Scalable Web Application with Load Balancing

Services used:
EC2, Application Load Balancer (ELB), Auto Scaling, RDS, S3, Route 53
Skills covered:
VPC, Subnets, Security Groups, EC2 Launch Template, User Data, Auto Scaling

---

PROJECT GOAL

Deploy a **highly available and scalable web application** that:

- Automatically **scales EC2 instances** based on traffic
- Uses **Load Balancer** to distribute traffic
- Stores data in **RDS**
- Serves static content from **S3**
- Uses **Route 53** for custom domain routing

Project Overview: Scalable Web Application with Load Balancing

This project focuses on deploying a **highly available and scalable web application** using AWS cloud services. The application is designed to automatically handle varying user traffic by distributing requests across multiple servers and scaling resources based on demand.

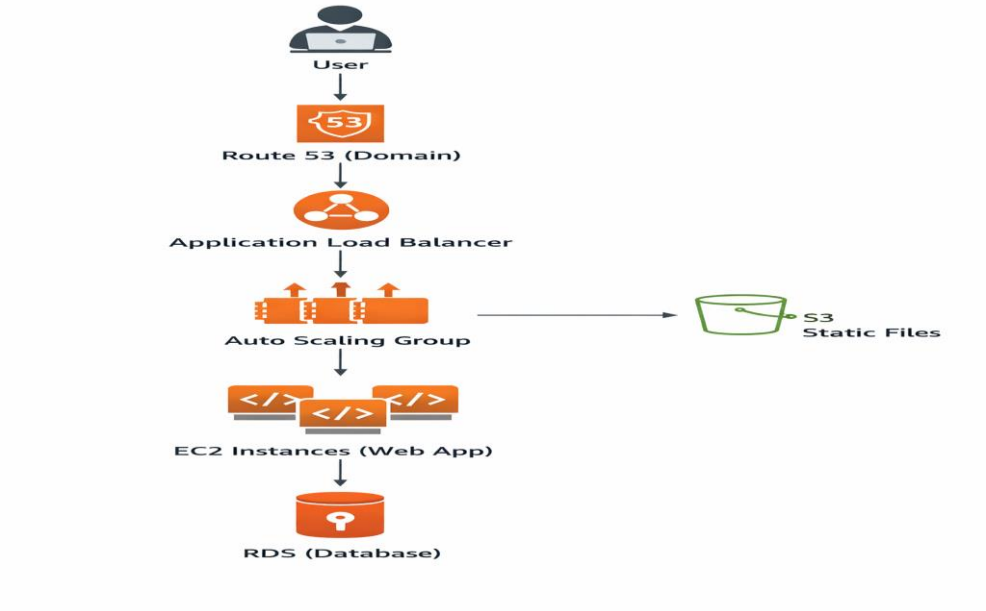
The solution uses **Amazon Route 53** to route user requests to an **Application Load Balancer**, which evenly distributes traffic to EC2 instances managed by an **Auto Scaling Group**. The EC2 instances host the web application and automatically launch or terminate based on CPU utilization. Application data is securely stored in an **Amazon RDS** database, while static content such as images, CSS, and JavaScript files are served from **Amazon S3**.

The entire setup is deployed inside a custom **VPC** with properly configured subnets, route tables, and security groups to ensure network isolation and security. This architecture provides fault tolerance, improved performance, cost efficiency, and scalability, making it suitable for real-world production workloads and enterprise-level applications.

ARCHITECTURE OVERVIEW

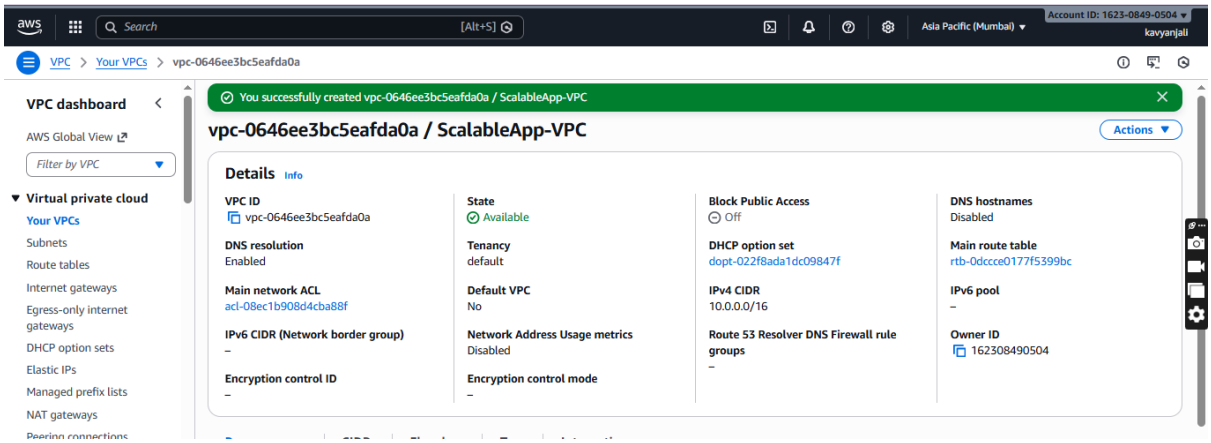
User
↓
Route 53 (Domain)
↓
Application Load Balancer
↓
Auto Scaling Group
↓
EC2 Instances (Web App)
↓
RDS (Database)

Static files → S3



STEP 1: Create VPC

1. Open **AWS Console**
2. Go to **VPC → Your VPCs**
3. Click **Create VPC**
4. Choose:
    - Name: ScalableApp-VPC
    - IPv4 CIDR: 10.0.0.0/16
5. Click **Create VPC**



STEP 2: Create Subnets

Create 2 Public Subnets (for ALB + EC2)

Go to **VPC → Subnets → Create subnet**

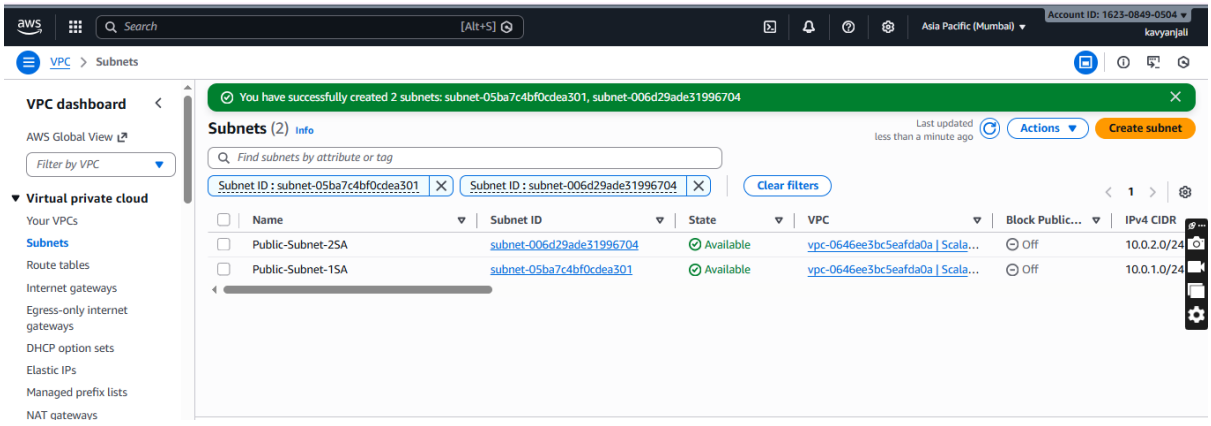
Subnet 1

- Name: Public-Subnet-1
- AZ: ap-south-1a
- CIDR: 10.0.1.0/24

Subnet 2

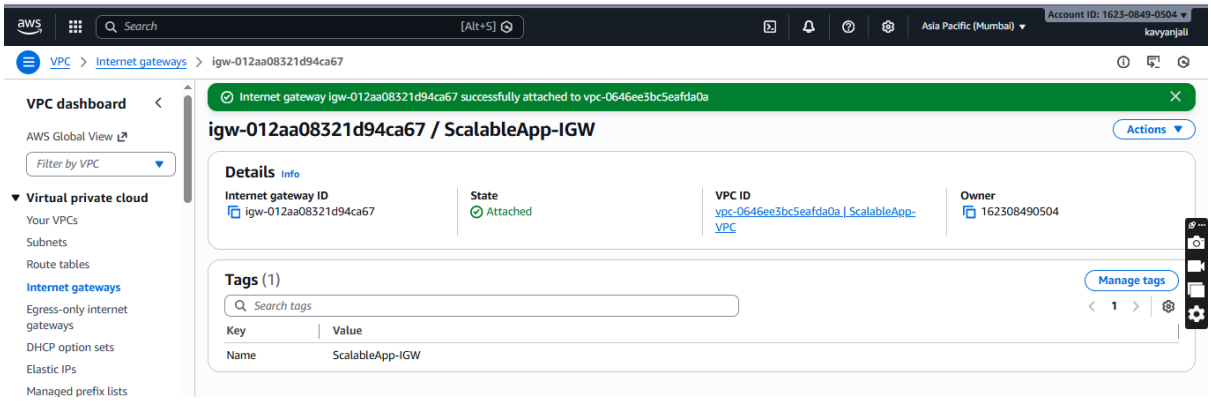
- Name: Public-Subnet-2
- AZ: ap-south-1b
- CIDR: 10.0.2.0/24

Click **Create subnet**



STEP 3: Internet Gateway

1. Go to **VPC → Internet Gateways**
2. Click **Create internet gateway**
    - Name: ScalableApp-IGW
3. Attach it to ScalableApp-VPC

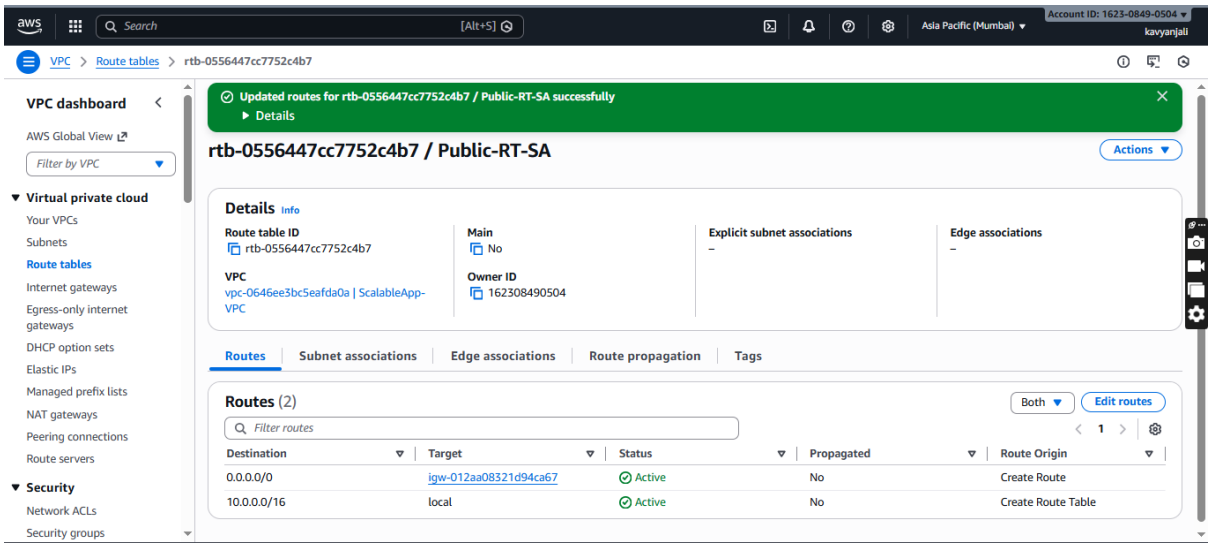


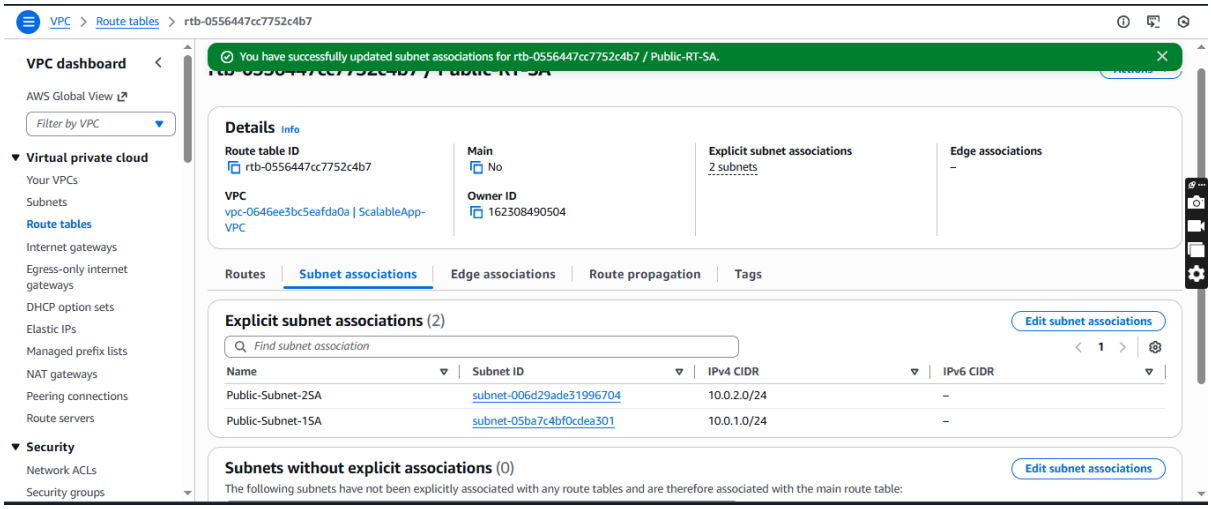
STEP 4: Route Table

1. Go to **VPC → Route Tables**
2. Create route table:

- Name: Public-RT
- VPC: ScalableApp-VPC

3. Edit routes:
   - Destination: 0.0.0.0/0
   - Target: Internet Gateway
4. Associate both public subnets

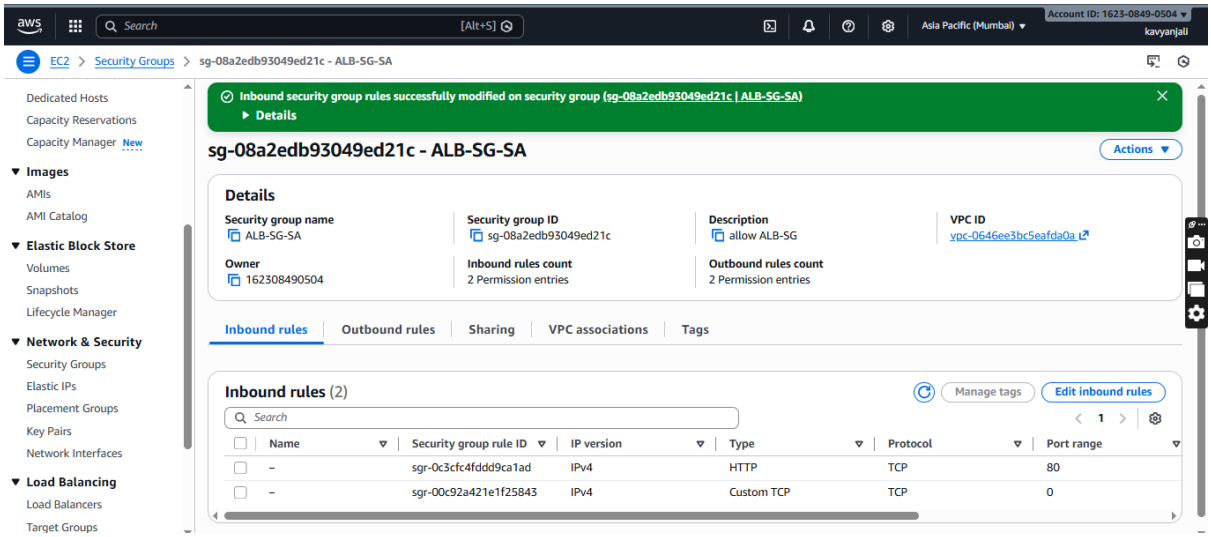




STEP 5: Security Groups

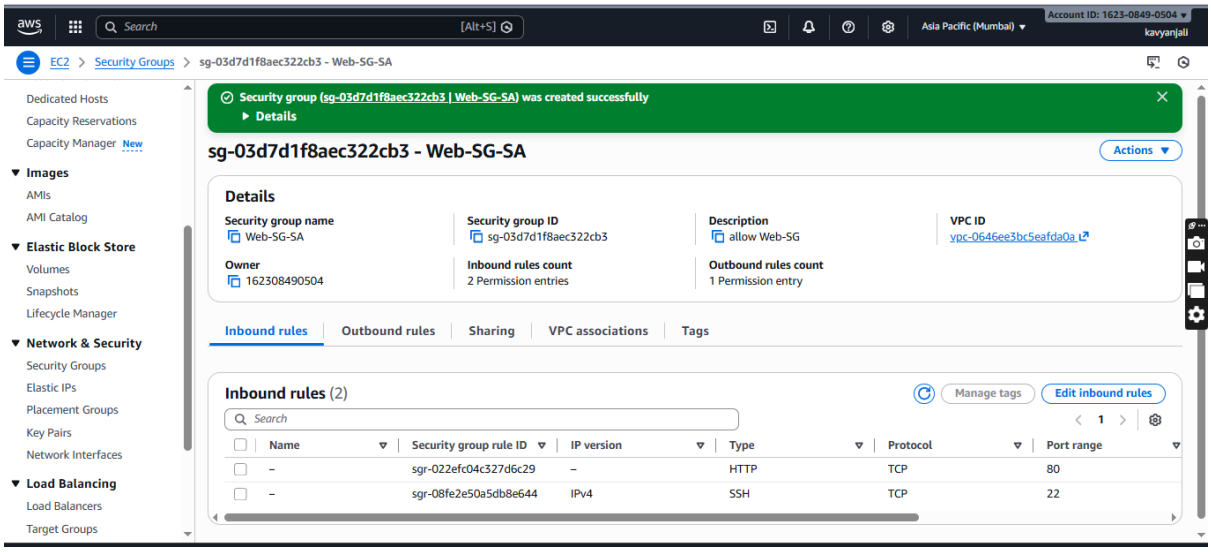
1.ALB Security Group

- Name: ALB-SG
- Inbound:
  - HTTP (80) → 0.0.0.0/0



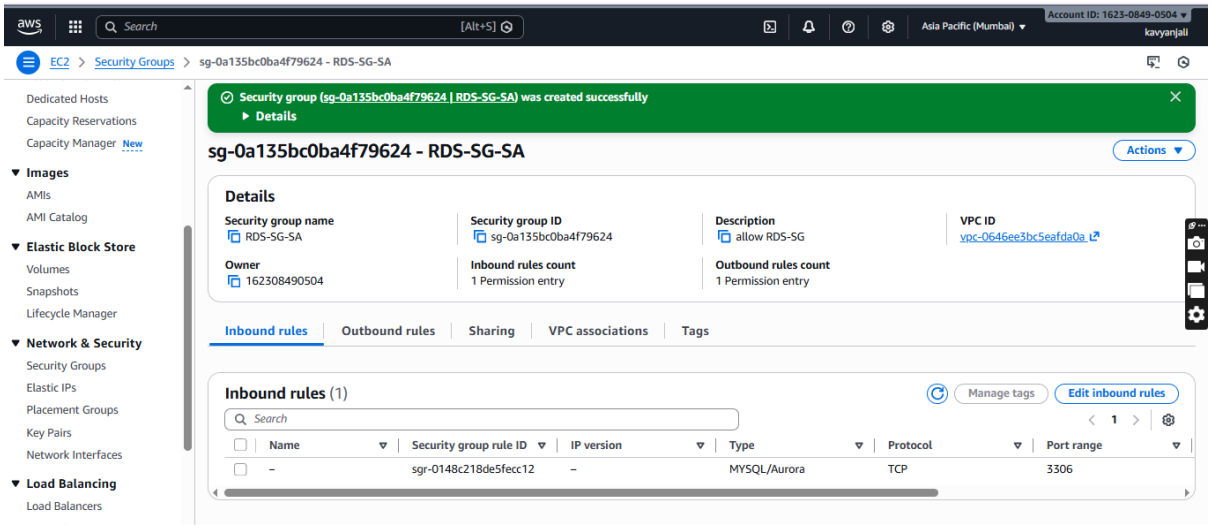
2. EC2 Security Group

- Name: Web-SG
- Inbound:
  - HTTP (80) → ALB-SG
  - SSH (22) → Your IP



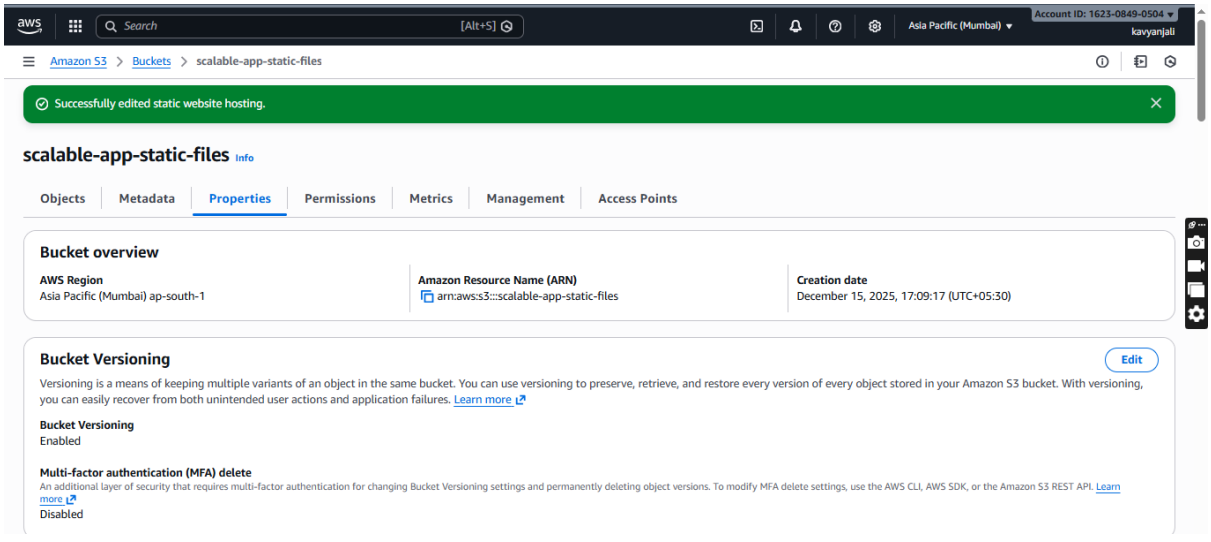
3. RDS Security Group

- Name: RDS-SG
- Inbound:
  - MySQL (3306) → Web-SG



STEP 6: Create S3 Bucket (Static Files)

1. Go to **S3 → Create bucket**
2. Name: scalable-app-static-files
3. Disable **Block Public Access**
4. Enable **Static Website Hosting**
5. Upload:
    - CSS
    - Images
    - JS files

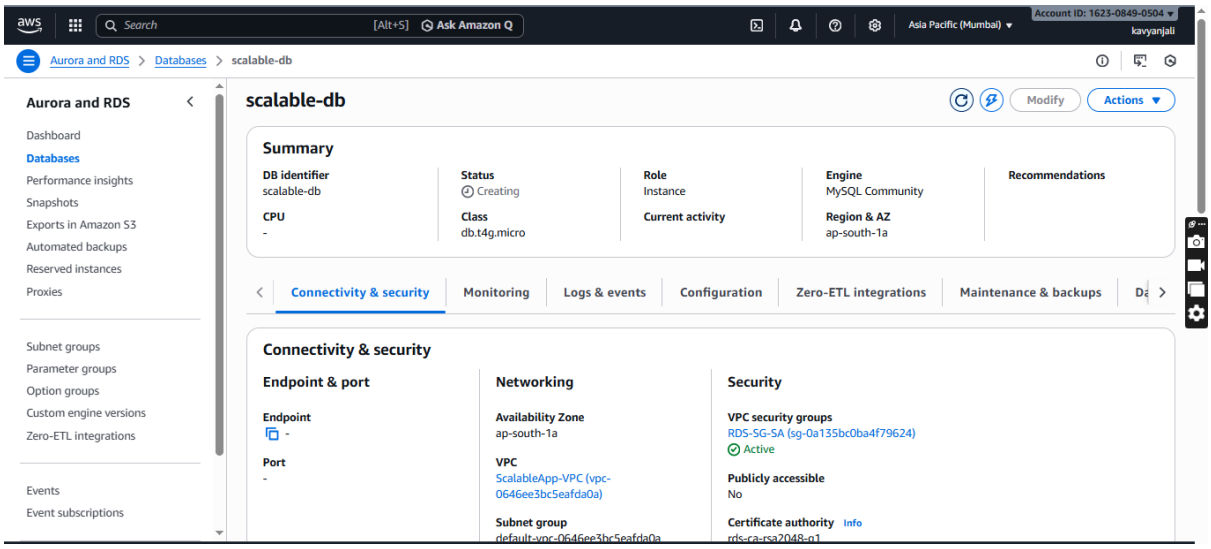


STEP 7: Create RDS Database

1. Go to **RDS → Create database**
2. Choose:

- Engine: **MySQL**
- Template: **Free tier**

3. Settings:
   - DB name: scalable-db
   - Username: admin
   - Password: StrongPassword123
4. Connectivity:
   - VPC: ScalableApp-VPC
   - Security Group: RDS-SG
   - Public access: No
5. Click **Create database**



STEP 8: Create EC2 Launch Template

1. Go to **EC2 → Launch Templates**
2. Click **Create launch template**
3. Name: WebApp-LT

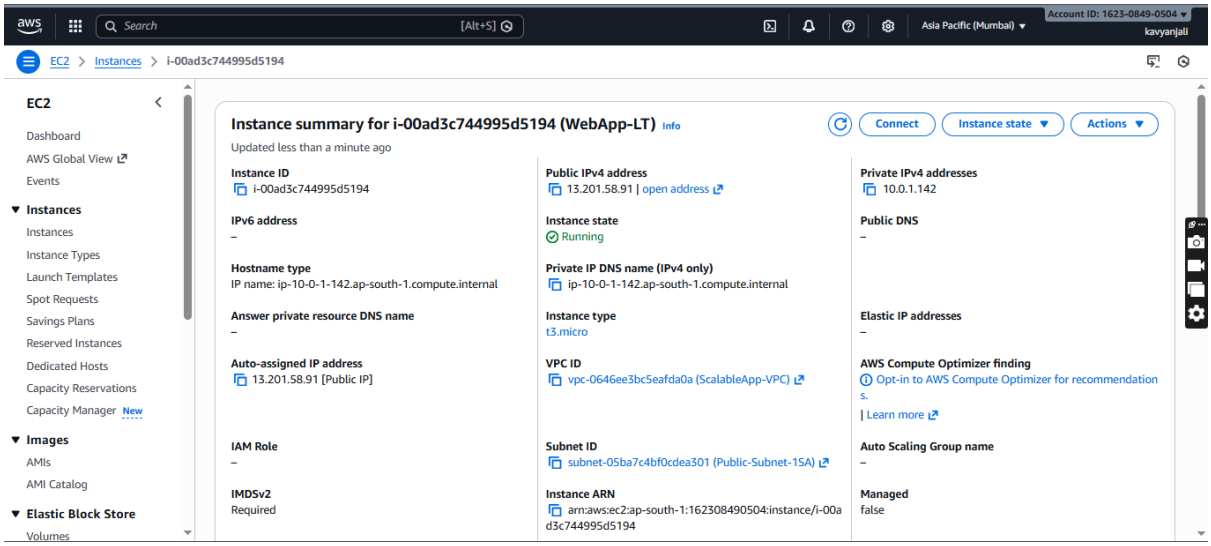
Configuration:

- AMI: Amazon Linux 2
- Instance type: t2.micro
- Key pair: Select your key
- Security Group: Web-SG

User Data Script:

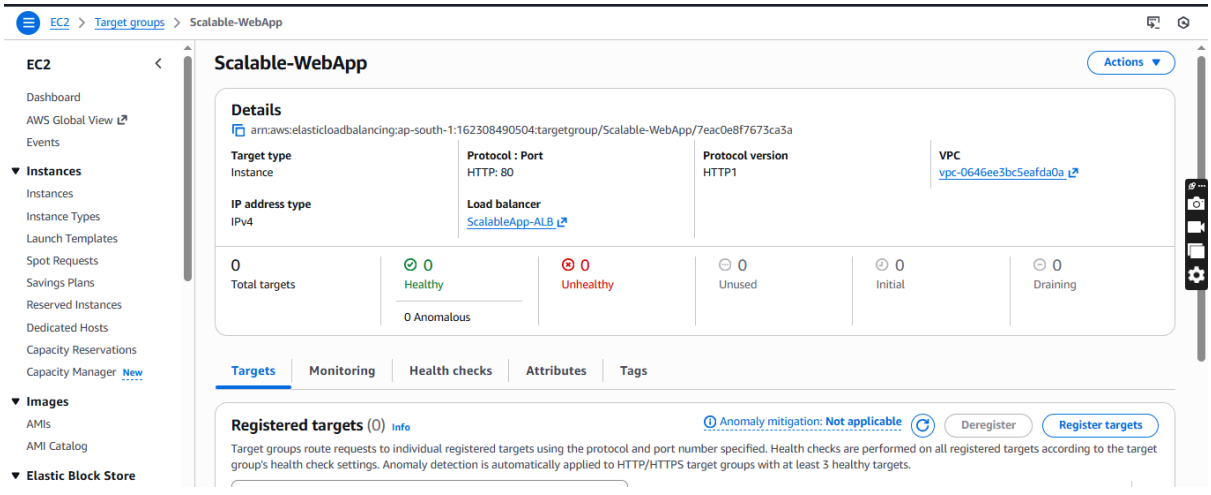
```
#!/bin/bash
yum update -y
yum install httpd -y
systemctl start httpd
systemctl enable httpd
echo "<h1>Scalable Web App Running on EC2</h1>" > /var/www/html/index.html
```

Click **Create launch template**



STEP 9: Create Target Group

1. Go to **EC2 → Target Groups**
2. Click **Create target group**
3. Choose:
    - Type: Instances
    - Protocol: HTTP
    - Port: 80
    - VPC: ScalableApp-VPC
4. Health check path: /
5. Create target group



STEP 10: Create Application Load Balancer

1. Go to **EC2 → Load Balancers**
2. Click **Create Load Balancer**
3. Choose **Application Load Balancer**

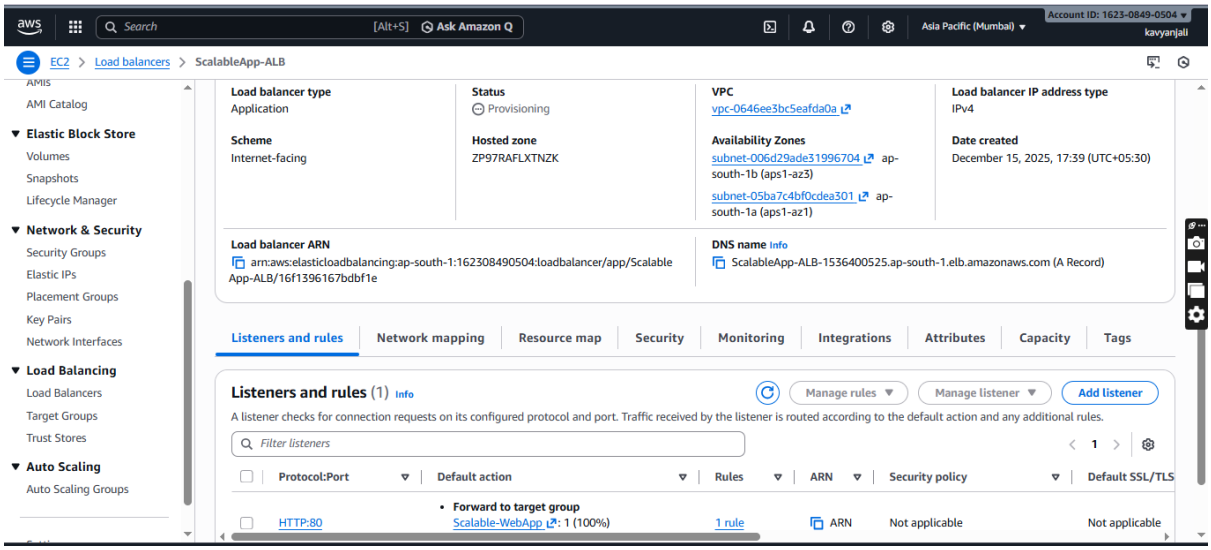
Settings:

- Name: ScalableApp-ALB
- Scheme: Internet-facing
- Subnets: Public Subnet 1 & 2
- Security Group: ALB-SG

Listener:

- HTTP → Forward to Target Group

Click **Create**



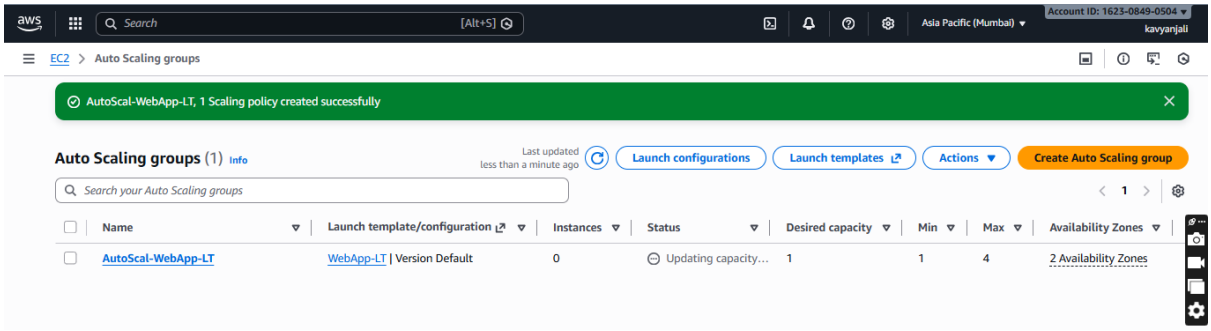
STEP 11: Create Auto Scaling Group

1. Go to **EC2 → Auto Scaling Groups**
2. Click **Create Auto Scaling group**
3. Choose launch template: WebApp-LT
4. VPC: ScalableApp-VPC
5. Subnets: Both public subnets
6. Attach to:
    - Existing Load Balancer
    - Select Target Group

Scaling Policy:

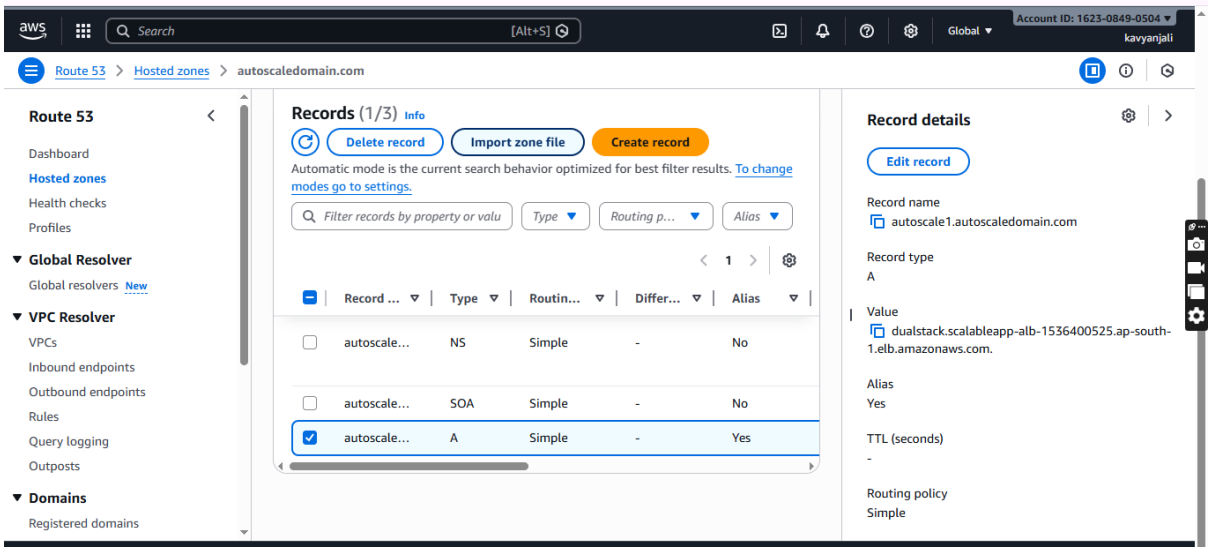
- Min: 1
- Desired: 2
- Max: 4
- Target Tracking:
    - CPU Utilization: 50%

Create Auto Scaling Group



STEP 12: Configure Route 53

1. Go to **Route 53 → Hosted Zones**
2. Create hosted zone:
    - Domain: autoscaledomain.com
3. Create record:
    - Type: A (Alias)
    - Alias target: Application Load Balancer
    - Routing: Simple



STEP 13: Testing & Verification

□ 13.1 Verify Auto Scaling Group Instances

1. Go to **EC2 Console**
2. Click **Auto Scaling Groups**
3. Select your ASG (e.g., WebApp-ASG)
4. Open **Instance management**
5. Confirm:
    - Desired number of instances are **Running**
    - Instances are **InService**

□ 13.2 Verify Target Group Health

1. Go to **EC2 → Target Groups**
2. Select your target group
3. Open **Targets** tab
4. Confirm all instances show **Healthy**

□ 13.3 Verify Load Balancer

1. Go to **EC2 → Load Balancers**
2. Select your **Application Load Balancer**
3. Copy **DNS name: https://scalableapp-alb-1536400525.ap-south-1.elb.amazonaws.com/**
4. Paste DNS name in a web browser
5. Confirm web page loads successfully

□ 13.4 Verify Auto Scaling (Optional Test)

1. Go to **EC2 → Auto Scaling Groups**
2. Select your ASG
3. Click **Edit**
4. Increase **Desired capacity** (e.g., from 2 → 3)
5. Save
6. Check **Instance management**
7. Confirm a **new EC2 instance launches**

□ 13.5 Verify S3 Static Content (If Used)

1. Go to **S3**
2. Open your bucket
3. Click **Properties**
4. Copy **Static website endpoint**
5. Open endpoint in browser
6. Confirm static files load

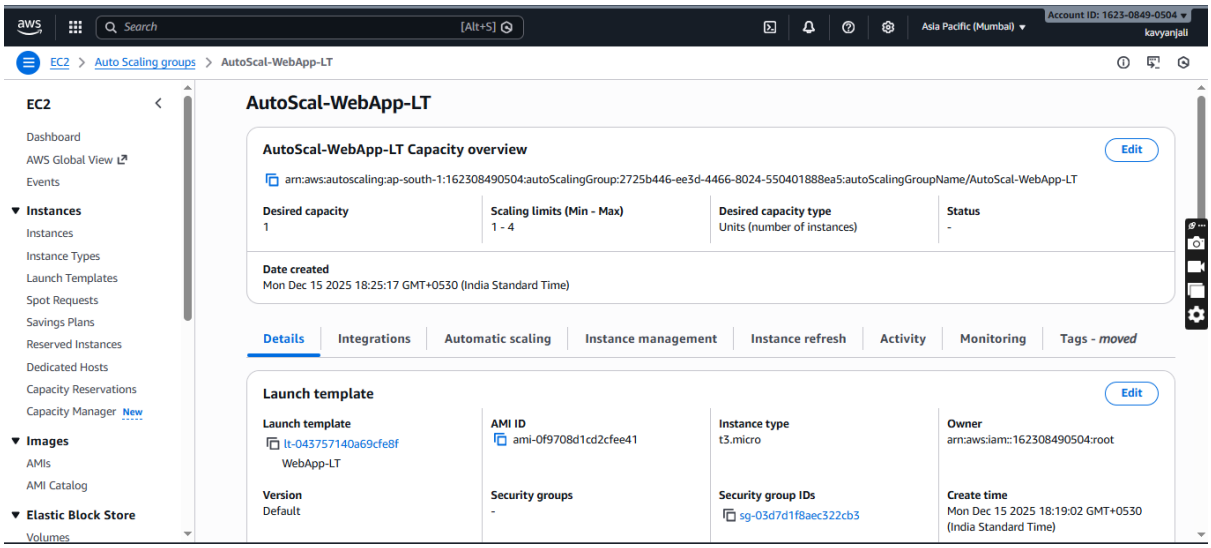
□ 13.6 Verify RDS Connectivity

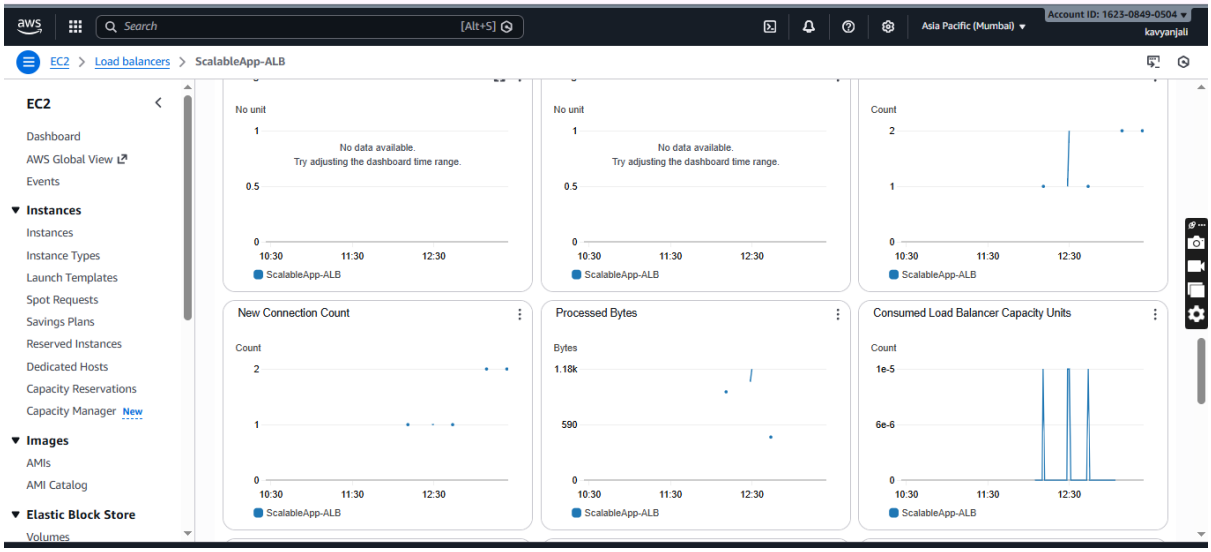
1. Go to **RDS**
2. Select your database
3. Confirm **Status = Available**
4. Note **Endpoint** for application use

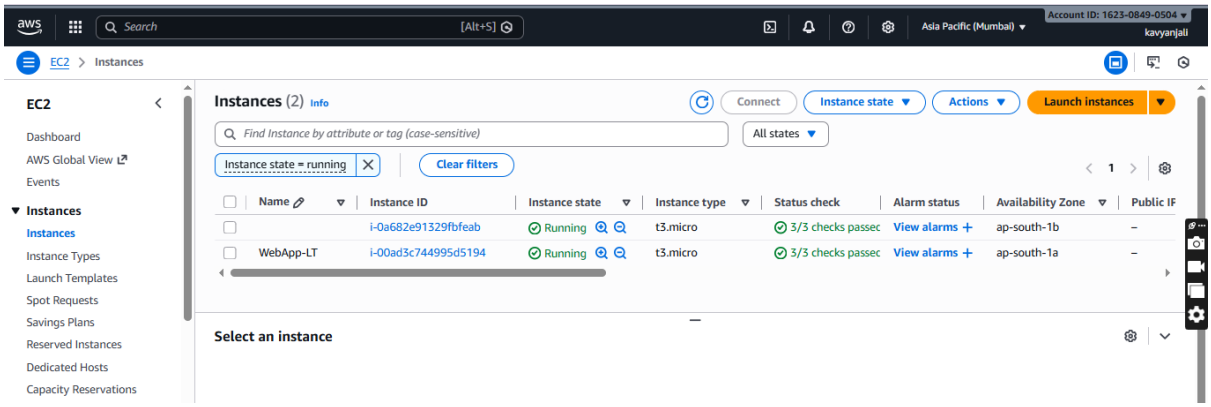
STEP 13 COMPLETE WHEN:

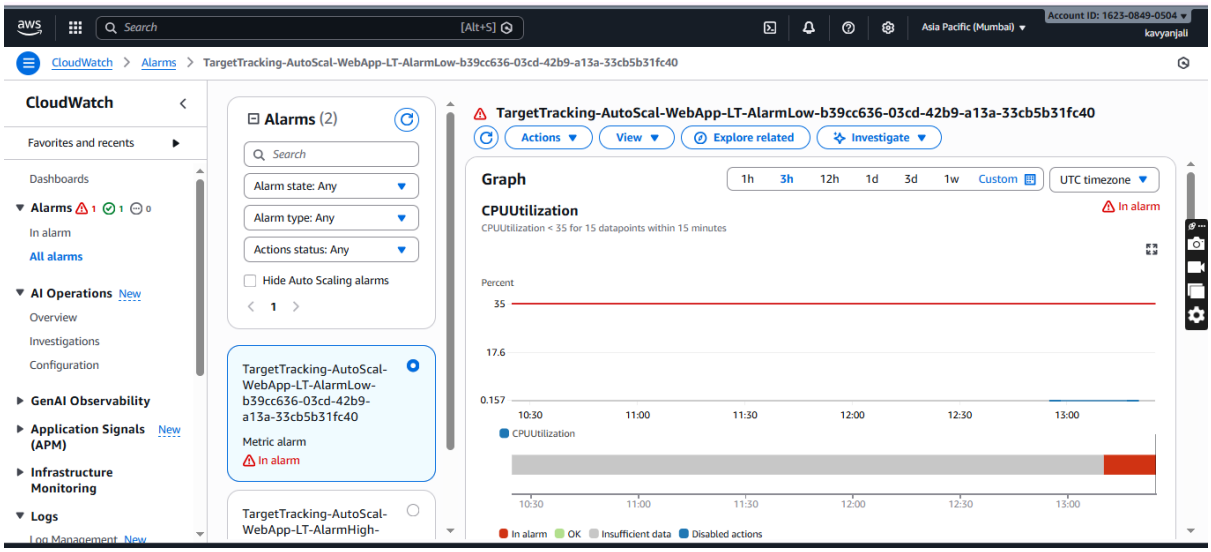
- ALB DNS opens website
- Target group shows **Healthy And Unhealthy**

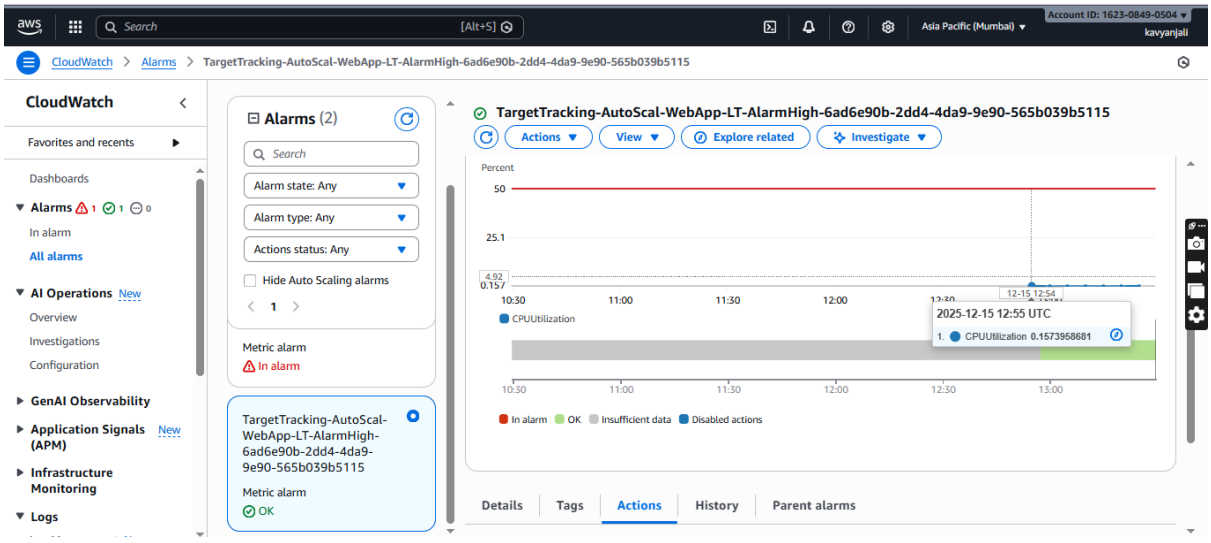
- ASG launches/terminates instances correctly
- RDS is available
- S3 static content accessible











Summary

- A Virtual Private Cloud (VPC) was created to isolate and secure the application network.
- Public subnets were configured across multiple Availability Zones to ensure high availability.
- An Internet Gateway and route table enabled internet access.
- Security Groups were configured to control traffic between the Load Balancer, EC2 instances, and RDS database.
- An S3 bucket was created to host static website assets.
- An RDS MySQL database was deployed within the VPC for persistent data storage.
- An EC2 Launch Template with a user data script was used to automatically configure web servers.
- An Application Load Balancer was configured to distribute traffic across EC2 instances.
- An Auto Scaling Group was set up to automatically scale EC2 instances based on CPU utilization.
- Route 53 was configured to route domain traffic to the Load Balancer.

Conclusion

This setup successfully deploys a **scalable, highly available, and fault-tolerant web application** using AWS managed services. The architecture ensures automatic traffic distribution, dynamic scaling based on demand, secure database access, and reliable static content delivery. This solution improves performance, availability, and operational efficiency while maintaining cost optimization through Auto Scaling.
