

AWS APPLICATION LOAD BALANCER AND AUTO SCALING GROUP OVERVIEW

Core Function and Component Roles

This system is built to make our application **highly reliable** and able to **handle any amount of user traffic** automatically.

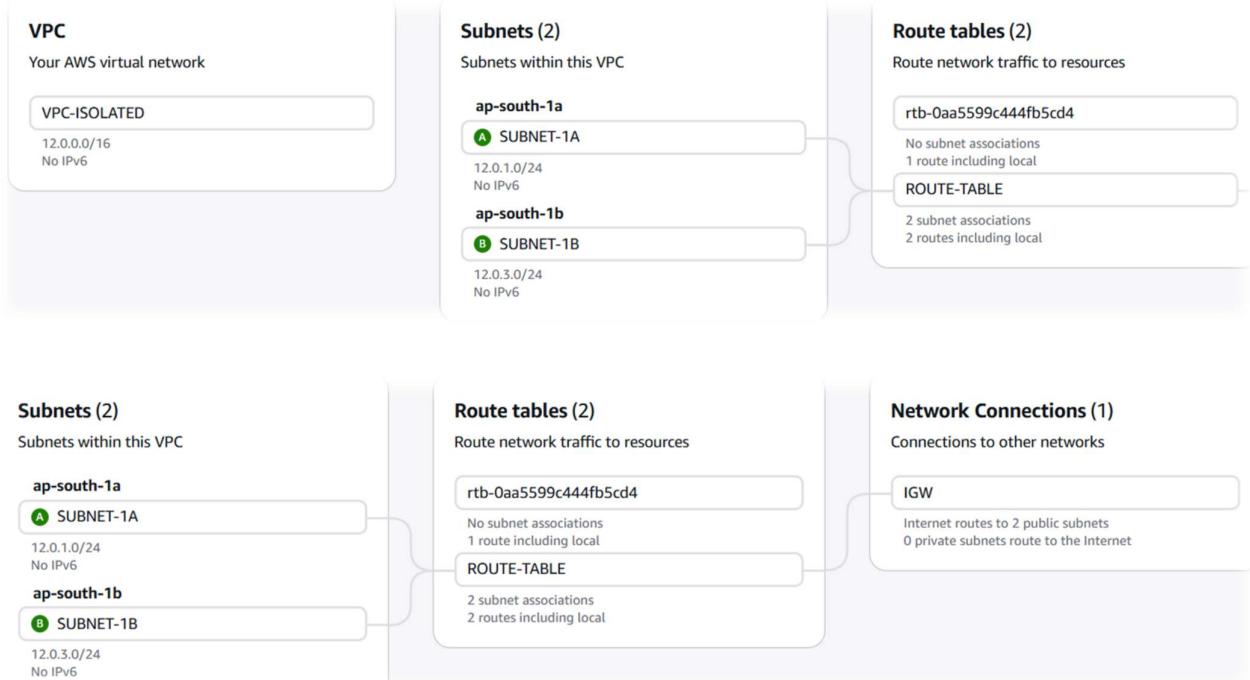
- The **Application Load Balancer (ALB)** is the traffic cop that receives all requests and sends them only to the healthy servers.
- The **Auto Scaling Group (ASG)** is the smart manager that automatically creates more servers when traffic is high and removes them when traffic is low.

CONFIGURATION STEPS IN AWS

The process is broken down into three main phases: Setting up a **VPC setup** and then setting up **Load Balancer** and the **Auto Scaler** to scale our servers horizontally.

1. Network Setup (VPC and Routing)

- **VPC Creation:** Create a VPC (VPC-ISOLATED) with a desired IP range (e.g., 12.0.0.0/16). Ensure Default Tenancy is selected.
- **Internet Gateway (IGW):** Create an IGW (e.g., IGW) and Attach it to your new VPC.
- **Subnets:** Create multiple subnets (e.g., 12.0.1.0/24) across different Availability Zones (AZs) in ap-south-1 for redundancy.
- **Route Table:**
 - Create a Route Table and associate all created subnets.
 - Ensure you add a route for 0.0.0.0/0 (all internet traffic) pointing to the IGW.



2. Load Balancer Setup (ALB)

- Target Group Creation:**
 - Create a Target Group (e.g., TARGET-GROUP-APACHE) for instances using the HTTP protocol.
 - Ensure you select your VPC-ISOLATED.
- ALB Security Group:**
 - Create a new Security Group (e.g., ALB-SECURITY-GRP-HTTP).
 - Ensure it allows Inbound HTTP (Port 80) traffic from Anywhere (0.0.0.0/0).
- ALB Creation:**
 - Create an Internet-facing Application Load Balancer (ALB).
 - Ensure you select your VPC and the public subnets across multiple AZs.
 - Configure the Listener (Port 80) to forward traffic to the Target Group you created.
 - Wait until the ALB state is Provisioned.



3. Auto Scaling Setup (ASG)

- **ASG Security Group:**
 - Create a Security Group for the EC2 instances.
 - Ensure it allows Inbound HTTP (Port 80) and SSH (Port 22) traffic.
- **Launch Template Creation:**
 - Create a Launch Template (e.g., MY-LAUNCH-TEMPLATE) which acts as the instance blueprint.
 - Select the required AMI, Instance Type (e.g., T3.Micro), and your Key Pair.
 - Ensure you select the ASG Security Group.
 - In Network Configuration, Enable "Auto-assign Public IP."
 - In User Data, add the following script to install and configure Apache

Launch template version details			
Version	Description	Date created	Created by
1 (Default)	-	2025-11-19T15:32:54.000Z	arn:aws:iam::654290653206:root
Instance details	Storage	Resource tags	Network interfaces
AMI ID ami-02b8269d5e85954ef	Instance type t3.micro	Availability Zone -	Availability Zone Id -
Key pair name NEW-KEY-PAIR	Security groups -	Security group IDs sg-0ff1cb4dacff7d8b	
Advanced details			

```
#!/bin/bash
yes | sudo apt update
yes | sudo apt install apache2
# Create HTML file with server details
cat <<EOF | sudo tee /var/www/html/index.html
<h1>Server Details</h1>
<p><strong>Hostname:</strong> $(hostname)</p>
<p><strong>IP Address:</strong> $(hostname -I | awk '{print $1}')</p>
EOF
sudo systemctl restart apache2
```

- **ASG Creation:**

- Create the Auto Scaling Group (e.g., AUTO-SCALING-GROUP) and select the Launch Template.
- Ensure you select the private subnets across the Availability Zones.
- Under Load Balancing, Attach to the existing Load Balancer and select your Target Group.
- Turn on Elastic Load Balancing Health Checks.
- Define the Group Size (e.g., Min: 2, Max: 5, Desired: 2).
- Define a Scaling Policy (e.g., Target Tracking to keep Average CPU Utilization at 60%).
- The ASG will now automatically launch the initial instances.

AUTO-SCALING-GROUP Capacity overview			
Edit			
Desired capacity 1	Scaling limits (Min - Max) 1 - 5	Desired capacity type Units (number of instances)	Status -

Verification

- **Check Load Balancing:**

Copy the ALB's DNS Name and paste it into a browser. Verify that upon refreshing, the displayed Hostname/IP Address changes, confirming the ALB routes traffic properly.

- **Check Auto Scaling:**

Terminate one of the running EC2 instances manually. Verify that the ASG automatically launches a replacement instance to maintain the desired capacity.