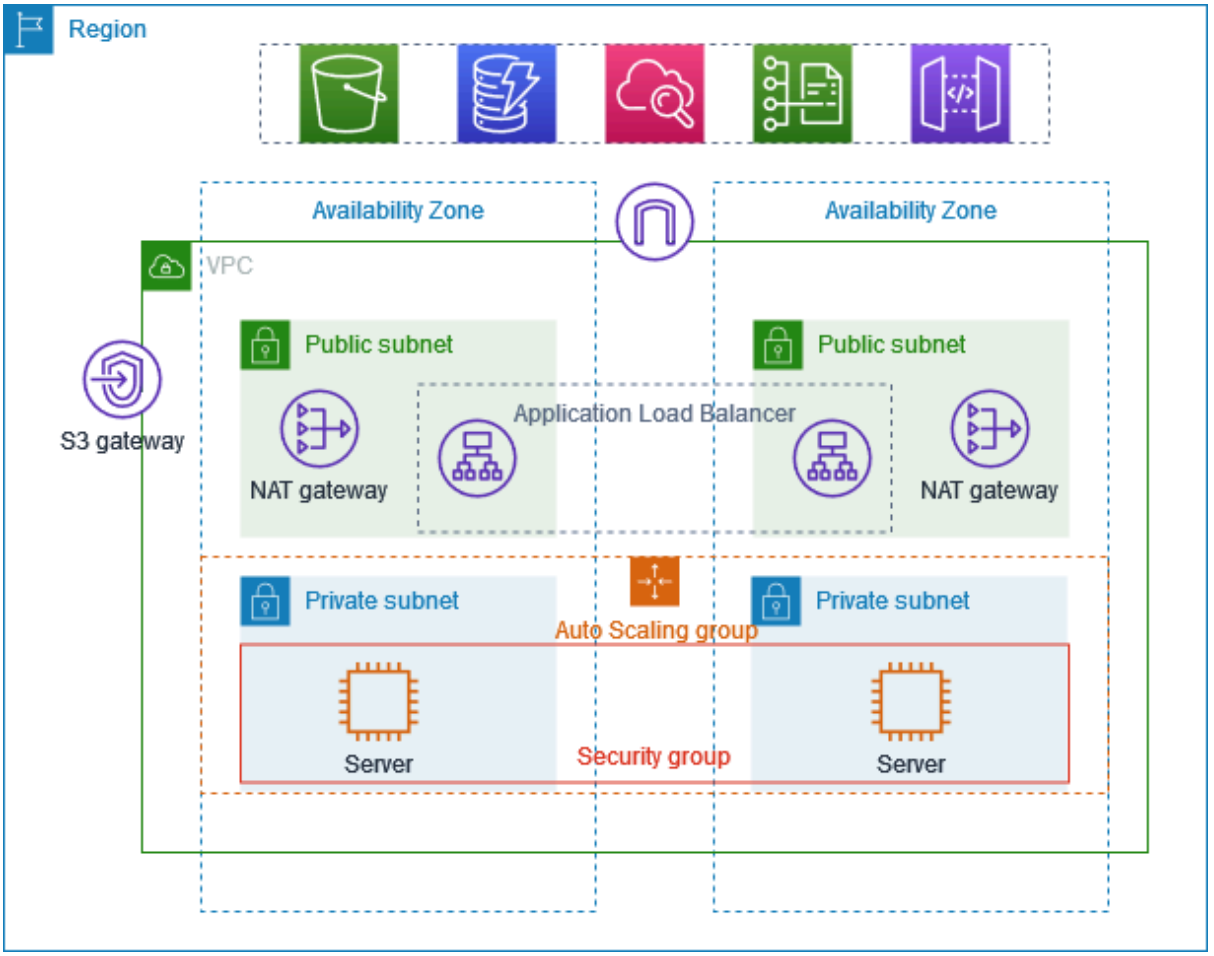
AWS VPC and Private Subnet Application Deployment

Project Overview

This project demonstrates deploying an application in a **private subnet** using AWS services, ensuring security, scalability, and high availability. The architecture includes:

- **Bastion Host** for secure access to private instances.
- **Auto Scaling Group (ASG)** with EC2 instances in private subnets.
- **Application Load Balancer (ALB)** for distributing traffic.
- **NAT Gateway** for outbound internet access from private subnets.
- **Security Groups & Route Tables** for network segmentation.



AWS Services Used

- **Virtual Private Cloud (VPC)** – Custom network for secure deployment.
- **Subnets** – Public and private subnets for network separation.
- **Internet Gateway (IGW)** – Allows public subnet resources to access the internet.
- **NAT Gateway** – Enables private instances to securely access the internet.
- **EC2 Instances** – Compute resources for the application.
- **Bastion Host** – Acts as a secure intermediary to access private instances.
- **Security Groups** – Firewall rules to control access.
- **Application Load Balancer (ALB)** – Distributes traffic across instances.
- **Auto Scaling Group (ASG)** – Ensures high availability and scalability.

Implementation Steps

1. Create a VPC

1. Navigate to the **AWS VPC Dashboard**.
2. Click **Create VPC**, provide a **name** (e.g., `aws-prod`).
3. Choose an **IPv4 CIDR Block** (e.g., `10.0.0.0/16`).
4. Click **Create VPC**.

2. Create Public and Private Subnets

1. In the **VPC Dashboard**, go to **Subnets**, click **Create Subnet**.
2. Choose the **created VPC** (`aws-prod`).
3. **Create two public subnets** in different Availability Zones (e.g., `10.0.1.0/24`, `10.0.3.0/24`).
4. **Create two private subnets** in different Availability Zones (e.g., `10.0.2.0/24`, `10.0.4.0/24`).
5. Ensure subnets are **properly tagged** for identification.

3. Attach an Internet Gateway (IGW) to the VPC

1. In the **VPC Dashboard**, go to **Internet Gateways**.
2. Click **Create Internet Gateway**, provide a **name** (e.g., `aws-prod-igw`).
3. Click **Attach to VPC**, select `aws-prod`, and confirm.

Update Route Tables:

1. **Create a new Route Table** for the **Public Subnets**.
2. Add a route: `0.0.0.0/0 → Internet Gateway`.
3. Associate this route table with the **public subnets**.

4. Create and Attach a NAT Gateway

1. Go to **NAT Gateways**.
2. Click **Create NAT Gateway**.
3. Assign it to one of the **public subnets** (e.g., `10.0.1.0/24`).
4. **Allocate an Elastic IP** and attach it to the NAT Gateway.

Modify the Private Subnet Route Tables:

1. Add a route: `0.0.0.0/0 → NAT Gateway`.
2. Associate this route table with the **private subnets**.

---

5. Create Security Groups

Bastion Host Security Group

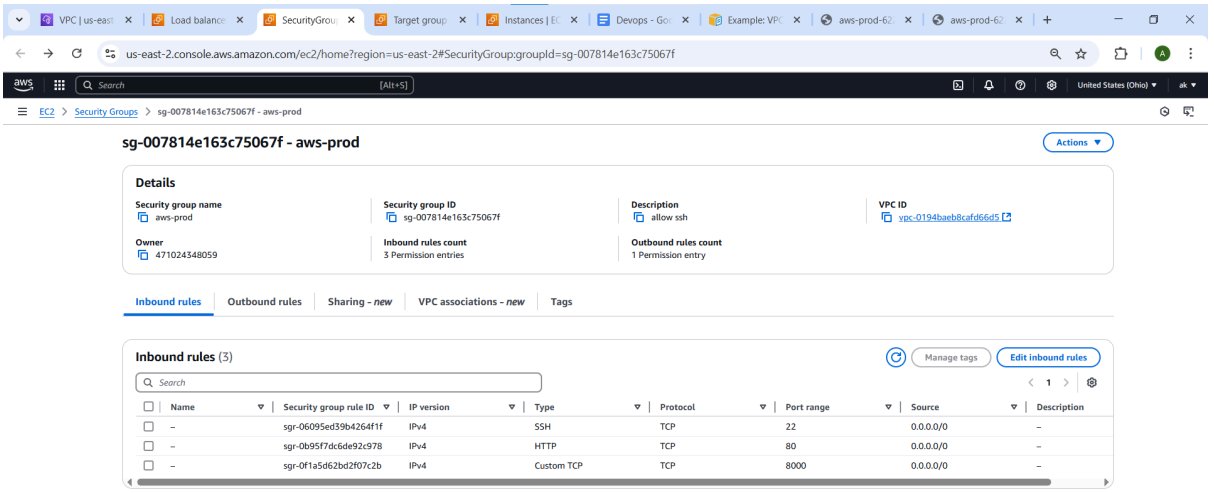
- **Inbound**: Allow SSH (`port 22`) **only from your IP**.

Private Instance Security Group

- **Inbound**: Allow SSH (`port 22`) **only from Bastion Host**.

Load Balancer Security Group

- **Inbound**: Allow HTTP (`port 80`) **from the internet**.
- **Inbound**: Allow application traffic (`port 8000`).

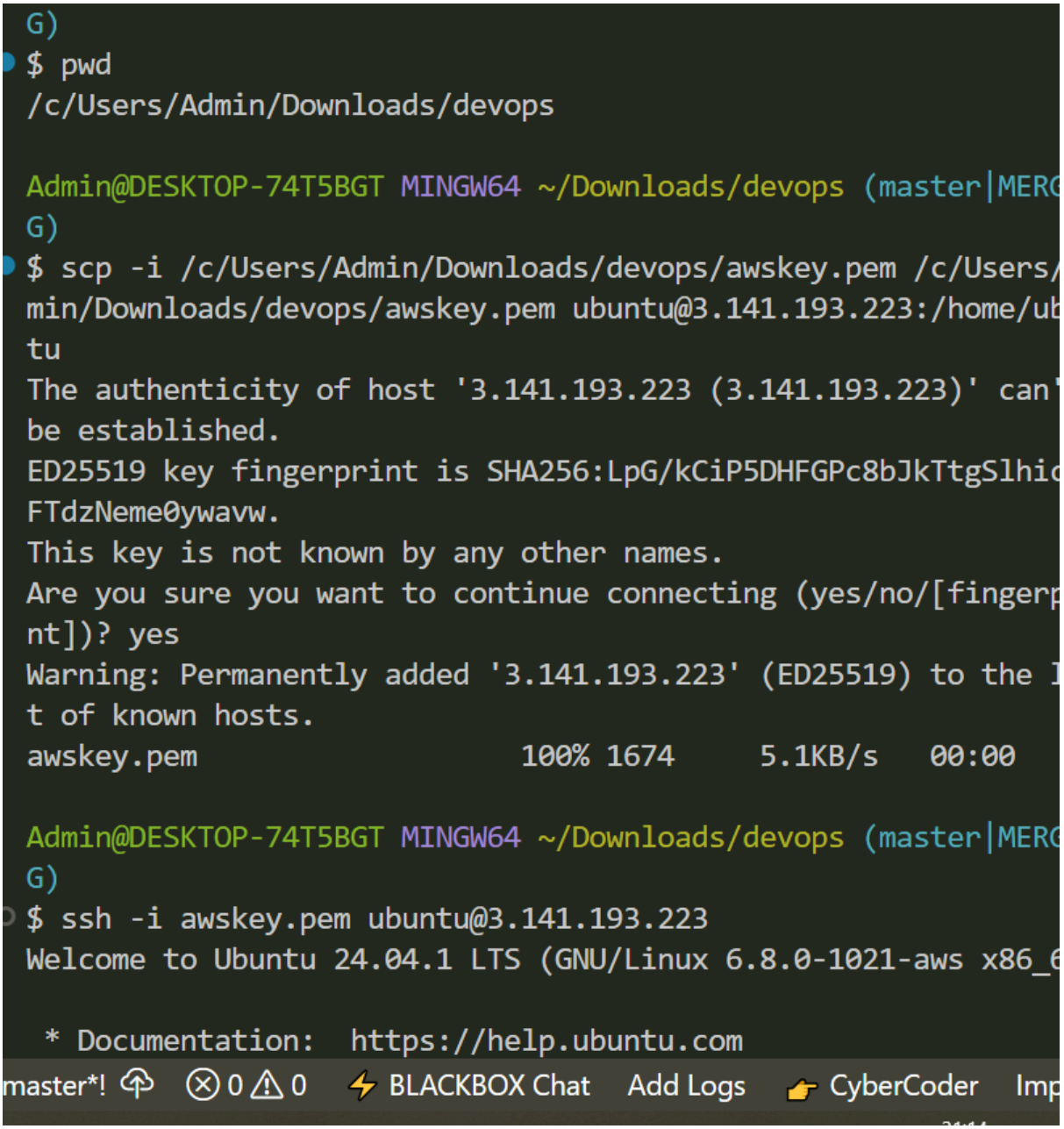


6. Deploy a Bastion Host in Public Subnet

1. Launch an **EC2 Instance** in the **public subnet**.
2. Choose **Ubuntu** as the AMI.
3. Select **T2.micro** instance type.
4. Attach **Bastion Host Security Group**.
5. **Enable Auto-assign Public IP**.

SSH into the Bastion Host:

ssh -i my-key.pem ubuntu@<bastion-public-ip>



---

7. Deploy Application Instances in Private Subnet

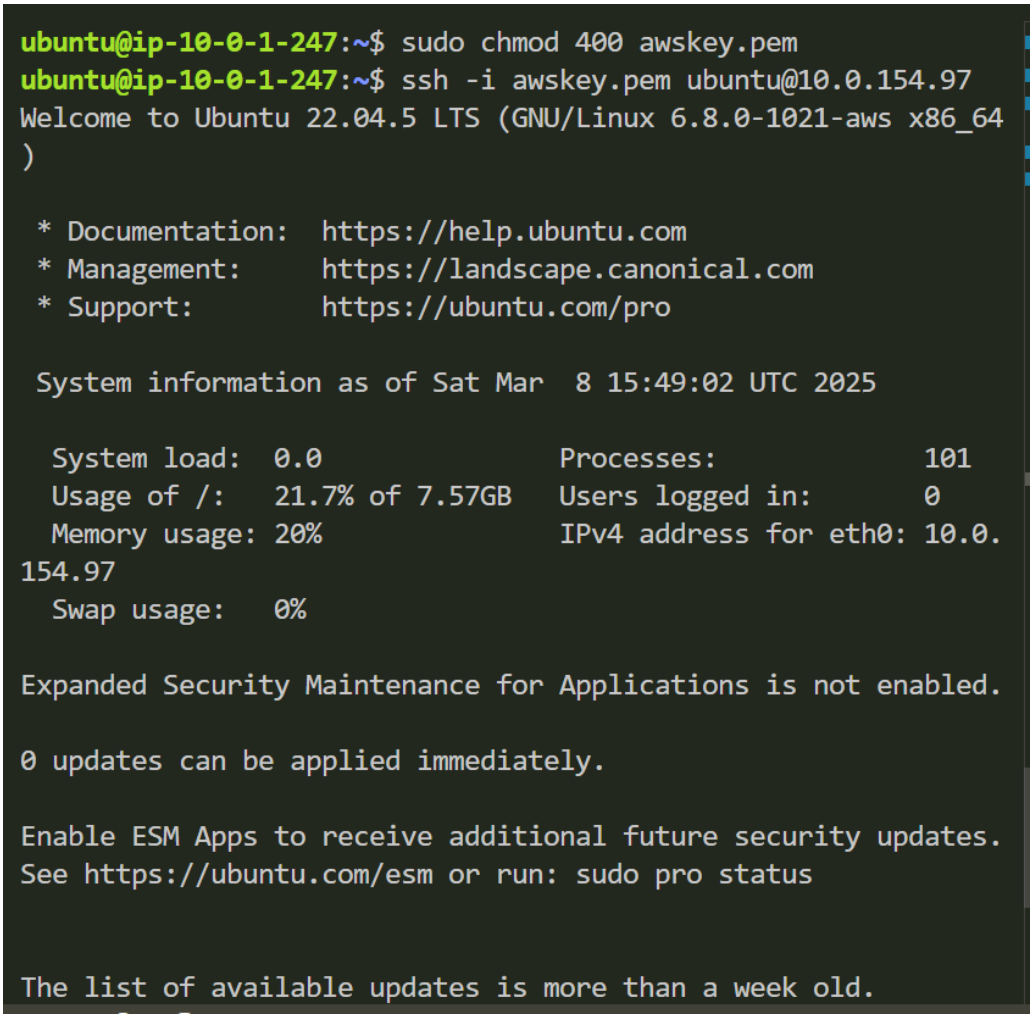
1. Launch **two EC2 instances** in the **private subnet**.
2. Attach **Private Instance Security Group**.

Copy SSH Key to Bastion Host:

scp -i my-key.pem my-key.pem ubuntu@<bastion-public-ip>:~

SSH from Bastion Host to Private Instance:

ssh -i my-key.pem ubuntu@<private-instance-ip>



---

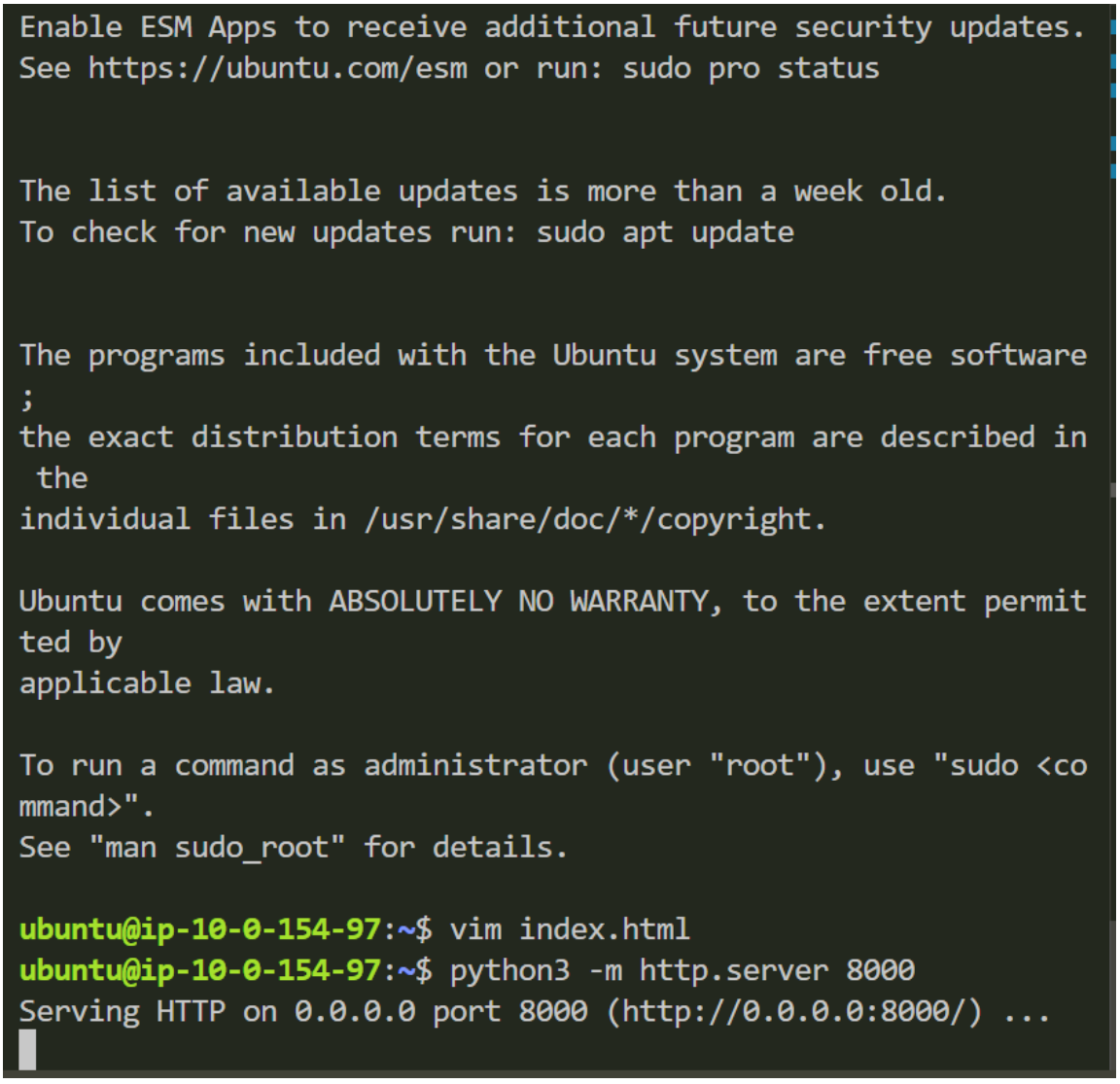
8. Install a Simple Python Web Application

On the **private EC2 instance**, create an `index.html` file:

<html>
<head><title>My first AWS Private to demonstrate app in private subnet</title></head>
<body>
  <h1>My first project using bastion-host</h1>
</body>
</html>

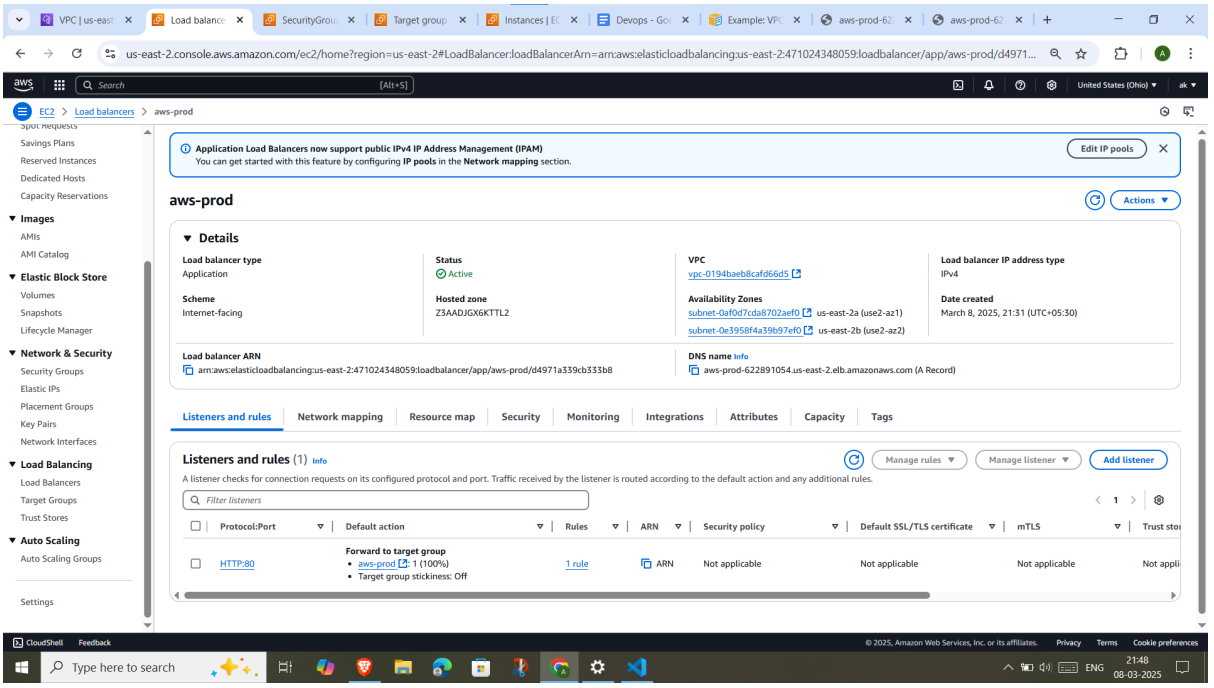
Start a **Python HTTP Server**:

python3 -m http.server 8000



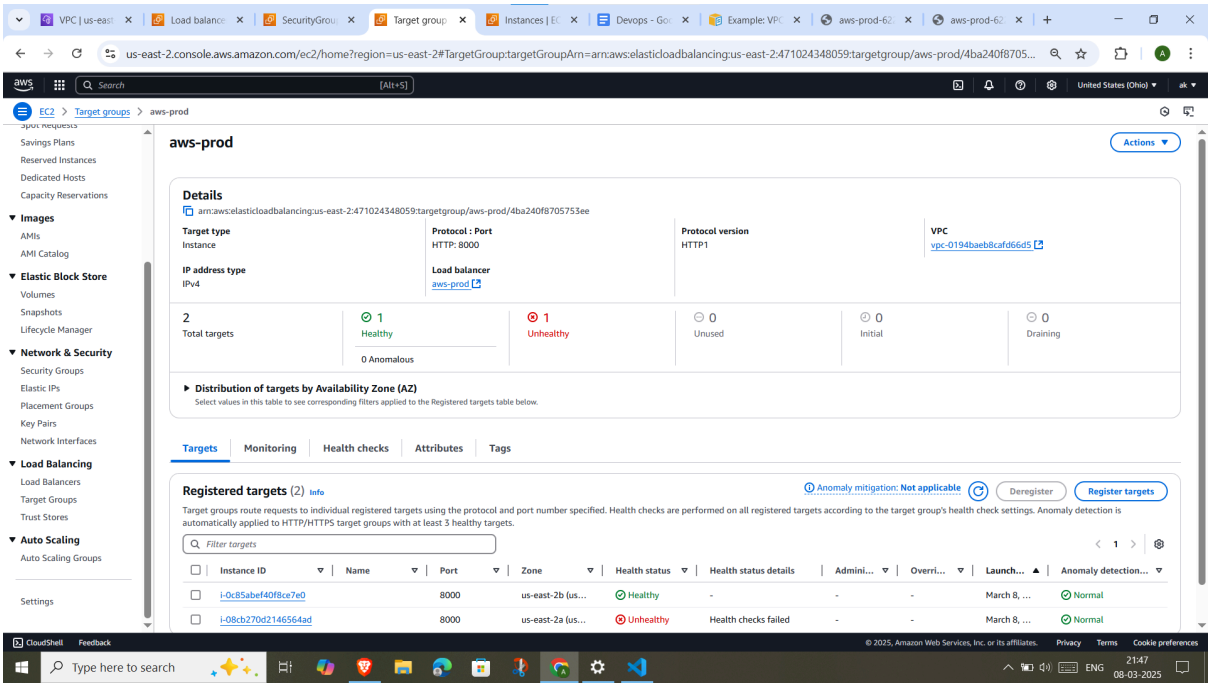
9. Create an Application Load Balancer (ALB)

1. Go to **EC2 > Load Balancers**.
2. Choose **Application Load Balancer**.
3. Set it as **Internet-facing**.
4. Assign it to the **Public Subnet**.
5. Attach **Load Balancer Security Group**.
6. Configure a **Target Group** to include **private EC2 instances**.
7. Configure **Health Checks** on **Port 80**.



10. Configure Auto Scaling Group (ASG)

1. Go to **EC2 > Auto Scaling Groups**.
2. Create a **Launch Template** with private instance settings.
3. Set **desired, minimum, and maximum** instances.
4. Attach the **Target Group** from Load Balancer.



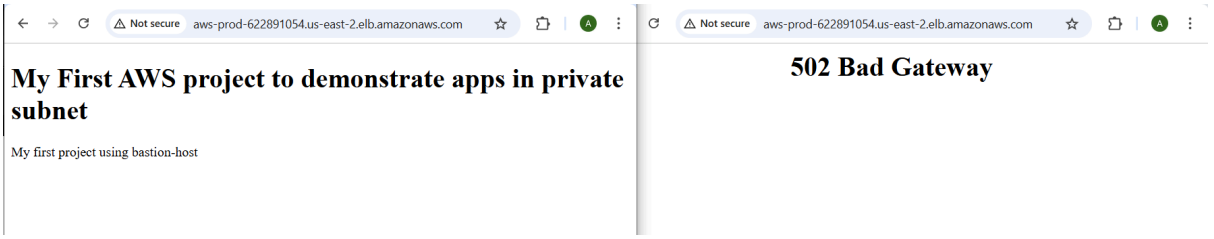
11. Test Load Balancer & Auto Scaling

1. **Get Load Balancer DNS Name** from EC2 Console.
2. Access the application in a browser:

http://<load-balancer-dns>

Verify Load Balancer

- **One instance responds** with `index.html` output.
- **Other shows "Bad Gateway"** (because it's missing `index.html`).
- **Target Group** shows **one healthy, one unhealthy** instance.



Key Takeaways

✅ **Bastion Host** allows secure SSH access to private instances.
✅ **Load Balancer** distributes traffic and enhances availability.
✅ **Auto Scaling** ensures high availability and scalability.
✅ **Security Groups** control access effectively.
✅ **NAT Gateway** provides secure outbound connectivity for private instances.