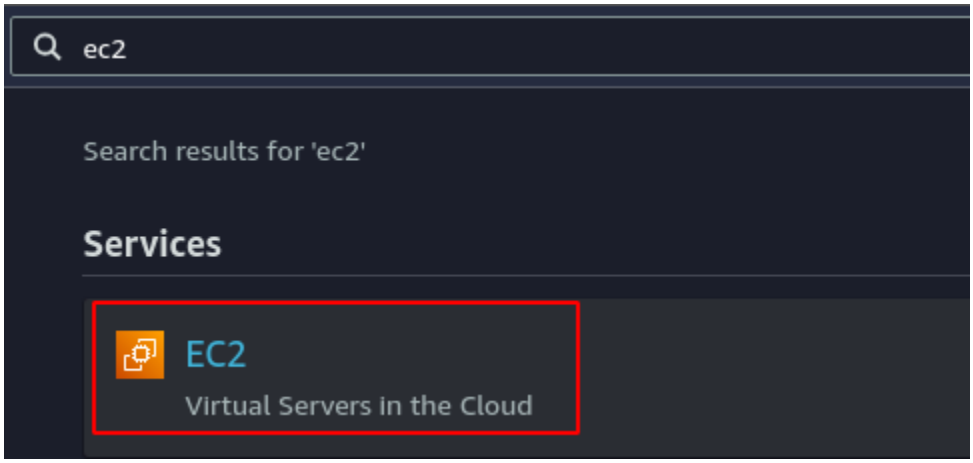
LOAD BALANCER IN AWS (APPLICATION LOAD BALANCER)

how an application load balancer can be created and configured using an AWS management console.

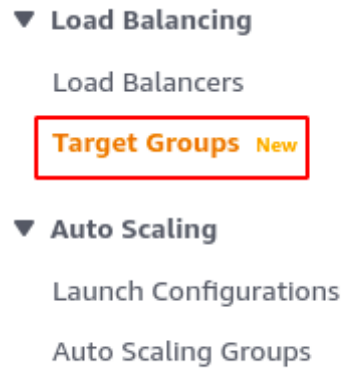
Creating target groups

Application load balancer receives traffic and forwards the traffic to the target groups. These target groups are the groups of the targets like EC2 instances in multiple availability zones.

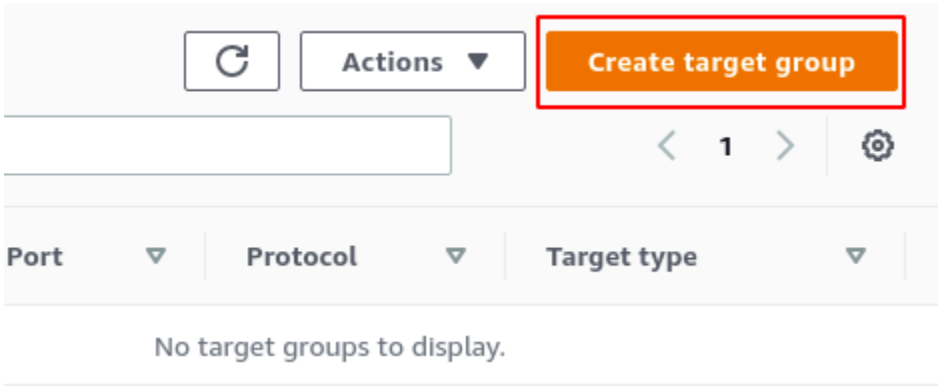
This section will create a target group and then register the EC2 instance to the target group. First, log into the AWS management console and go to the EC2 services.



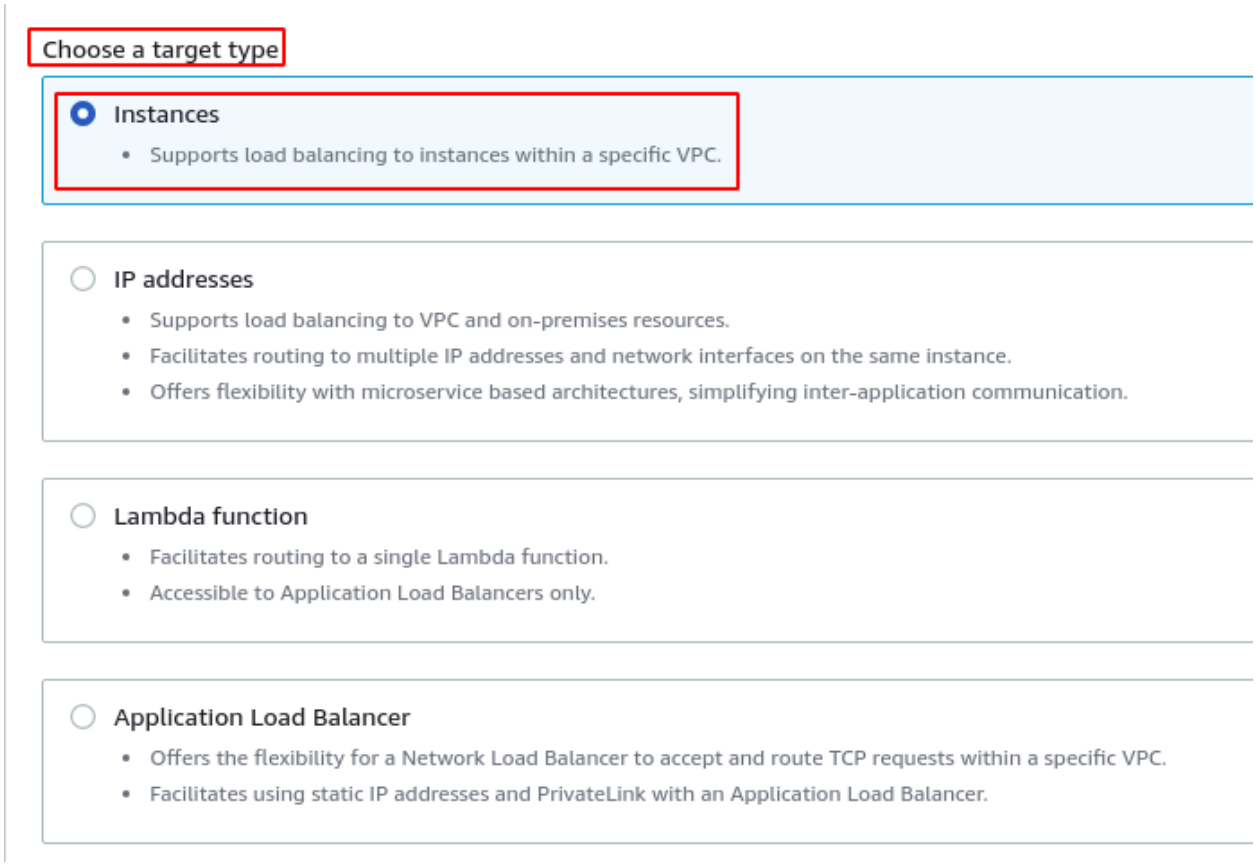
From the left side panel, go to the **Target Groups** section.



From the top right corner of the console, click on the **create target group button** to create a new target group.



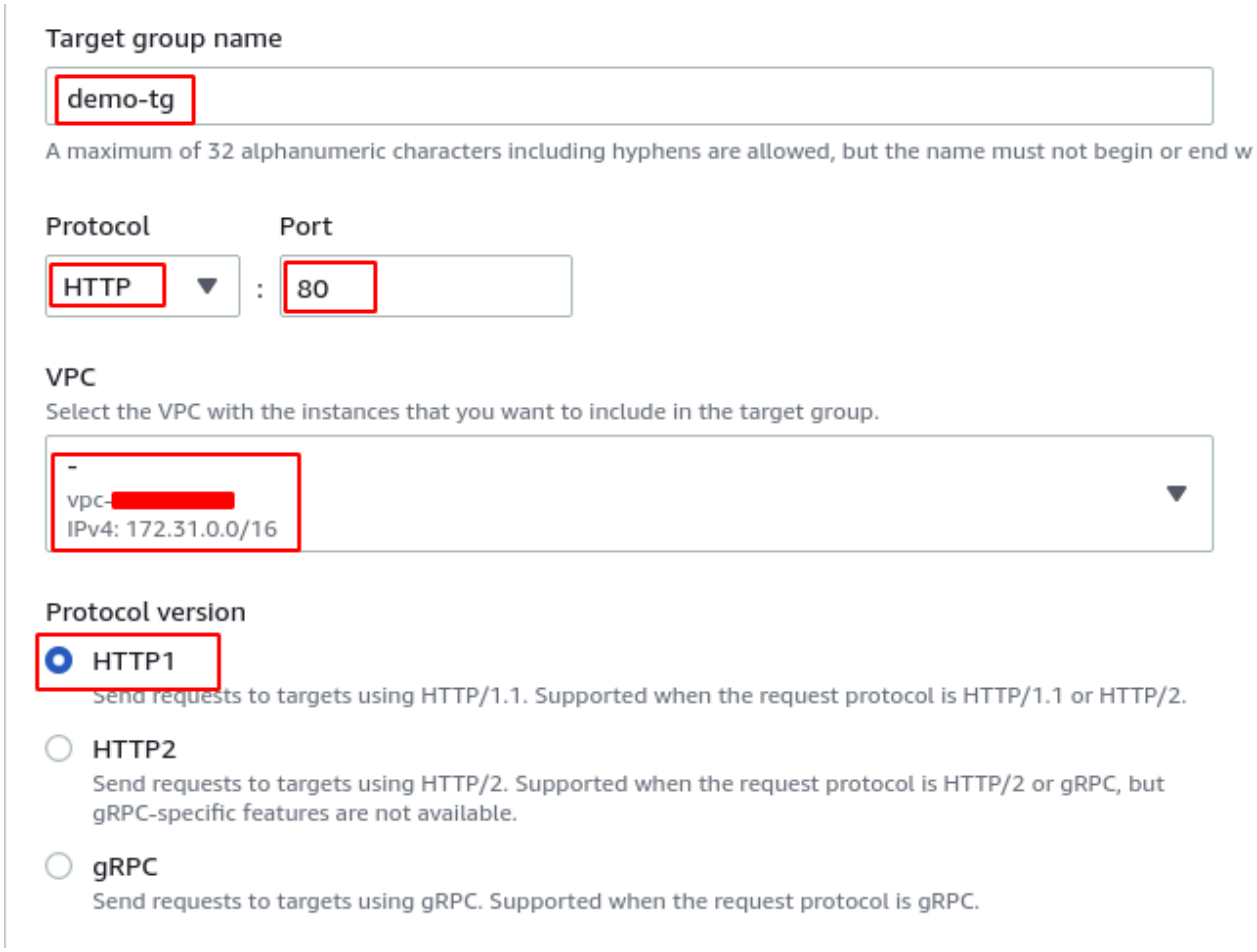
Now it will ask for the **target type** you want to register in this target group. For this demo, select the **Instances** as target types for this target group.



Now it will ask for the **target group name, port, protocol, HTTP Version, and VPC**. The target group name is a unique identifier that identifies the target group in a VPC.

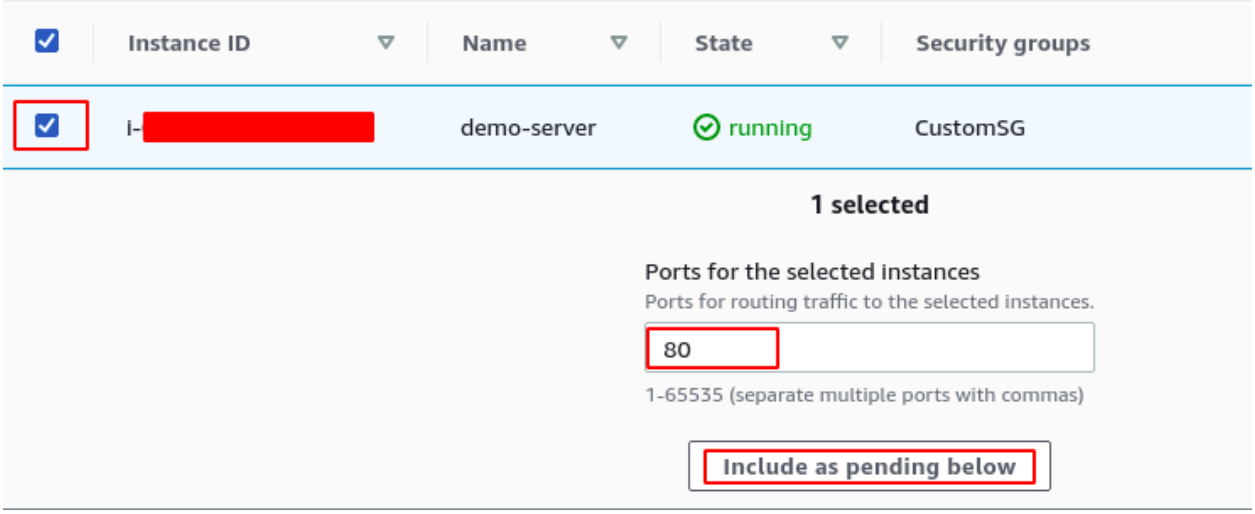
The port and protocol are the port number and protocol on which the target group will listen for incoming requests from the application load balancer. For the target groups that will receive traffic from ALB, the protocol must be HTTP or HTTPS.

The VPC is the virtual private cloud in which instances are running that will be registered with the target group.



Now keep the health check settings as default and click on the **Next** button to register the targets for the target group.

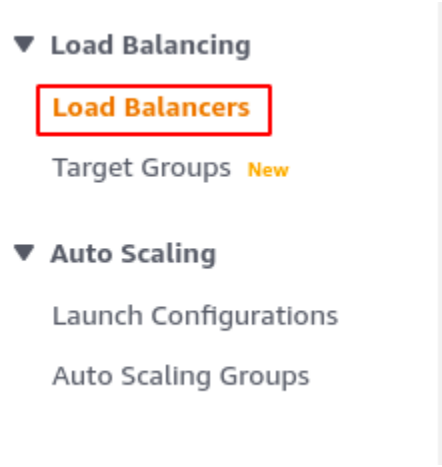
It will display all the instances in the VPC, and you can register any number of instances from here. For this demo, we will register only one EC2 instance in this target group. Select the instance you want to register and enter the port on which the server is running inside the instance. Click on the **include as pending below button** to register the target in the target group.



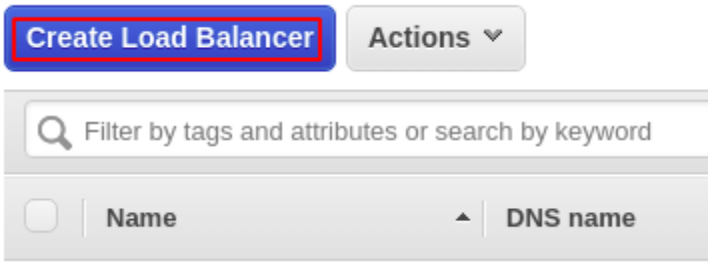
After registering the target, click on the **create target group button** to create the target group with EC2 instance registered.

Creating application load balancer

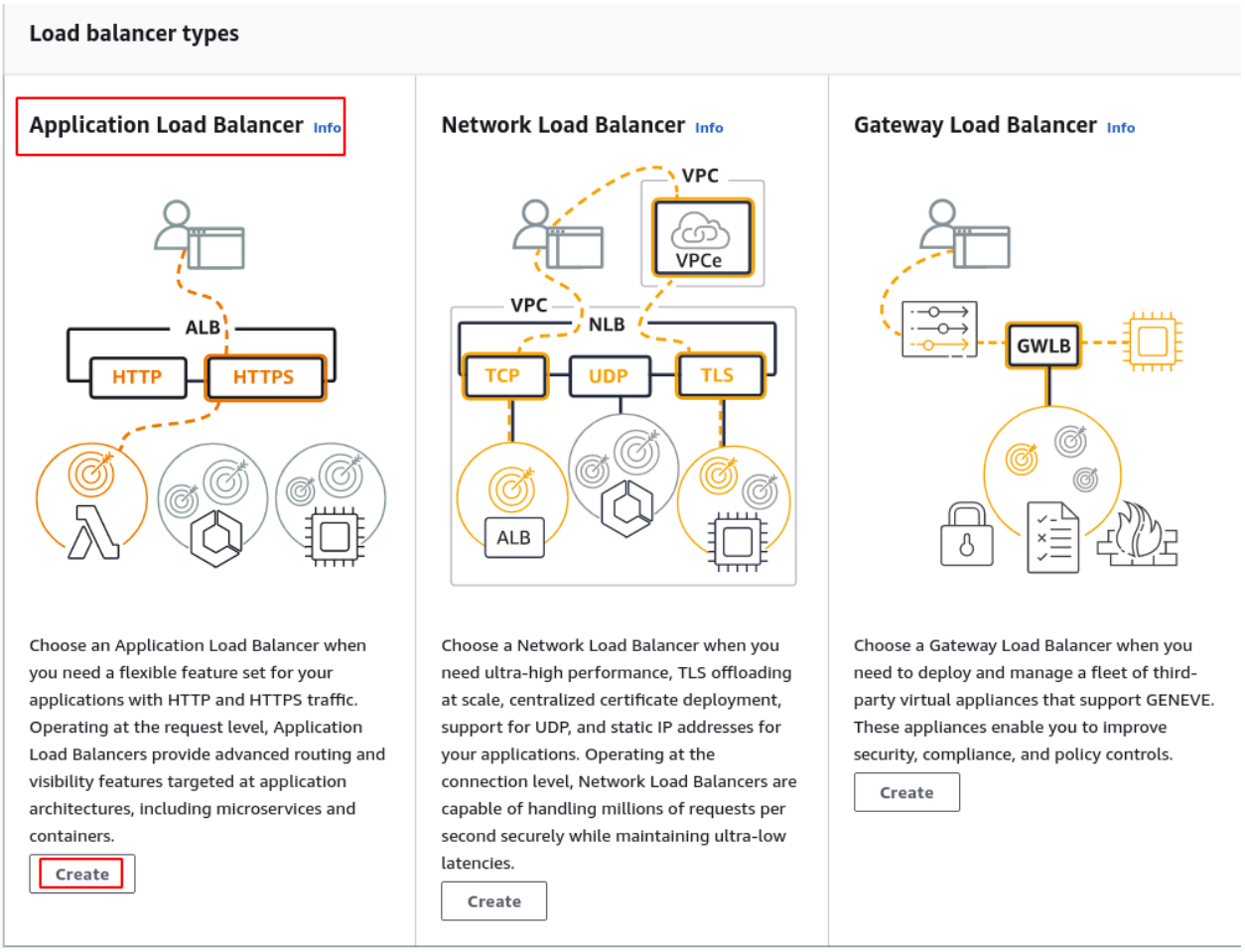
After creating the target group for the **load balancer**, now go to the Load balancers from the left side panel.



Click on the **create load balancer button**.



It will open a new web page to select the load balancer type to create. Select the **application load balancer** and click on the **create** button.

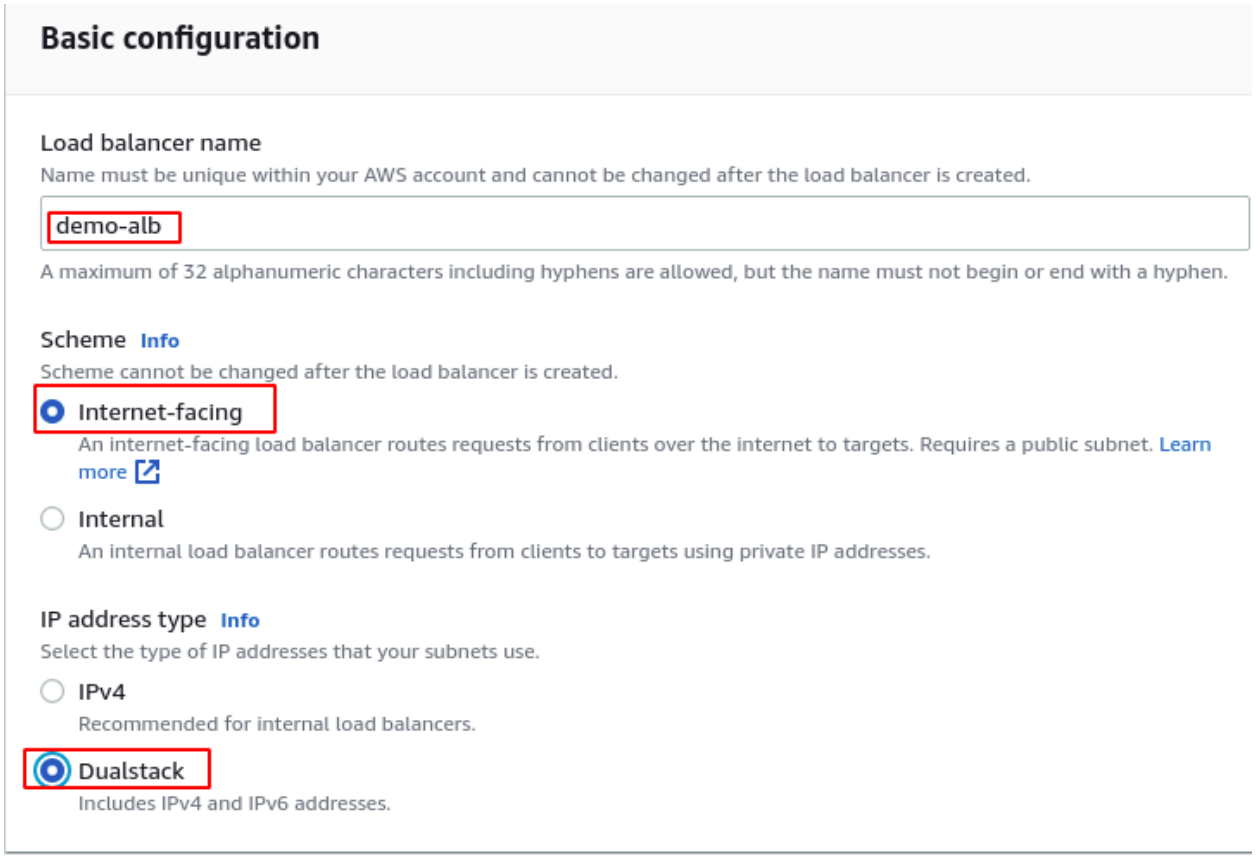


Now it will ask for the basic configuration of the application load balancer. Enter the **Load balancer name, scheme, and IP address type**. Load balancer name is a unique identifier for the application load balancer to be created.

Load balancer scheme defines whether load balancer will be internal or internet-facing. Internet-facing load balancers can accept connection requests from the public internet and route these requests to the target groups internally. On the other hand, internal load balancers do not have a publicly resolvable DNS name. They can only be accessed within the VPC and route the requests to the target groups internally.

The IP address type defines whether the end-users can send requests using **IPv4 or IPv6 addresses**. Setting the IP address type to IPv4 will allow the end-users to send the requests from IPv4 only. For internet-facing load balancers, it is recommended to use the **dualstack** IP address type.

For this demo, we will create an internet-facing application load balancer with **dualstack** IP address type.

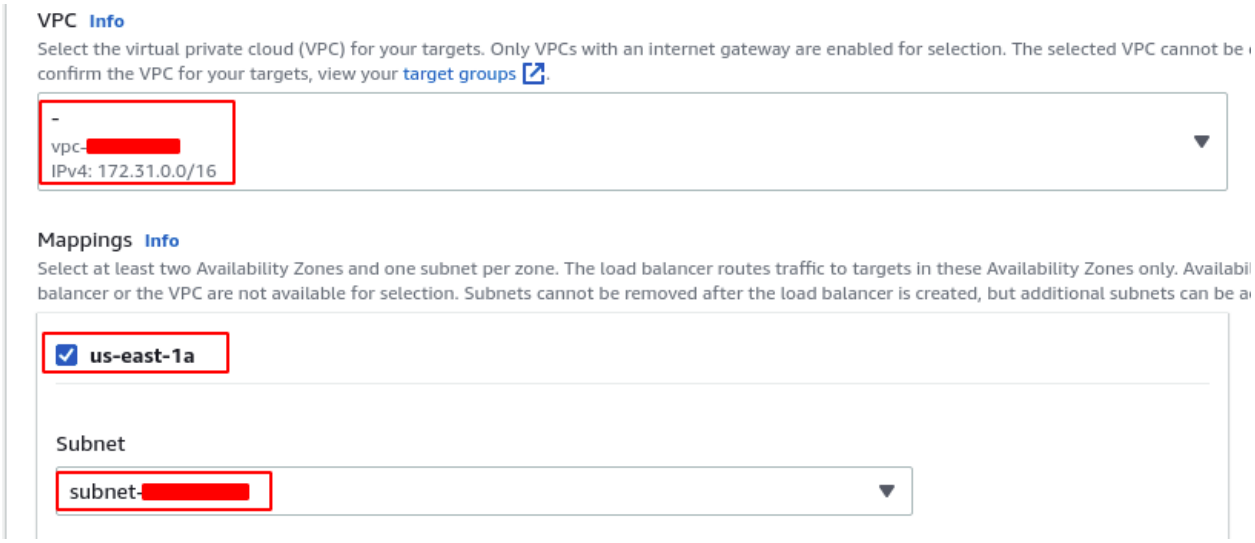


Now for networking, select the VPC, availability zones, and subnets from

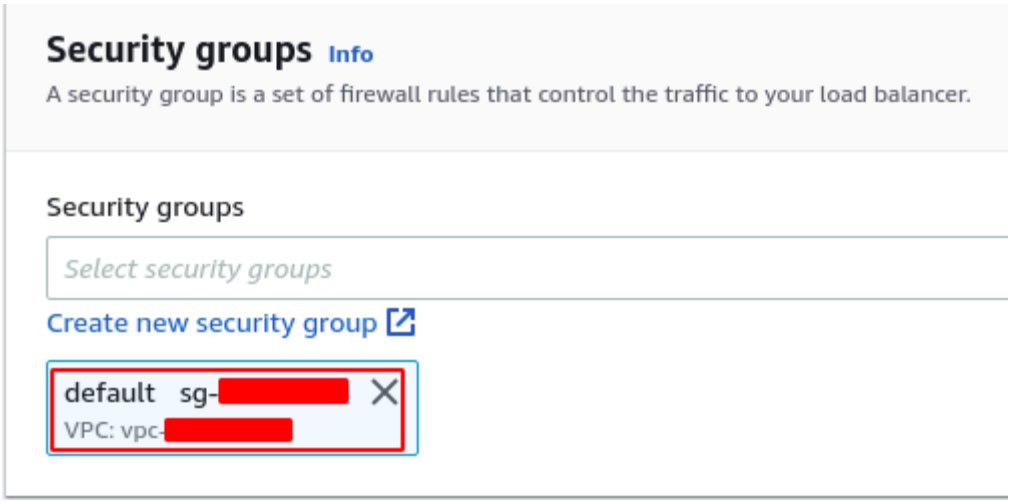
availability zones. The VPC must be the same as selected while creating the target group.

For availability zones, AWS recommends choosing at least two availability zones with at least one public subnet in it to configure for the internet-facing application load balancer.

For this demo, select 3 availability zones of default VPC. Availability zones of default VPC have public subnets to configure it with the internet-facing application load balancer.



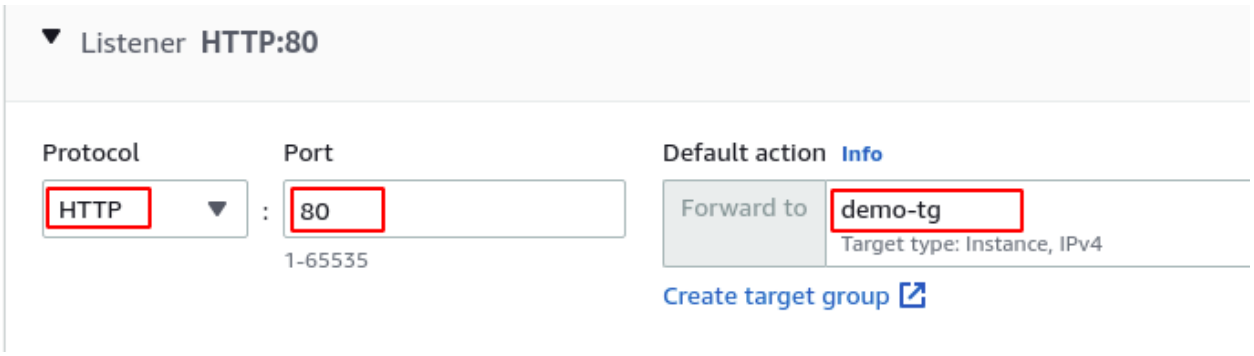
For the security group, select a security group from the VPC that will control inbound and outbound traffic from the application load balancer.



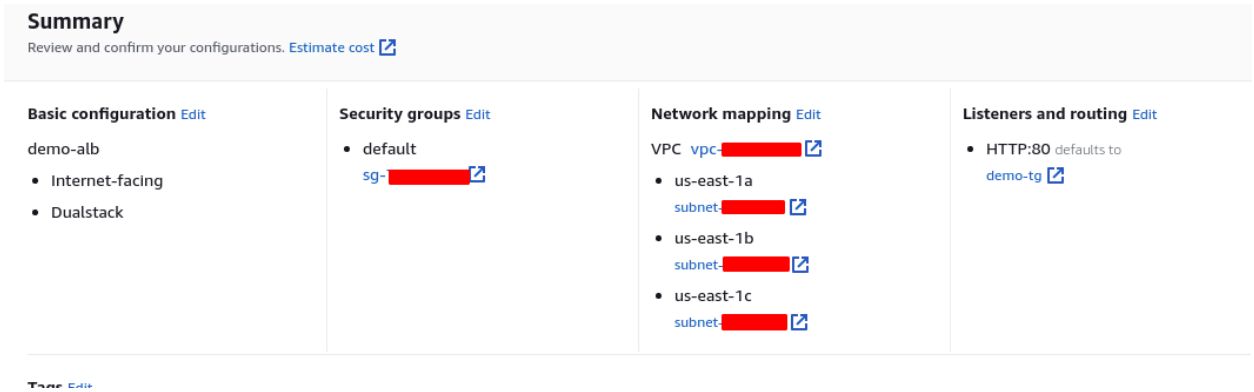
After configuring VPC and security group, now configure the default listener for the application load balancer.

For this demo, we will configure port 80 of the application load balancer to receive traffic from the end-users and forward the traffic to the **demo-tg target group**.

The **demo-tg target group** is configured to equally pass the traffic to port 80 of the EC2 instances inside it.



After configuring the load balancer, now have a look at the configuration summary and click on the create load balancer to **create the load balancer**.



After creating the load balancer, make sure of the following points.

- Security group attached to the load balancer have an inbound rule to allow traffic on the load balancer
- Security group attached to the EC2 instances have an inbound rule to allow traffic from the load balancer
- The port on which the application is running in the EC2 instance is configured properly in the target group
- EC2 instance and application load balancer must be in the same VPC
- The availability zone and subnet in which the EC2 instance is running must be mapped while creating the load balancer