon Machine Image) para o Auto Scaling

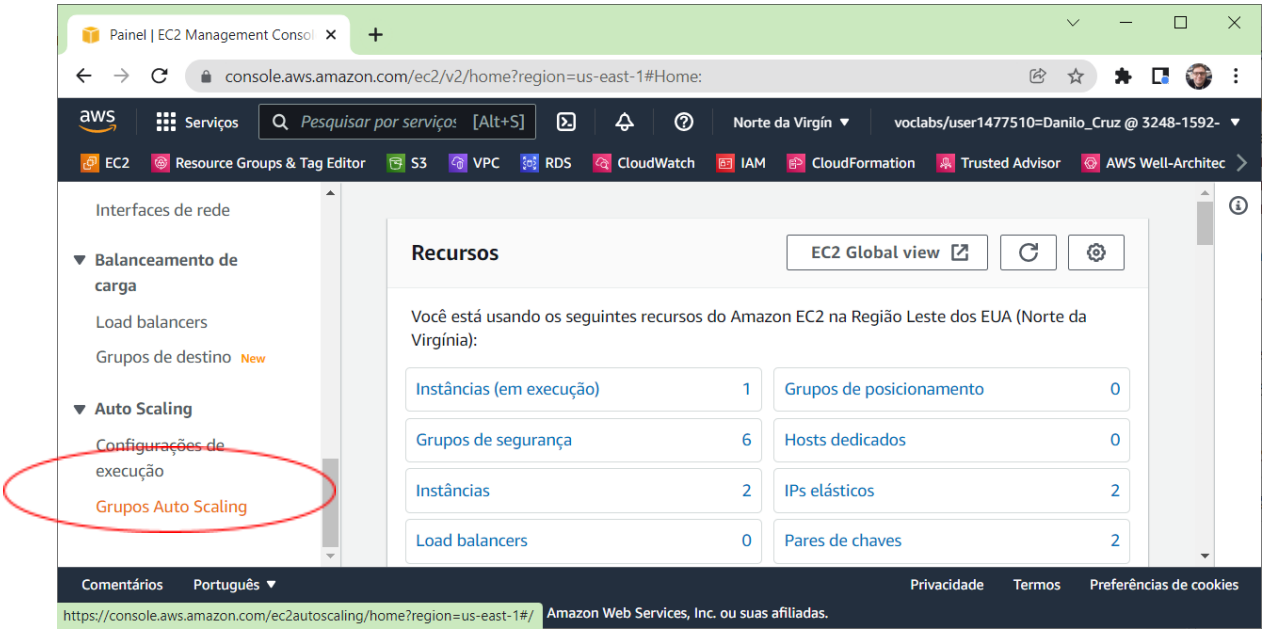
Laboratório 12: Criamos o Auto Scalling

Nesta tarefa, você criou regras de funcionamento do Auto Scalling onde identificou que 2 instâncias EC2 em zonas de disponibilidade diferente foram criadas.

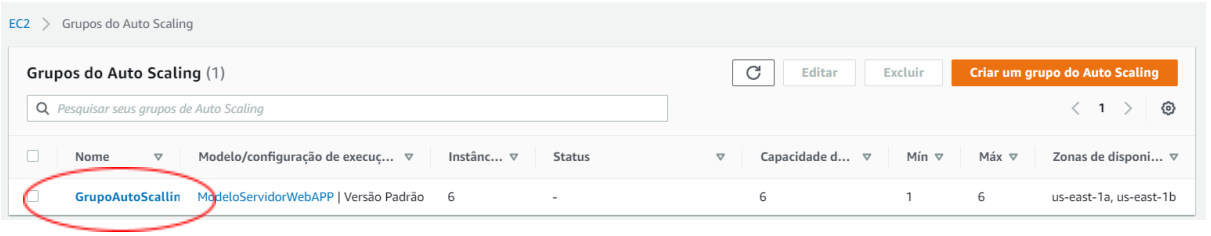
Laboratório 13 - Aplicando Auto-Scaling e ambiente com muitos acessos

Etapa 1 - Vamos criar uma polícia de escalabilidade automática

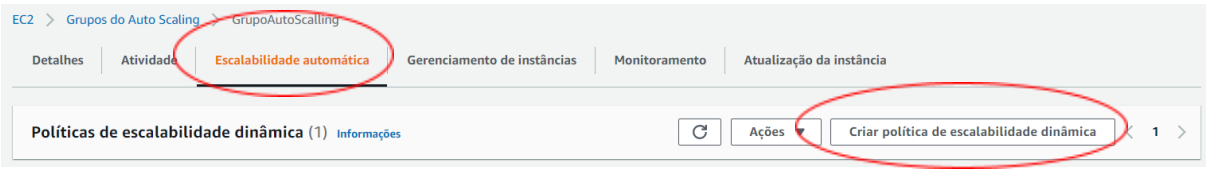
No Console **AWS**, em **EC2** clique em **Grupos do Auto Scaling:**



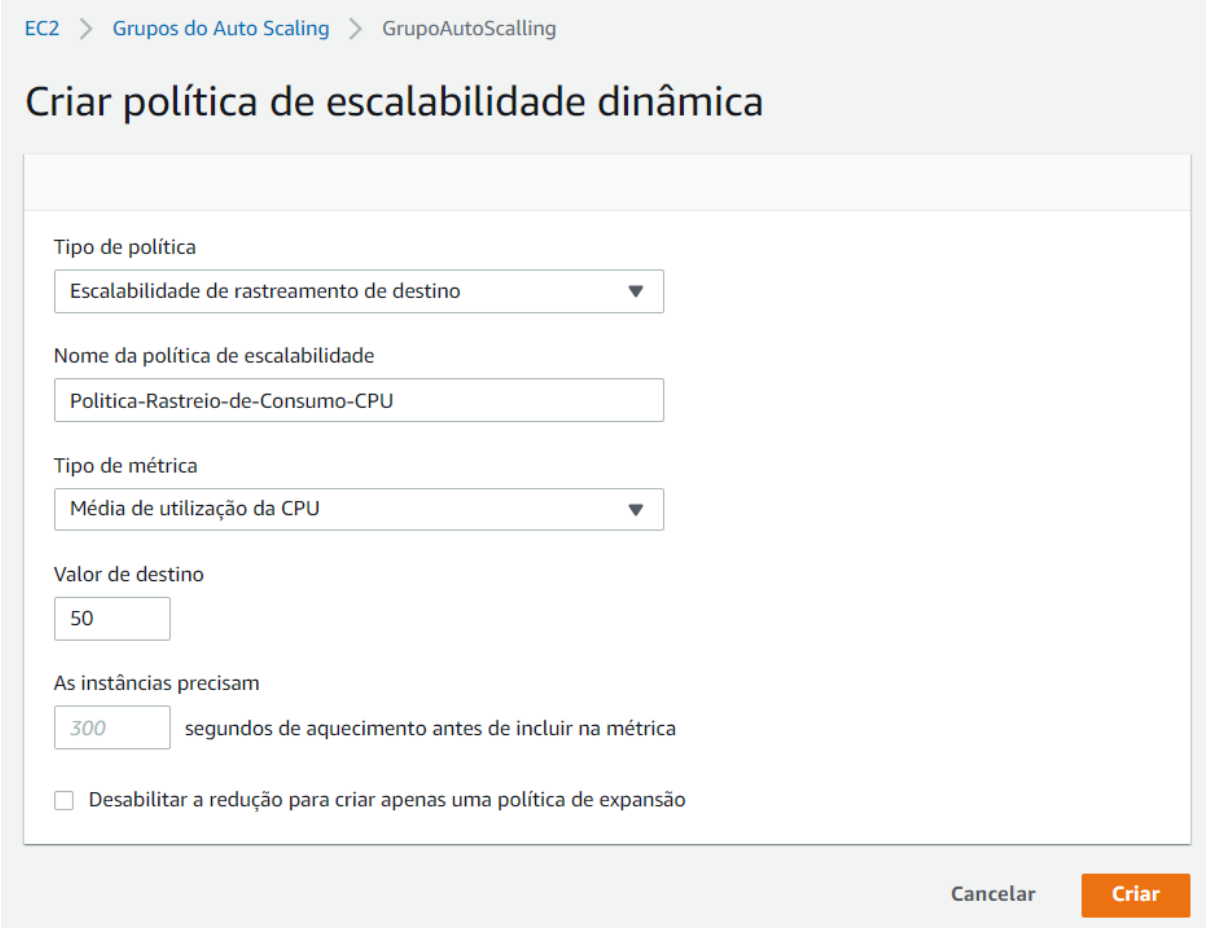
Clique em **GrupoAutoScaling**



Dentro do **GrupoAutoScaling** escolha a aba "**Escalabilidade automática**"



Clique em "**Criar política de escalabilidade dinâmica**"



Dentro de "Criar política de escalabilidade dinâmica

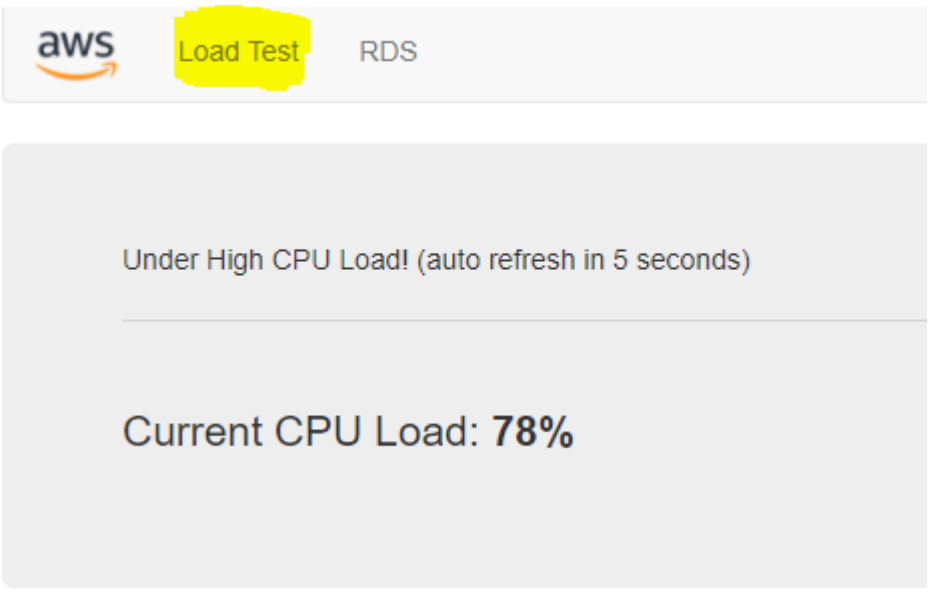
- Escolar o "Tipo de política" escolha: **Política de escalabilidade de rastreamento de destino**;
- Nome da política: **Politica-Rastreio-de-Consumo-CPU**

- Tipo de métrica: **Média de Utilização média da CPU**;
- Valor de destino: **50**

Isso informa ao Auto Scaling para manter uma utilização de CPU *média* em *todas as instâncias* de 50%.

O Auto Scaling adicionará ou removerá automaticamente a capacidade conforme necessário para manter a métrica no valor de destino especificado ou próximo a ele.

Acesse a guia do navegador com a aplicação Web APP de **cada instância**.



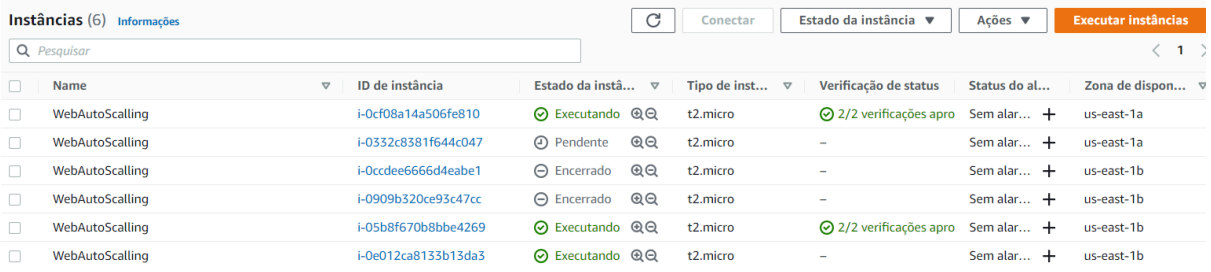
Clique em **Load Test** (Carregar teste) ao lado do logotipo da AWS.

A máquina irá começar a consumir muito processamento chegando em 100%.

Veja se a instância esta com IP Válido e DNS e faça acesso em cada uma pelo endereço:

http://ec2-107-22-24-147.compute-1.amazonaws.com/index.php

Veja que as instancias estão sendo criadas em média de 5 minutos.



Você deve ver duas novas instâncias chamadas de **WebAutoScalling**

Elas foram iniciadas pelo Auto Scaling.

Se as instâncias ou nomes não forem exibidos, aguarde 30 segundos e clique em atualizar no canto superior direito.

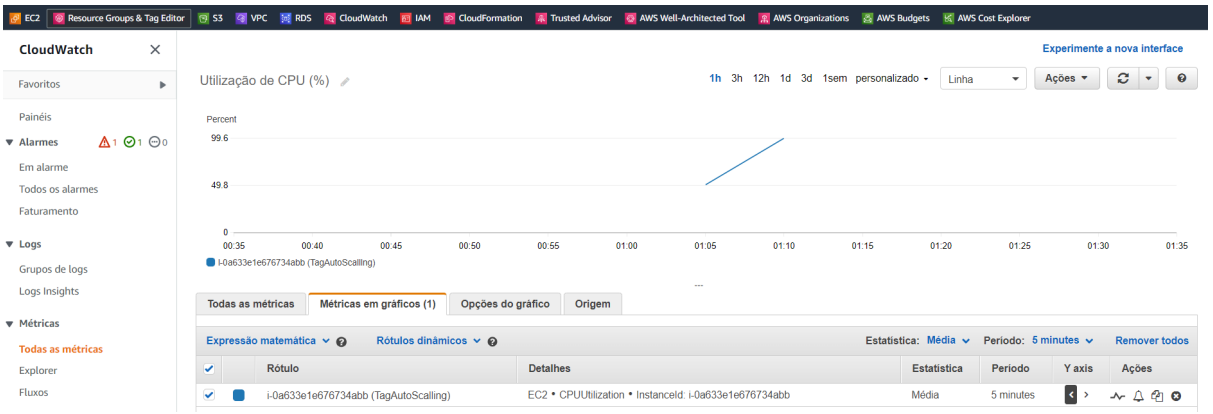
Primeiro, você confirmará que as novas instâncias foram aprovadas na verificação de integridade.

Tarefa 6: Testar o Auto Scaling

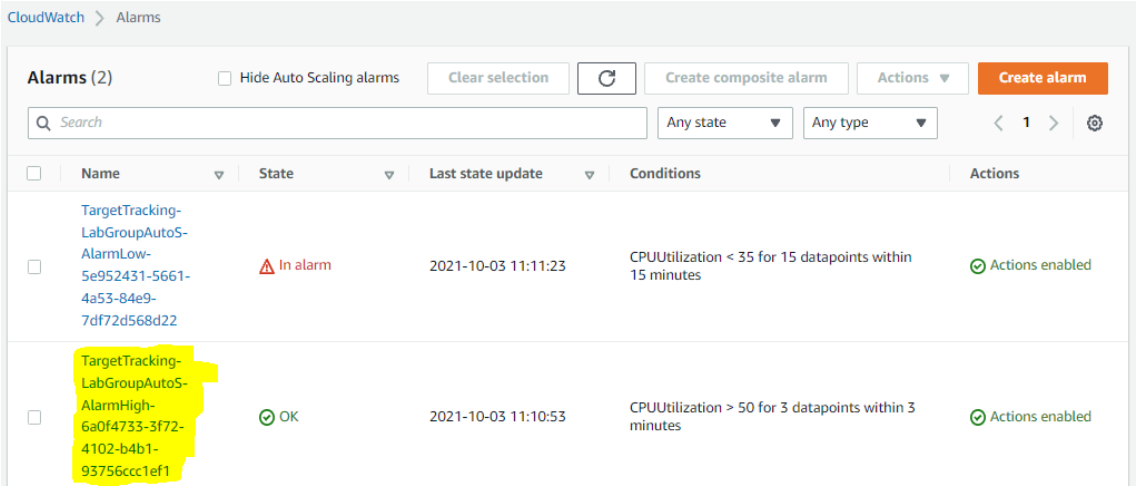
Você criou um grupo de Auto Scaling com um mínimo de duas instâncias e um máximo de seis instâncias. Atualmente, duas instâncias estão em execução porque o tamanho mínimo é duas e o grupo não está atualmente sob nenhuma carga. Agora, você aumentará a carga para fazer com que o Auto Scaling acrescente instâncias adicionais.

52. Volte para o Console de Gerenciamento da AWS, mas não feche a guia da aplicação. Você retornará a ele em breve.
53. No menu **Services** (Serviços), clique em **CloudWatch**.
54. No painel de navegação à esquerda, clique em **Alarms** (Alarmes) (*não em* **ALARM** (ALARME)).

Verifique no serviço do CloudWatch que aletas foram gerados.



Clique na guia **Alarmes**.



48. Aguarde até que o **Status** de ambas as instâncias mude para **Healthy** *(íntegro)*.

49. Clique em atualizar no canto superior direito para verificar se há atualizações.

O status **Healthy** *(íntegro)* indica que a instância passou na verificação de integridade.

Alarmes serão exibidos. Eles foram criados automaticamente pelo grupo de Auto Scaling.

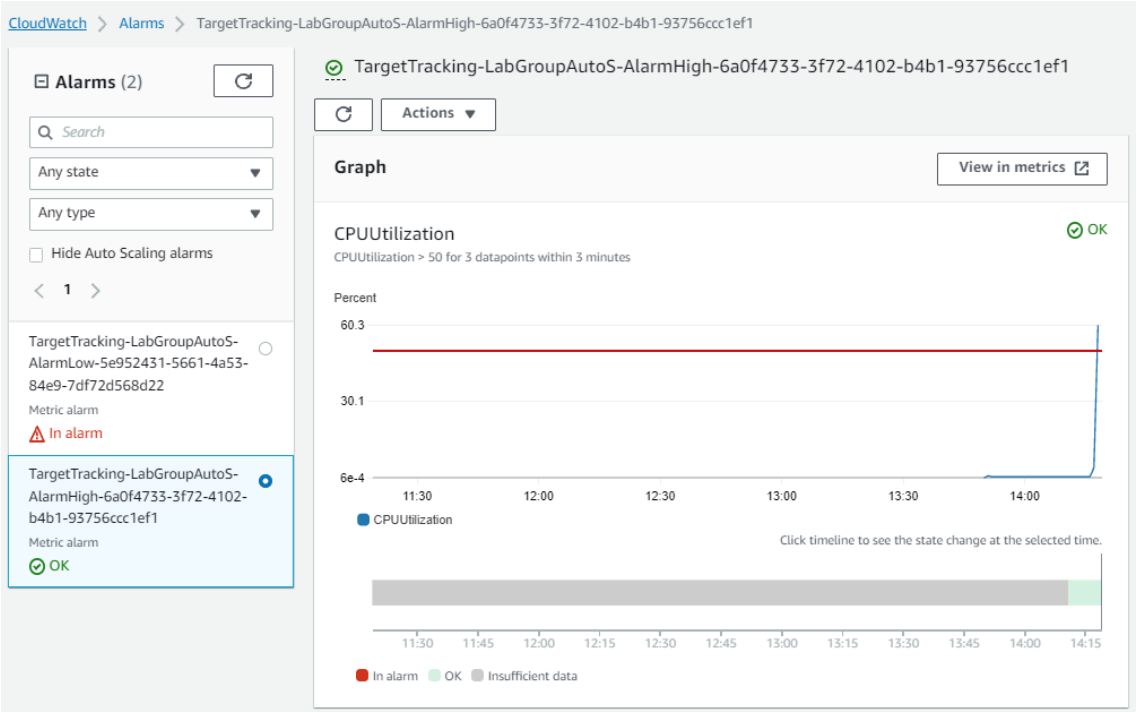
Eles manterão automaticamente a carga média da CPU próxima a 50%, permanecendo também dentro da limitação de ter duas a seis instâncias.

- No menu **Services** (Serviços), clique em **CloudWatch**.
- No painel de navegação à esquerda, clique em **All Alarms** (Todos os Alarmes) (*não em* **ALARM** (ALARME)) e confirme se vê dois alarmes.

55. Clique no alarme **OK**, que tem *AlarmHigh (Alarme alto)* no nome.

Se nenhum alarme estiver mostrando **OK**, aguarde um minuto e clique em Refresh (Atualizar) no canto superior direito até que o status do alarme mude.

O **OK** indica que o alarme *não* foi acionado.



Em menos de 5 minutos, o alarme **AlarmLow** (Alarme baixo) deverá mudar para **OK** e o status do alarme **AlarmHigh** (Alarme alto) deverá mudar para *ALARM* (ALARME).

Você pode clicar em Refresh (Atualizar) no canto superior direito a cada 60 segundos para atualizar a exibição.

Você deve ver o gráfico **AlarmHigh**(Alarme alto) indicando uma porcentagem crescente de CPU. Depois de passar da linha de 60% por mais de 3 minutos, o Auto Scaling acionará a adição de novas instâncias.

59. Aguarde até que o alarme **AlarmHigh** (Alarme alto) entre no estado *ALARM (ALARME)*.

Agora você pode visualizar as instâncias adicionais que foram executadas.

60. No menu **Services** (Serviços), clique em **EC2**.
61. No painel de navegação esquerdo, clique em **Instances** (Instâncias).

Agora deve haver mais de duas instâncias rotuladas **Lab Instance** (Instância do laboratório) em execução. As novas instâncias foram criadas pelo Auto Scaling em resposta ao alarme.

Remover Grupos de Auto Scaling, AMIs e Instâncias

Exercícios do Módulo

1. Você deseja redefinir a configuração de instâncias EC2 que você planeja iniciar manualmente e usando o Auto Scaling. Que recurso você deve usar??
a) Launch configuration (Configuração de lançamento).
b) Modelo CloudFormation.
c) Papel de instância.
d) Launch template (Modelo de lançamento).
e) Modelo Lambda.

2. Qual o parâmetro de grupo do Auto Scaling define o limite para o número de instâncias que o Auto Scaling cria?
a) Tamanho do grupo.
b) Capacidade desejada.
c) Capacidade mínima.
d) Capacidade limitante.
e) Capacidade máxima.

3. Um grupo de Autoscaling tem uma capacidade desejada de 7 e um tamanho máximo de 7. O que o Auto Scaling fará se alguém encerrar manualmente uma dessas instâncias?
a) Ele iniciará uma nova instância.
b) Isso mudará a capacidade desejada para 6.
c) Ele encerra uma instância.
d) Não iniciarão nenhuma nova instância.
e) Isso mudará a capacidade máxima para 6.

4. Como uma empresa pode habilitar elasticidade para um aplicativo em execução no Amazon EC2?
a) Usando balanceamento de carga elástica.
b) Usando o Amazon EC2 Auto Scaling.
c) Configurando multi-AZ.
d) Ao habilitar o failover no Amazon EC2.
e) Configurando um bucket no S3.

5. Como o Amazon EC2 Auto Scaling ajuda no custo-benefício?
a) Escolhendo o tipo de instância com melhor custo-benefício.
b) Equilibrando a carga entre instâncias uniformemente.
c) Ao iniciar e encerrar instâncias à medida que a demanda muda.

d) Por falha de aplicativo automaticamente.

e) Ao iniciar as instâncias à medida que a demanda diminui e encerrar instâncias à medida que a demanda cresce.

6. Dentre as afirmações a seguir, qual NÃO é verdade?

a) O Auto Scaling pode ajudar a diminuir os custos.

b) O Auto Scaling pode distribuir carga igualmente para as instâncias EC2.

c) O Auto Scaling te ajuda a ter menos problemas com instâncias ociosas e sem necessidade na sua topologia.

d) O Auto Scaling pode tanto criar novas instâncias quanto encerrar de acordo com a sua demanda.

e) O Auto Scaling tem número de instâncias mínimo, desejado e máximo totalmente customizáveis.

Gabarito

1. Letra D. É no Launch Template (Modelo de Lançamento) que você define essas informações, como a AMI usada, o tipo da instância, o par de chaves e mais.
2. Letra E. Ao definir a capacidade máxima, você está definindo o limite de instâncias que podem ser executadas.
3. Letra A. A capacidade desejada não mudaria nesse caso, sendo assim, o Auto Scaling iniciaria uma nova instância para satisfazer as 6 desejadas.
4. Letra B. O serviço responsável por aumentar a elasticidade dentre os citados é o Auto Scaling.
5. Letra C. Ele ajuda no custo-benefício por conseguir iniciar as instâncias à medida que a demanda cresce e encerrar à medida que a demanda diminui, por isso não se encaixa na letra e.
6. Letra B. Essa alternativa está incorreta, visto que o papel de balancear carga é do Load Balancer, e não do Auto Scaling.