# **Cloud Computing | Aula 15**

## **Material do Módulo 1 - Conhecendo as categorias de produtos da Nuvem**

Este módulo aborda os seguintes tópicos:

* **Tópico 1 - Introdução ao Load Balancer e Tipos de Load Balancer;**
* **Laboratório 1 - Criando ALB – Application Load Balancer;**
* **Exercícios do Módulo 1** - Por fim, você deverá concluir um teste de conhecimento que será usado para testar sua compreensão dos principais conceitos abordados neste módulo.

## **Referência Bibliográfica**

Introdução a Load Balancer -

<https://docs.aws.amazon.com/pt_br/elasticloadbalancing/latest/userguide/what-is-load-balancing.html>

Tipos de Load Balancer - <https://docs.aws.amazon.com/pt_br/AmazonECS/latest/userguide/load-balancer-types.html>

Erro 503 Service Temporary Unavailable -

<https://nadtakan-futhoem.medium.com/aws-load-balancer-503-service-temporarily-unavailable-e1e91c0dfcdb>

## **Tópico 1 - Introdução ao Load Balancer**

**O que é o Elastic Load Balancing?**

O Elastic Load Balancing distribui automaticamente seu tráfego de entrada entre vários destinos, como instâncias do EC2, contêineres e endereços IP, em uma ou mais zonas de disponibilidade. Ele monitora a integridade dos destinos registrados e roteia o tráfego apenas para os destinos íntegros.

O Elastic Load Balancing escala seu balanceador de carga conforme seu tráfego de entrada muda com o tempo. Ele pode ser dimensionado automaticamente para a vasta maioria das cargas de trabalho.

**Benefícios do load balancer**

Um load balancer distribui cargas de trabalho para vários recursos computacionais, como servidores virtuais. Usar um load balancer aumenta a disponibilidade e a tolerância a falhas dos aplicativos.

Adicione e remova recursos computacionais do load balancer conforme mudarem suas necessidades, sem perturbar o fluxo geral de solicitações para os aplicativos.

Configure as verificações de integridade, que monitoram a integridade dos recursos computacionais, para que o load balancer envie solicitações somente para as instâncias íntegras. Também é possível descarregar o trabalho de criptografia e descriptografia no load balancer, para que os recursos computacionais possam se concentrar no trabalho principal.

**Acesso ao Elastic Load Balancing**

Você pode criar, acessar e gerenciar seus load balancers usando qualquer uma das interfaces a seguir:

**AWS Management Console**— Fornece uma interface da Web que pode ser usada para acessar o Elastic Load Balancing.

**AWS Interface de linha de comando (AWS CLI)** — Fornece comandos para um amplo conjunto de serviços AWS, incluindo Elastic Load Balancing. A AWS CLI é compatível com Windows, macOS e Linux.

**AWS SDKs** fornece APIs específicas da linguagem e cuida de muitos dos detalhes da conexão, como calcular assinaturas, lidar com novas tentativas de solicitação e tratamento de erros.

**API de consulta:** fornece ações de API de baixo nível que são chamadas usando solicitações HTTPS. Usar a API de consulta é a maneira mais direta de acessar o Elastic Load Balancing. No entanto, a API de consulta requer que o aplicativo lide com detalhes de baixo nível, como gerar o hash para assinar a solicitação e tratamento de erros.

**Serviços relacionados**

O Elastic Load Balancing funciona com os serviços a seguir para melhorar a disponibilidade e a escalabilidade dos seus aplicativos.

* Amazon EC2
* Auto Scaling do Amazon EC2
* Amazon CloudWatch
* Amazon ECS
* Route 53
* AWS WAF

## **Tipos de Load Balancer**

**Recursos do Elastic Load Balancing**

O Elastic Load Balancing oferece suporte aos seguintes load balancers: **Application Load** **Balancers**, **Network Load Balancers**, **balanceadores de carga do Gateway**

**Application Load Balancers**

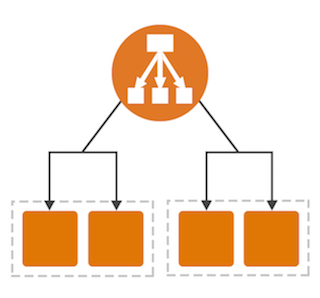
Um balanceador de carga da aplicação toma decisões de roteamento na camada da aplicação (HTTP/HTTPS), oferece suporte ao roteamento com base em caminho e pode encaminhar solicitações para uma ou mais portas em cada instância de contêiner no cluster.

Os balanceadores de carga da aplicação oferecem suporte ao mapeamento de porta de host dinâmico.

Por exemplo, se a definição de contêiner da tarefa especifica a porta 80 para uma porta de contêiner NGINX e a porta 0 para a porta do host, a porta do host é escolhida dinamicamente com base no intervalo de portas temporário da instância de contêiner (como entre 32768 e 61000 na AMI otimizado para Amazon ECS mais recente).

Quando a tarefa é inicializada, o contêiner NGINX é registrado no balanceador de carga da aplicação como uma combinação de ID e porta da instância, e o tráfego é distribuído para o ID e para a porta da instância correspondentes a este contêiner.

Esse mapeamento dinâmico permite várias tarefas de um único serviço na mesma instância de contêiner. Para obter mais informações, consulte o Guia do usuário para balanceadores de carga da aplicação.



Fonte: <https://docs.aws.amazon.com/pt_br/AmazonECS/latest/userguide/load-balancer-types.html>

**Network Load Balancers**

Um Network Load Balancer funciona na quarta camada do modelo Open Systems Interconnection (OSI).

Ele pode processar milhões de solicitações por segundo. Após o load balancer receber uma solicitação de conexão, ele seleciona um destino no grupo de destino para a regra padrão.

Ele tenta abrir uma conexão TCP para o destino selecionado na porta especificada na configuração do listener.

Quando você habilitar uma Zona de disponibilidade para o load balancer, o Elastic Load Balancing criará um nó do load balancer na Zona de disponibilidade.

Por padrão, cada nó do load balancer distribui tráfego aos destinos registrados somente na sua zona de disponibilidade.

Se você habilitar o balanceamento de carga entre zonas, cada nó do load balancer distribuirá o tráfego aos destinos registrados em todas as zonas de disponibilidade habilitadas. Para obter mais informações, consulte as zonas de disponibilidade.

Se você habilita várias zonas de disponibilidade para o seu load balancer e garante que cada grupo de destino tem pelo menos um destino em cada zona de disponibilidade habilitada, isso aumenta a tolerância a falhas dos aplicativos. Por exemplo, se um ou mais grupos de destino não têm um destino íntegro em uma zona de disponibilidade, removemos o endereço IP da sub-rede correspondente do DNS, mas os nós do load balancer em outras zonas de disponibilidade permanecerão disponíveis para rotear o tráfego. Se um cliente não honrar o tempo de vida (TTL) e enviar solicitações para o endereço IP depois que ele for removido do DNS, as solicitações falharão.

Para o tráfego TCP, o load balancer seleciona um destino usando um algoritmo de hash de fluxo baseado no protocolo, no endereço IP de origem, na porta de origem, no endereço IP de destino, na porta de destino e no número de sequência do TCP. As conexões TCP de um cliente têm diferentes portas de origem e números de sequência e podem ser direcionadas para destinos diferentes. Cada conexão TCP individual é roteada para um único destino para a vida útil da conexão.

Para o tráfego UDP, o load balancer seleciona um destino usando um algoritmo de hash de fluxo baseado no protocolo, no endereço IP de origem, na porta de origem, no endereço IP de destino e na porta de destino. Um fluxo UDP tem a mesma origem e o mesmo destino, portanto, ele é roteado de forma consistente para um único destino durante toda sua vida útil. Diferentes fluxos UDP têm diferentes portas e endereços IP de origem. Assim, eles podem ser roteados para destinos diferentes.

O Elastic Load Balancing cria uma interface de rede para cada Zona de disponibilidade que você habilitar. Cada nó de load balancer na Zona de disponibilidade usa essa interface de rede para obter um endereço IP estático. Quando você criar um load balancer voltado para a Internet, opcionalmente, poderá associar um endereço IP elástico por sub-rede.

Ao criar um grupo de destino, você especifica o tipo de destino, o que determina como você registra vários destinos.

Por exemplo, você pode registrar IDs de instância, endereços IP ou um Application Load Balancer.

Se você registrar destinos pelo ID da instância, os endereços IP de origem dos clientes serão preservados e fornecidos para seus aplicativos. Se você registrar destinos pelo endereço IP, os endereços IP de origem serão os endereços IP privados dos nós do load balancer.

Se você registrar um Application Load Balancer como um destino, os endereços IP de origem dos clientes serão preservados e fornecidos para seus aplicativos.

Você pode adicionar e remover destinos do load balancer conforme mudarem suas necessidades, sem perturbar o fluxo geral de solicitações para seu aplicativo. O Elastic Load Balancing escala seu balanceador de carga à medida que o tráfego para sua aplicação muda com o tempo. O Elastic Load Balancing pode ser escalado para a vasta maioria de workloads automaticamente.

Você pode configurar verificações de integridade, que são usadas para monitorar a integridade dos destinos registrados, para que o load balancer possa enviar solicitações apenas para os destinos íntegros.

**Balanceadores de carga do Gateway**

Os balanceadores de carga de gateway permitem que você implante, escale e gerencie dispositivos virtuais, como firewalls, sistemas de detecção e prevenção de intrusões e sistemas de inspeção profunda de pacotes.

Eles combinam um gateway de rede transparente (ou seja, um único ponto de entrada e saída para todo o tráfego) e distribui o tráfego enquanto escala os dispositivos virtuais segundo a demanda.

Um balanceador de carga de gateway opera na terceira camada do modelo Open Systems Interconnection (OSI), a camada de rede. Ele escuta todos os pacotes IP em todas as portas e encaminha o tráfego para o grupo de destino especificado na regra do listener.

Ele mantém a perdurabilidade dos fluxos para um dispositivo de destino específico usando 5 tuplas (para fluxos TCP/UDP) ou 3 tuplas (para fluxos não TCP/UDP). O balanceador de carga de gateway e suas instâncias de dispositivos virtuais registradas trocam tráfego de aplicações usando o protocolo GENEVE na porta 6081. Ele oferece suporte a um tamanho de unidade de transmissão máxima (MTU) de 8.500 bytes.

Os balanceadores de carga de gateway usam endpoints do balanceador de carga de gateway para trocar tráfego com segurança através dos limites da VPC. Um endpoint do balanceador de carga de gateway é um endpoint da VPC que fornece conectividade privada entre dispositivos virtuais na VPC do provedor de serviços e servidores de aplicações na VPC do consumidor de serviços. Implante o balanceador de carga de gateway na mesma VPC dos dispositivos virtuais.

Registre os dispositivos virtuais em um grupo de destino para o balanceador de carga de gateway.

## **Laboratório 1 - Criando ALB – Application Load Balancer**

**1 - Criar Grupo de Destino**

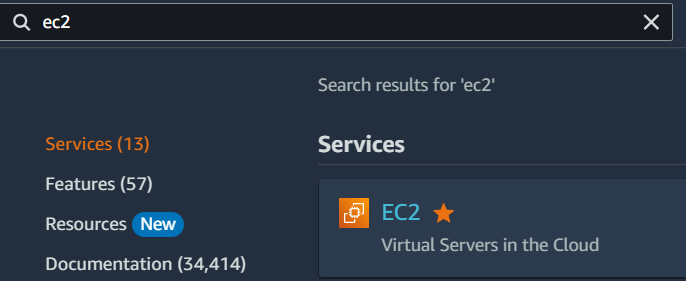
O roteamento configura para onde enviar as solicitações que chegam no balanceador de carga.

Com o ambiente dos Lab de Auto Scaling pronto, prosseguiremos agora adicionando balanceadores de carga de diferentes tipos, caso ainda não tenha feito ou tenha excluído/interrompido esse laboratório, refaça-o, pois será nossa base.

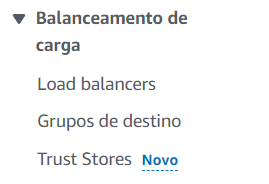
Nesta tarefa, você criará um balanceador de carga que pode balancear o tráfego entre várias instâncias EC2 e zonas de disponibilidade.

Você irá criar um **Criar grupo de destino** que será usado pelo Load Balancer.

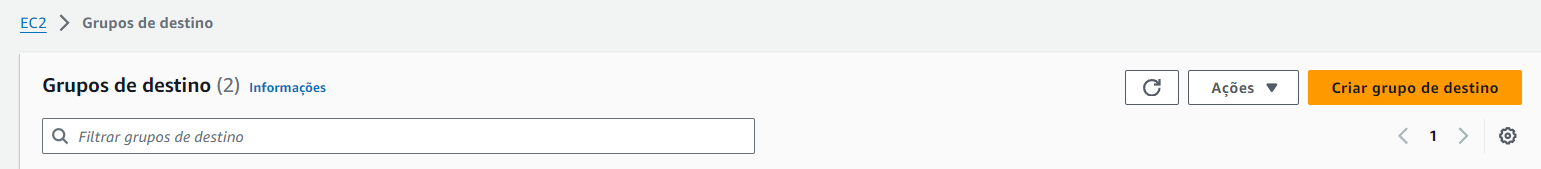
Primeiro, abra o Painel da EC2.



Procure no menu por **Grupos de destino**.



Clique em **Criar grupo de destino**



Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Preencher o nome: **LabGroupRouting**

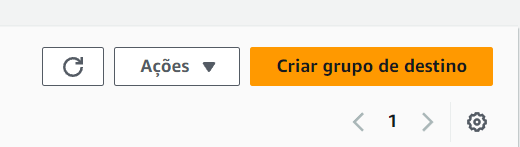
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Não mudar nada e clicar em **Próximo** e depois clique em **Criar grupo de destino**.



Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

O Grupo de destino **LabGroupRouting** foi criado.

**2 - Criar Load Balancer**

No painel de navegação à esquerda, dentro da categoria **Balanceamento de carga** clique em **Load Balancers.**



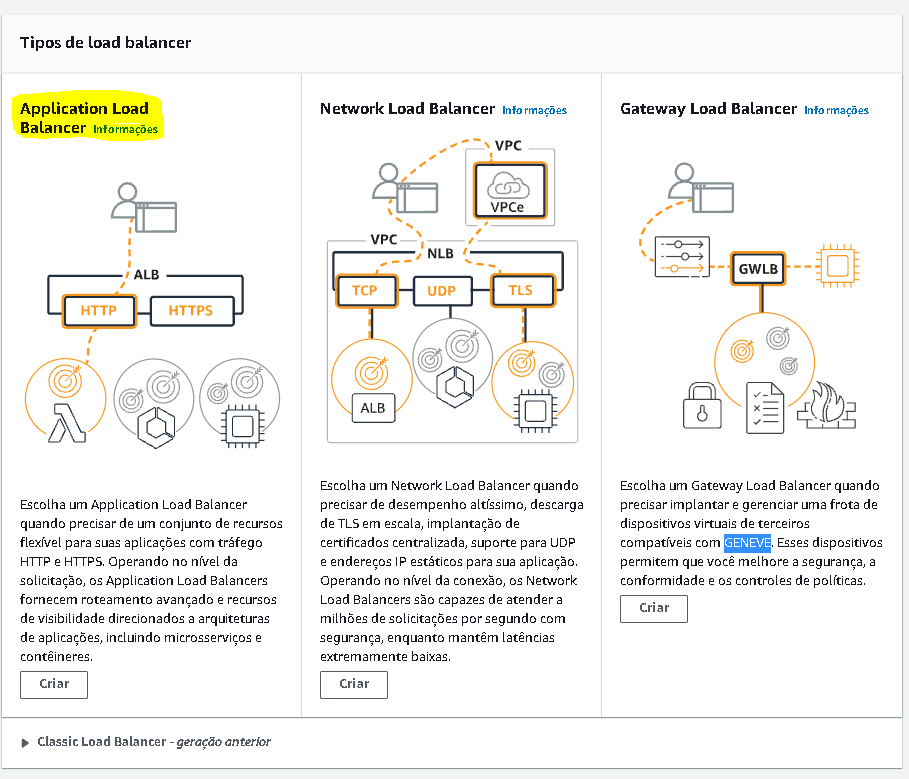
1. Clique em **Criar Load Balancer**

****

Vários tipos diferentes de balanceador de carga são exibidos.

Você usará um ***Application Load Balancer*** que opera no nível de solicitação (camada 7), roteando o tráfego para os destinos (instâncias EC2, contêineres, endereços IP e funções do Lambda) com base no conteúdo da solicitação.

Para obter mais informações, consulte: [Comparação de balanceadores de carga](https://aws.amazon.com/elasticloadbalancing/features/#compare)



Em **Tipos de Load Balancer**, clique em Criar em configuração básica:

* **Name** (Nome): **Lab-ELB-ALB**

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

* **VPC:** **Lab-VPC**

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

* **Zonas de disponibilidade**): selecione ambas para ver as sub-redes Públicas disponíveis;
* Selecione **Sub-Rede-Publica-1** e **Sub-Rede-Publica-2**;

Isso configura o balanceador de carga para operar em várias zonas de disponibilidade.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Um *Web Security Group (Grupo de segurança da Web)* já foi criado para você, o que permite o acesso HTTP.

1. Selecione **GrupoSec-ServidoresLinux** e desmarque **default** (padrão).

Selecione o “**Grupo de Destino**” criando anteriormente: **LabGroupRouting**

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Clique em **Criar load balancer**

Aguarde Provisionar

Interface gráfica do usuário, Texto, Aplicativo, chat ou mensagem de texto

Descrição gerada automaticamente

O balanceador de carga mostrará um estado de *provisioning (provisionamento)*. Não há necessidade de esperar até que ele esteja pronto. Siga para a próxima etapa.

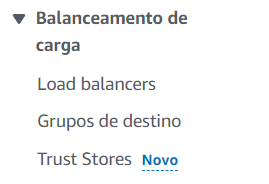
Agora foi provisionado e está “**Ativo**”

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

**3 – Associar Load Balancer ao Grupo de Destino**

No menu procure por **Grupos de destino**.



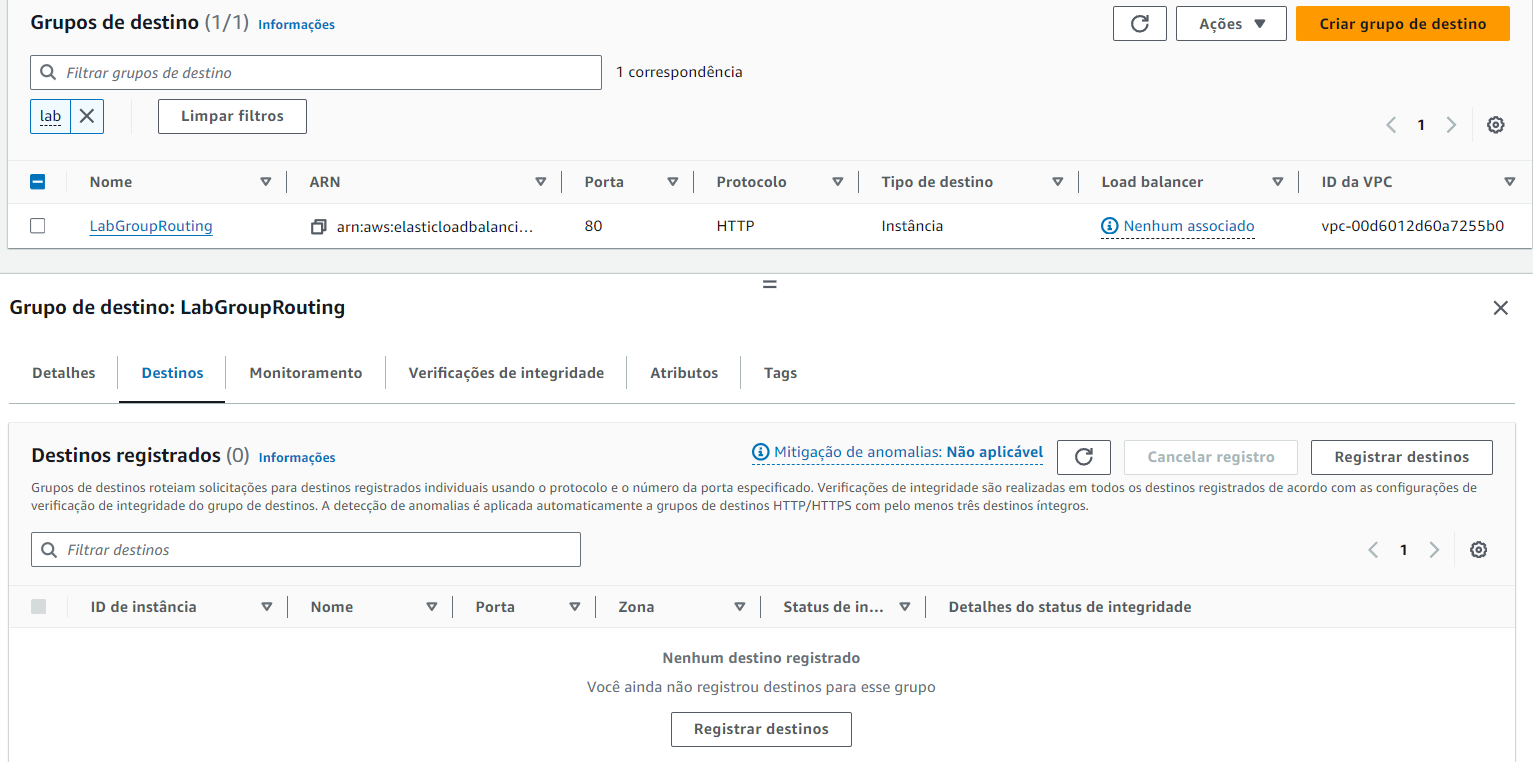
Selecione o grupo de destino **LabGroupRouting**.

Interface gráfica do usuário, Texto, Aplicativo

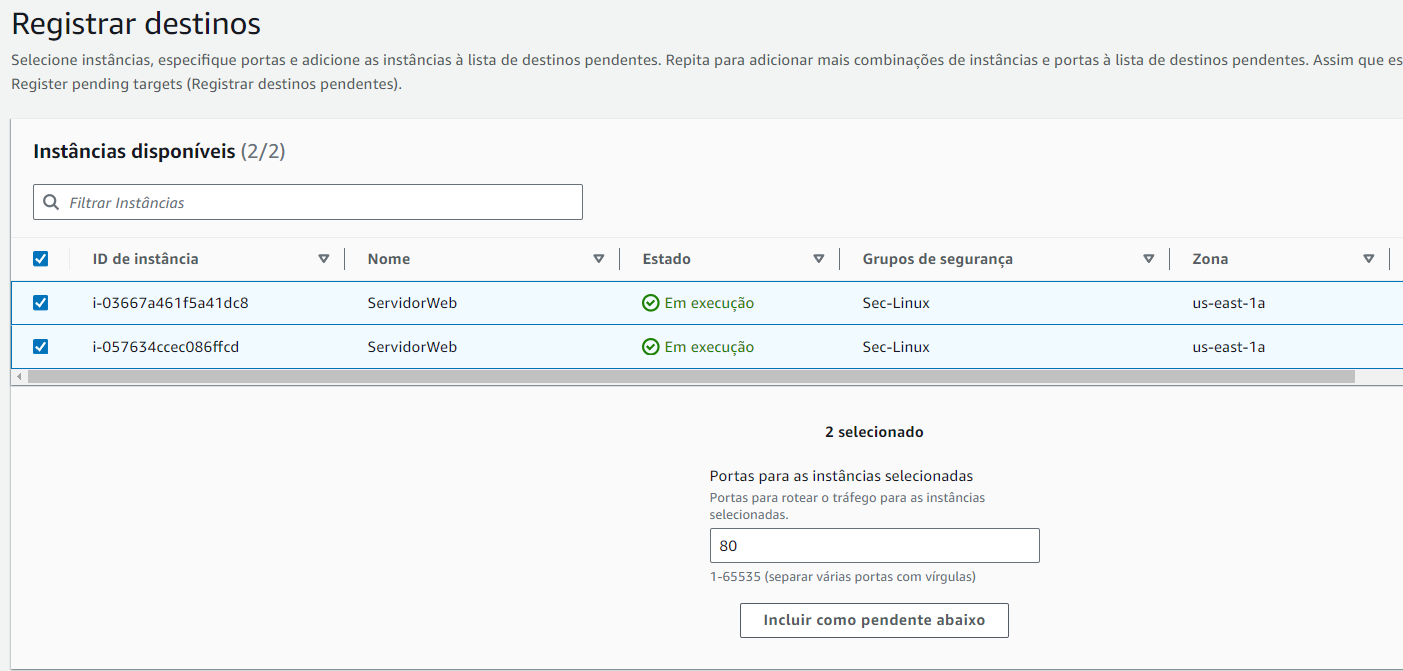
Descrição gerada automaticamente

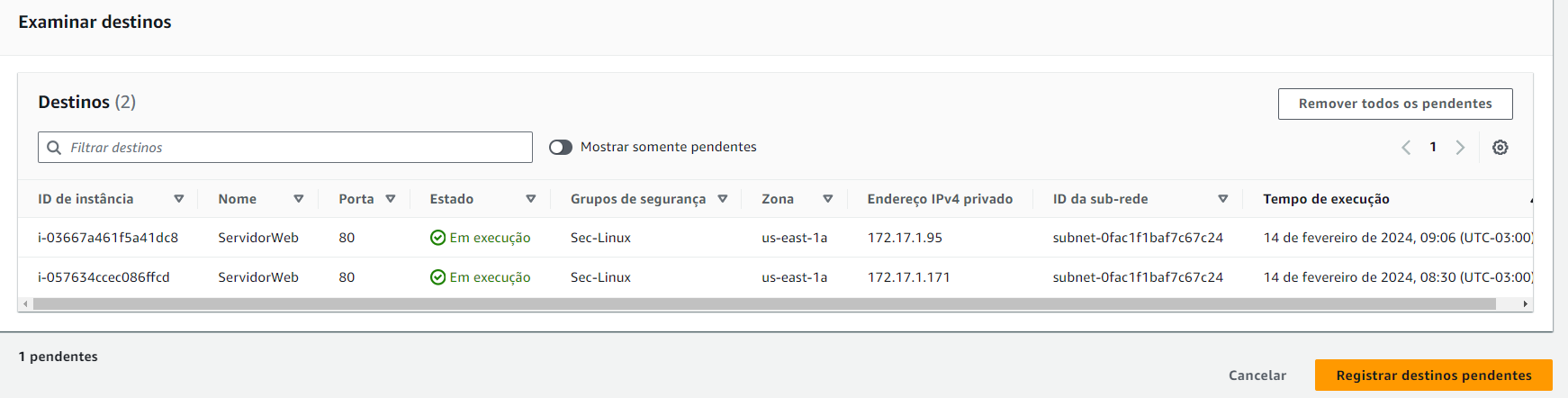
Veja que não existe nenhum Load Balancer associado.

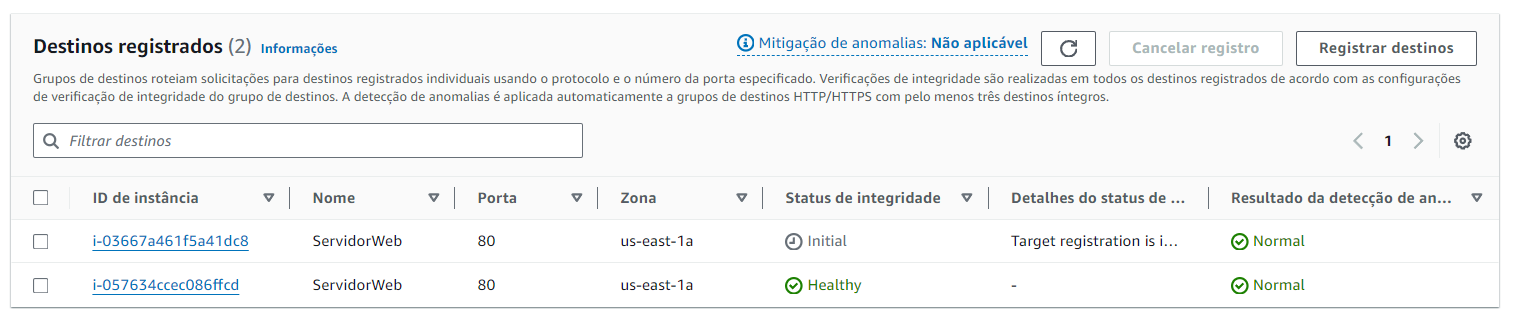
Selecione o grupo de destino **LabGroupRouting**.

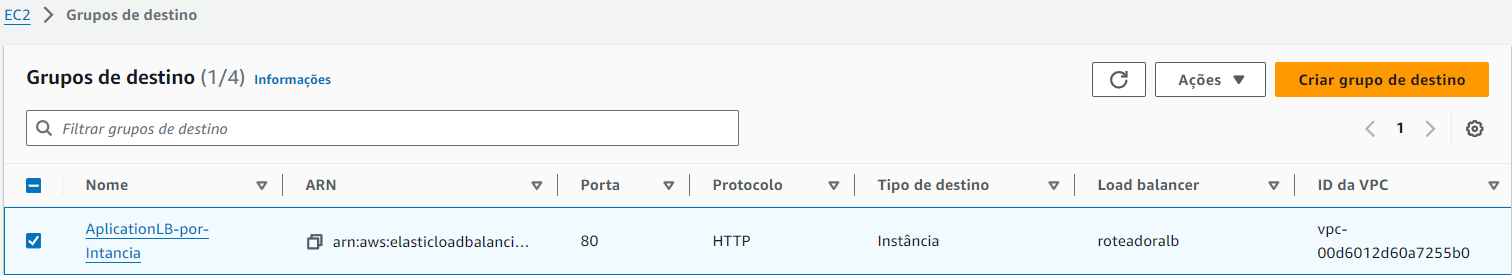


Selecione a aba Destinos e clique em Registrar destinos

Selecione as duas InstânciasClique em Incluir como pendentes a baixo

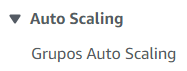
Clique em Registrar destinos pendentes

Aguarde sair de **Initial** para **Healthy** 

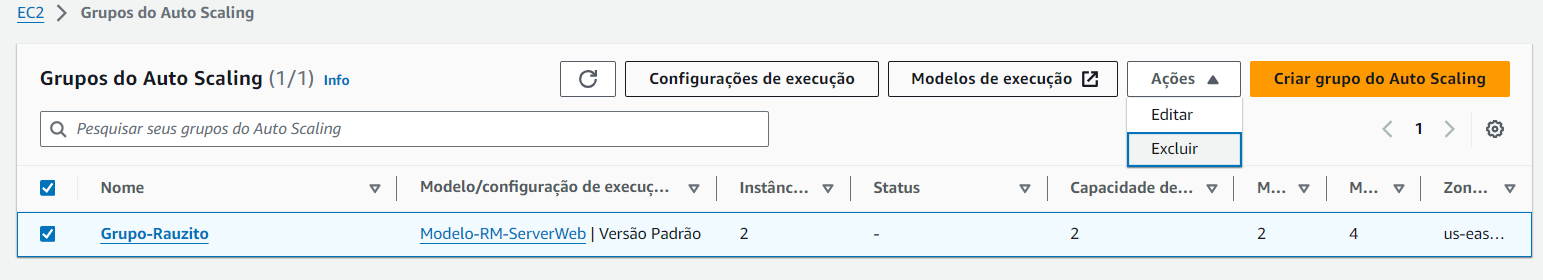
Veja que agora existe um **Load balancer** associado ao **Grupo de destino**

**4 – Associar Auto Scaling ao Grupo de Destino do Load Balancer**

No menu procure por **Grupos Auto Scaling**.



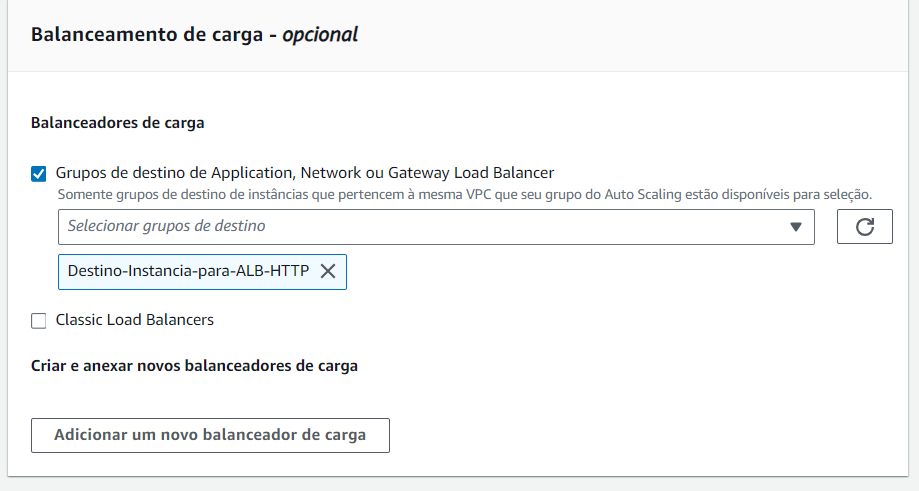
Selecione o **Grupo de Auto Scaling** e em **Ações** escolha: **Editar**

****

No quinto bloco **“Balanceamento de Carga – *opcional*”** habilite

**Grupos de destino de Application, Network ou Gateway Load Balance**

* depois escolha o Grupo de Destino ex.: **Destino-Instancia-para-ALB-HTTP**



**5 – Teste de conexão as instâncias pelo Load Balancer**

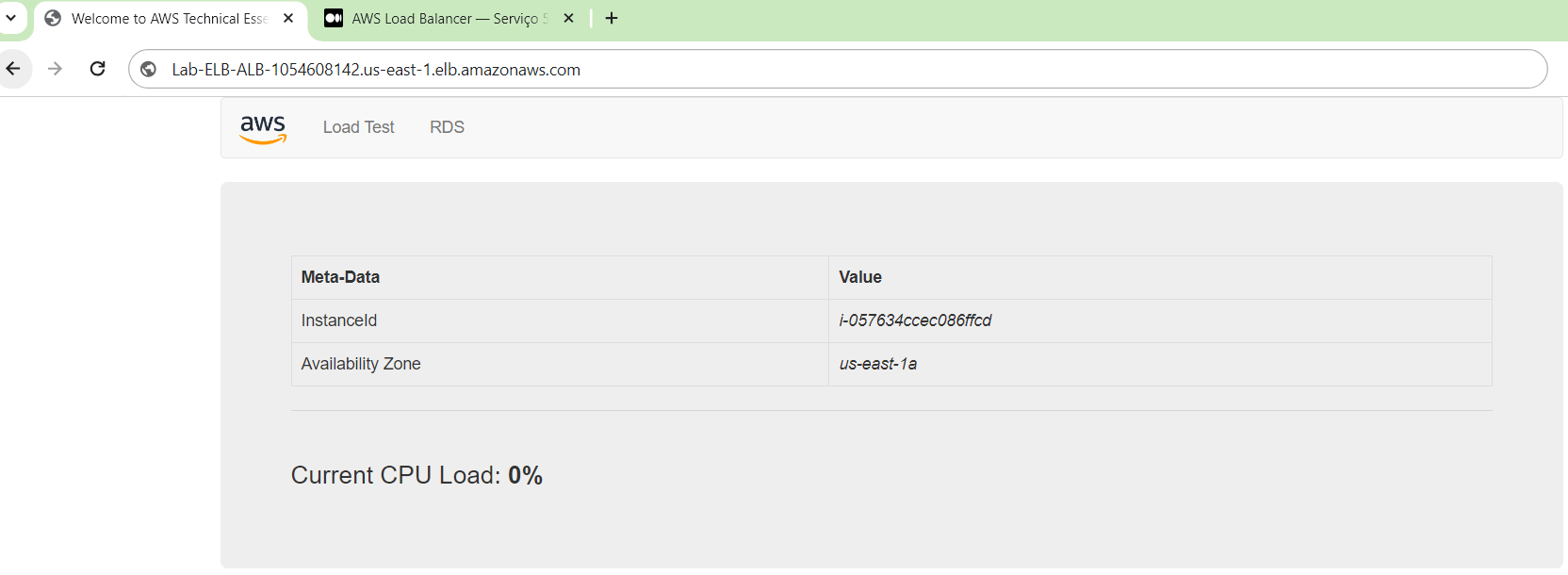
Selecionando o ELB “Lab-ELB-ALB” logo abaixo temos o endereço de acesso:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Lab-ELB-ALB-1054608142.us-east-1.elb.amazonaws.com

Acesse o navegador de sua preferência.

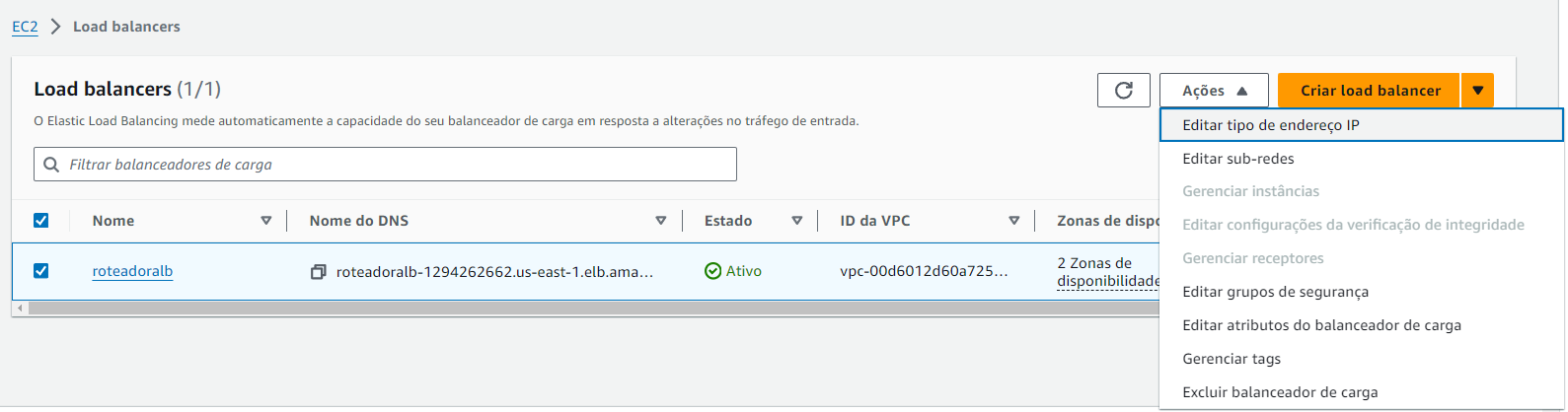


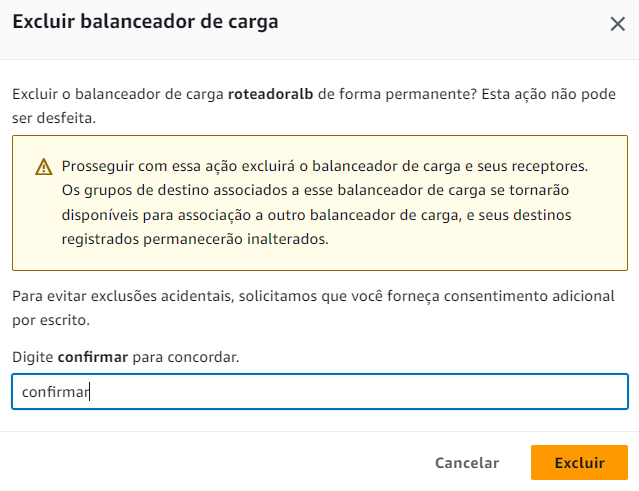
**6 – Excluir Load Balancer e Grupo de Destino**

No painel de navegação à esquerda, dentro da categoria **Balanceamento de carga** clique em **Load Balancers.**

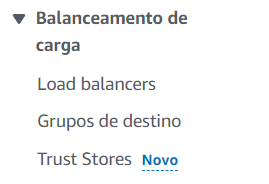


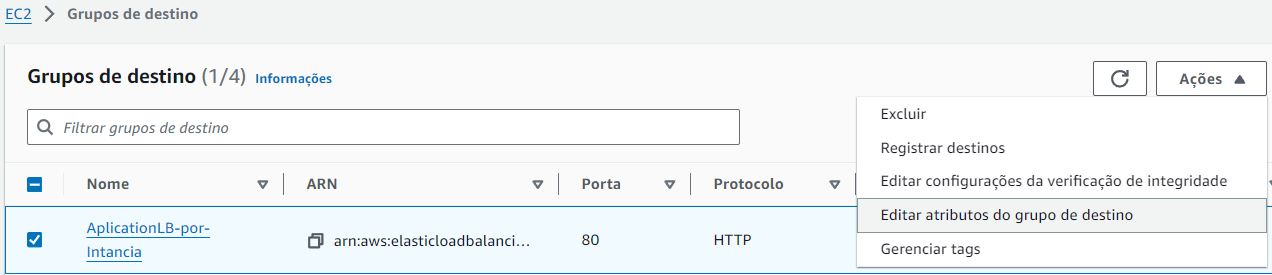
Selecione o balanceador de Carca ex.: “**roteadoralb**” e em **Ações** escolha a ultima opção **Excluir balanceador de carga**.

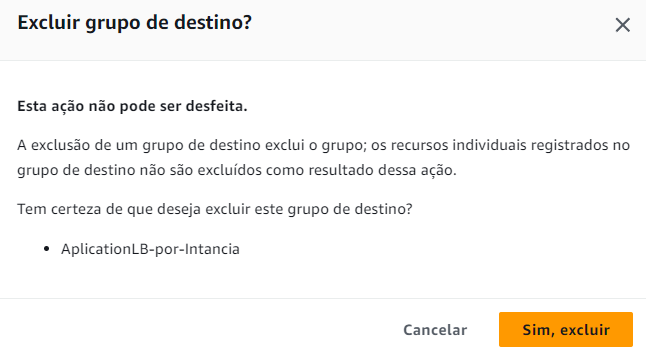


Escreva “**confirmar**” e clique em **Excluir**

Procure no menu por **Grupos de destino**.



Selecione o Grupo de destino ex.: “**AplicationLB**” e em **Ações** escolha a primeira opção **Excluir**



Clique em **Sim, excluir**