

Designing a Highly Available 3-Tier Web Application Architecture on AWS

What is AWS?

Amazon Web Services (AWS) is a comprehensive and widely adopted cloud platform offering over 200 fully featured services from data centers globally. These services include computing power, storage options, and various tools for databases, machine learning, analytics, and more. AWS helps businesses scale and grow by providing flexible, reliable, scalable, easy-to-use, and cost-effective cloud computing solutions.

Why Use AWS?.....

1. Scalability:

- Automatic Scaling: AWS can automatically scale resources up or down based on demand, ensuring that applications run smoothly without overprovisioning or underutilizing resources.
- Elastic Load Balancing: Distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses, improving availability and fault tolerance.

2. Cost Efficiency:

- Pay-as-You-Go: AWS follows a pay-as-you-go pricing model, meaning you only pay for the services you use, reducing upfront capital expenditure.
- Cost Management Tools: AWS provides tools like AWS Cost Explorer and AWS Budgets to monitor and optimize spending.

3. Security:

- Data Protection: AWS offers robust security services and features, including encryption, threat detection, and identity management.
- Compliance Certifications: AWS meets a wide range of compliance standards such as GDPR, HIPAA, and SOC.

4. Global Reach:

- Global Infrastructure: AWS operates in multiple geographic regions worldwide, ensuring low latency and high availability.
- Content Delivery Network (CDN): Amazon CloudFront delivers content globally with low latency and high transfer speeds.

Real-Time Use Cases

1. Web Hosting:

- Scalable Websites: Companies can host scalable websites using Amazon EC2, Amazon S3, and Amazon RDS.
- Static and Dynamic Content: AWS supports both static (via Amazon S3) and dynamic content (using Amazon EC2 and databases).

2. Data Backup and Storage:

- Reliable Storage: AWS offers reliable storage solutions such as Amazon S3 and Amazon Glacier for backups, data archiving, and disaster recovery.
- High Durability: Amazon S3 provides 99.999999999% (11 9's) of data durability.

3. Big Data and Analytics:

- Data Processing: AWS services like Amazon EMR, AWS Glue, and Amazon Redshift help process and analyze large datasets.
- Real-Time Analytics: Amazon Kinesis allows for real-time data streaming and analytics.

4. DevOps:

- Continuous Integration and Delivery (CI/CD): AWS offers tools like AWS CodePipeline, AWS CodeBuild, and AWS CodeDeploy for automating code deployments.
- Infrastructure as Code (IaC): Services like AWS CloudFormation and AWS CDK allow developers to define infrastructure using code.

What is 3 tier architecture?.....

A 3-tier architecture is a well-established software architecture that separates applications into three logical and physical computing tiers. Each tier is responsible for specific tasks and can be developed, maintained, and scaled independently. The three tiers are:

1. Presentation Tier (Client Tier):

- This is the user interface layer, where the user interacts with the application. It can be a web browser, desktop application, or mobile app.
- It is responsible for displaying information to the user and collecting user input.
- Technologies commonly used: HTML, CSS, JavaScript, React, Angular, Vue.js.

2. Application Tier (Logic Tier or Middle Tier):

- This is the business logic layer, where data processing occurs. It acts as an intermediary between the presentation tier and the data tier.
- It handles business rules, computations, and data manipulation.
- Technologies commonly used: Java, .NET, Python, Node.js, Ruby on Rails.

3. Data Tier (Database Tier or Storage Tier):

- This is the database layer, where data is stored and managed. It includes databases and data storage systems.
- It handles data retrieval, storage, and management.
- Technologies commonly used: MySQL, PostgreSQL, MongoDB, Oracle, SQL Server.

Benefits of 3-Tier Architecture:

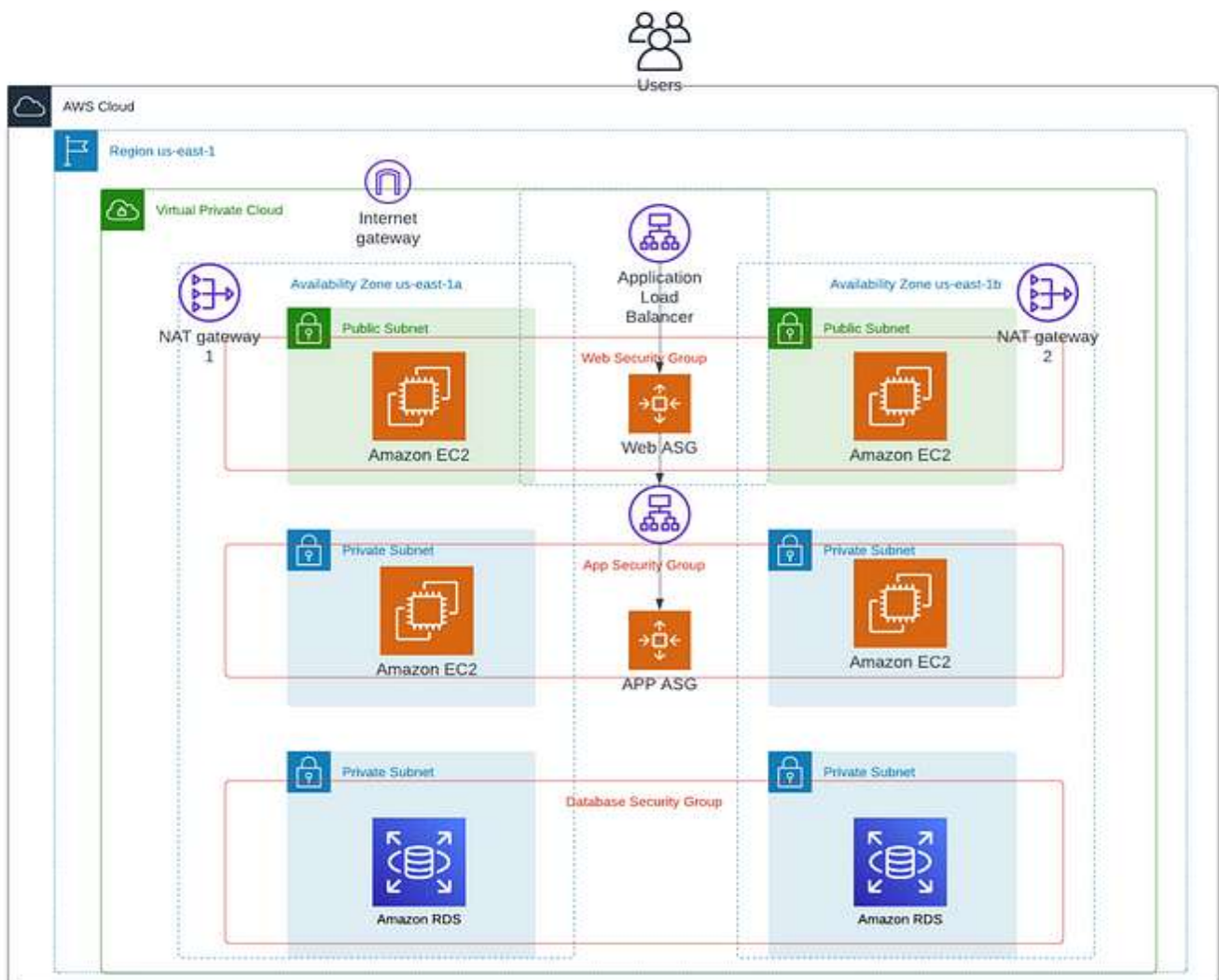
- Separation of Concerns: Each tier can be developed and maintained independently, making the application more modular and easier to manage.
- Scalability: Each tier can be scaled independently based on load and performance requirements.
- Maintainability: Changes in one tier do not necessarily impact the other tiers, making it easier to update and maintain.
- Reusability: Business logic and data access layers can be reused across different applications and projects.

- Flexibility: Different technologies can be used in each tier, allowing for the best tools to be chosen for specific tasks.

Example Scenario:

For a simple e-commerce application:

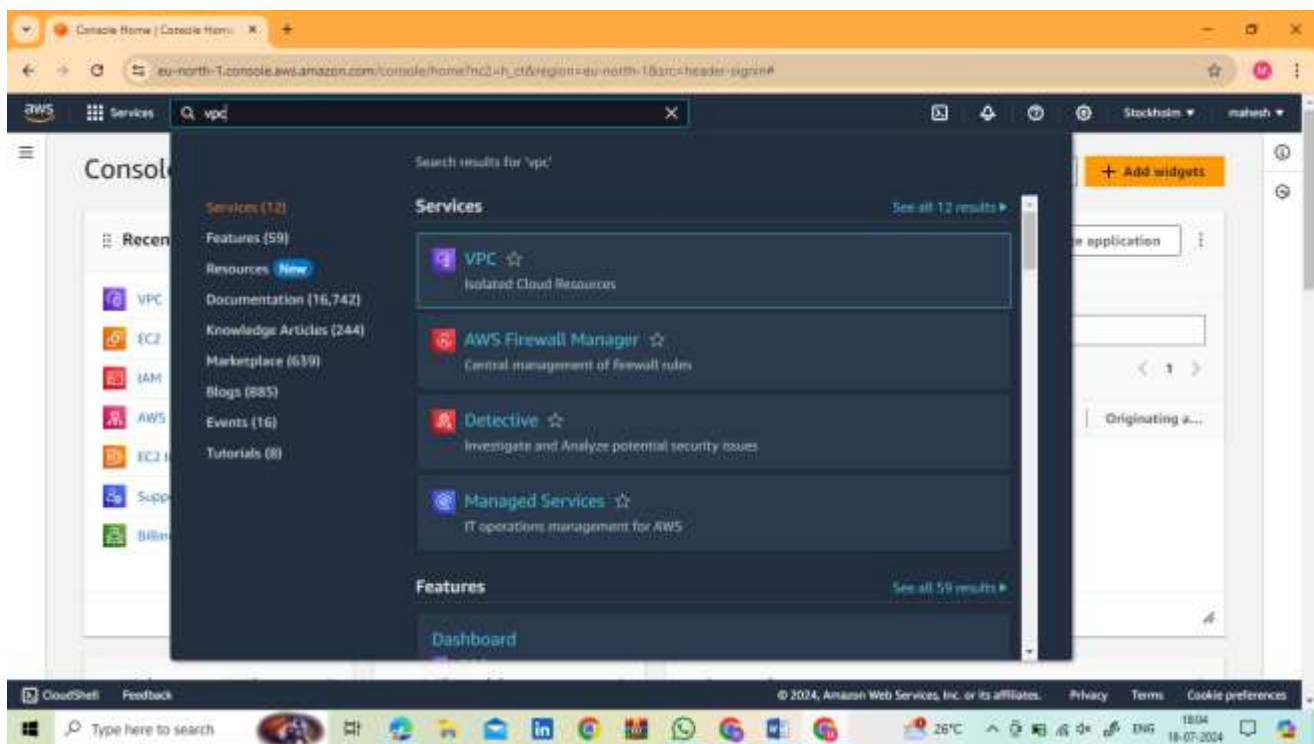
- Presentation Tier: The website or mobile app where users browse products and make purchases.
- Application Tier: The server that processes orders, manages user sessions, and handles business logic such as calculating totals and applying discounts.
- Data Tier: The database that stores product information, user details, order history, and inventory levels.

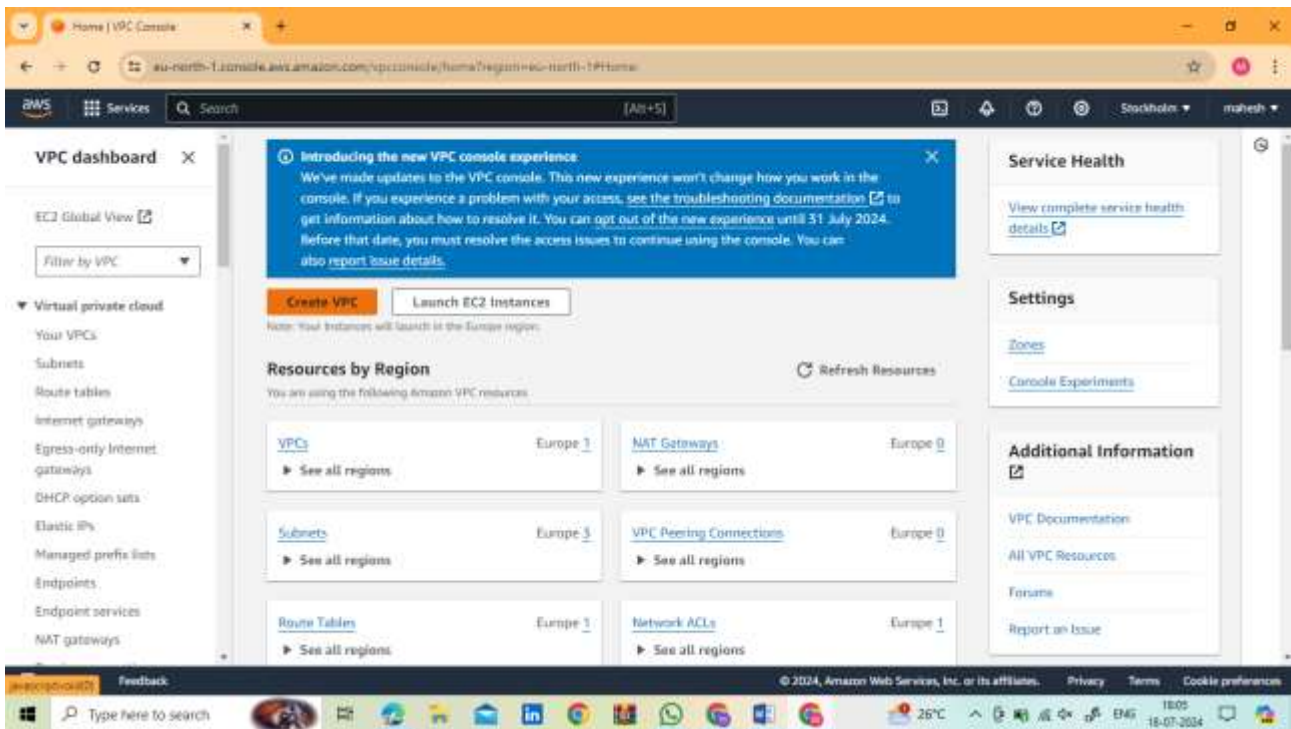


Step1-----creating Vpc

Based on the architecture diagram as a reference, our initial step involves creating a new VPC containing 2 public subnets and 4 private subnets.

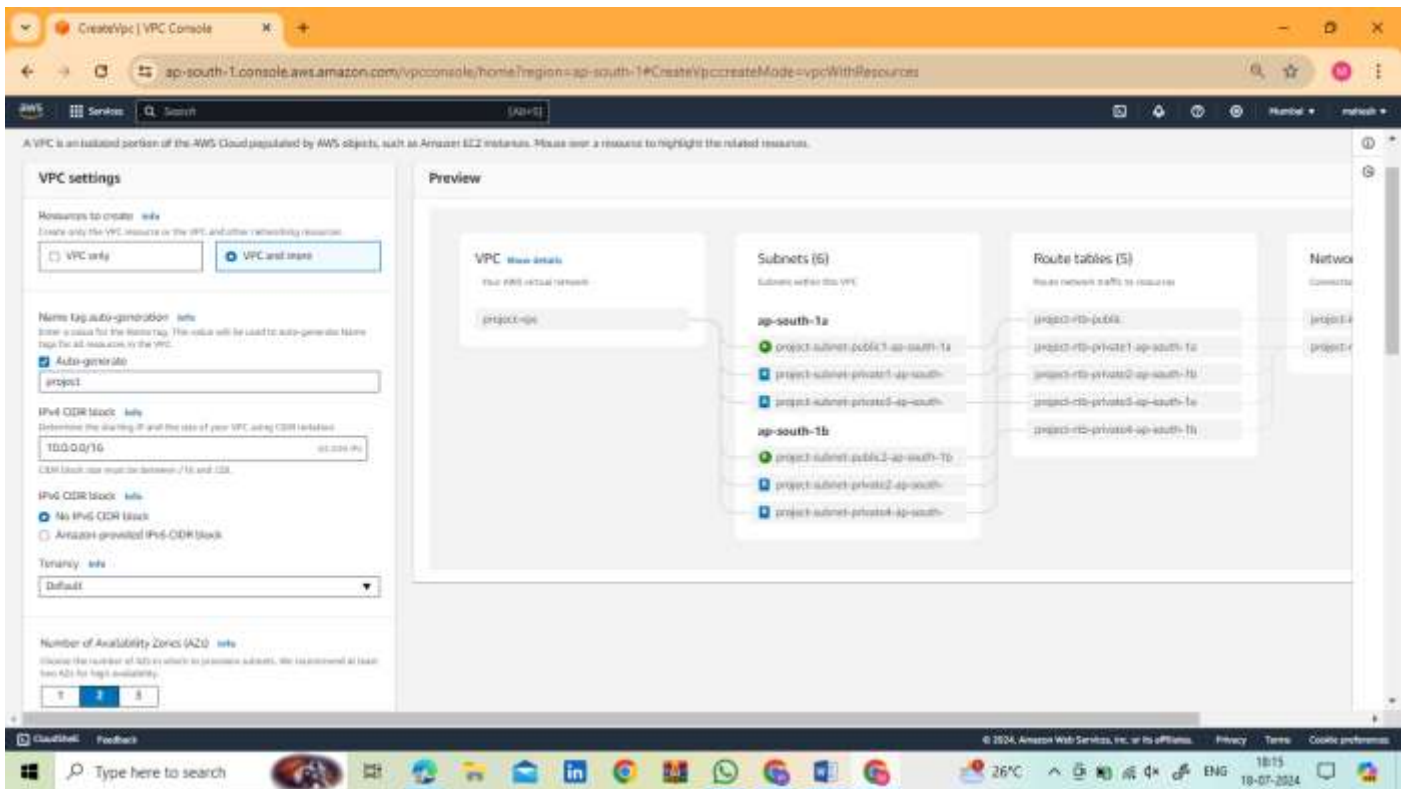
To begin, please log into the AWS Management Console and click on the "Create VPC" button.





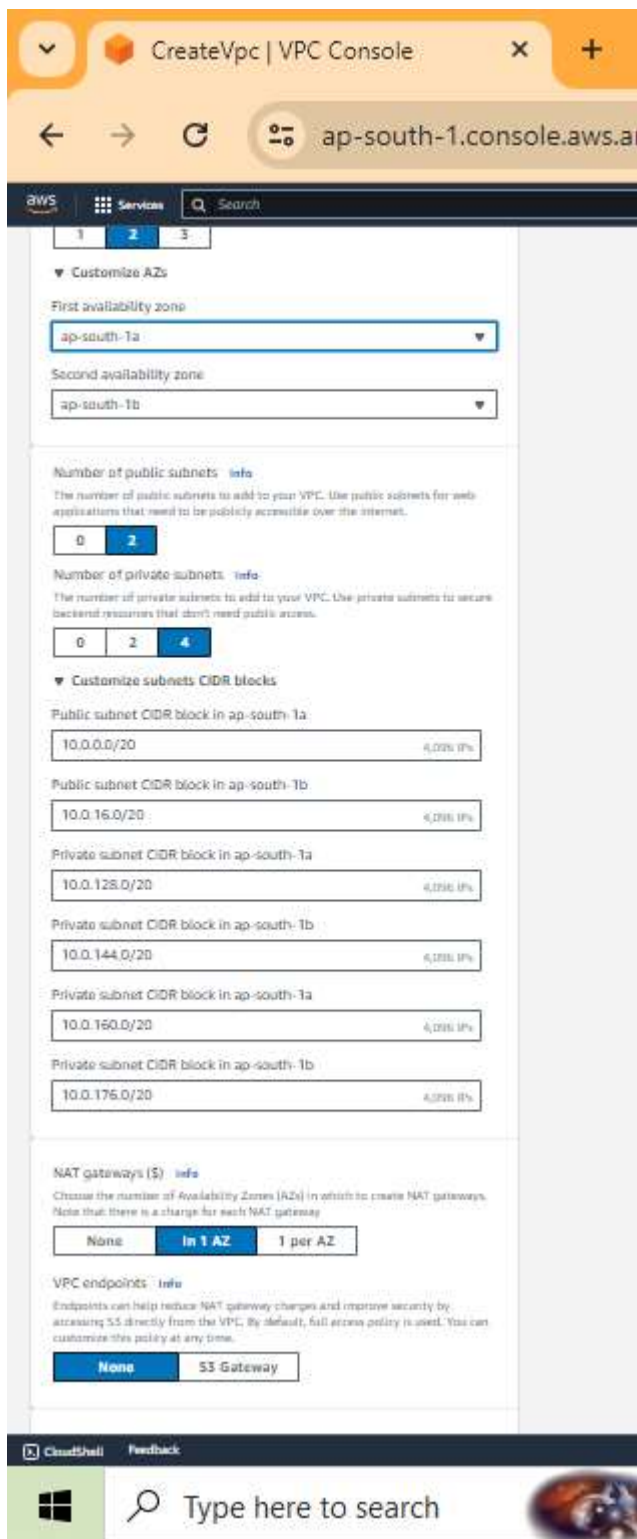
Let's proceed with creating a VPC that includes multiple public and private subnets across 2 Availability Zones. Ensure the following settings:

- Choose "VPC and more" option.
- Name your VPC.
- Use the auto-assigned IPv4 CIDR block of "10.0.0.0/16".
- Select no IPv6 support.
- Choose default Tenancy.
- Specify 2 public subnets and 4 private subnets.

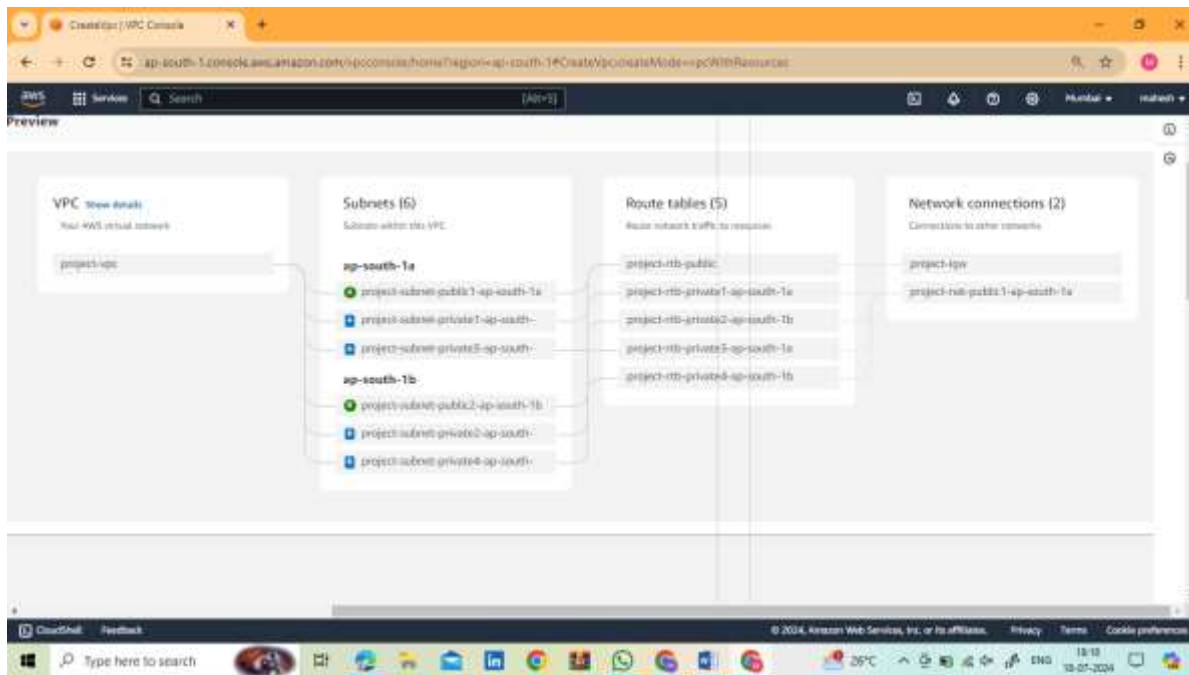


Before creating the VPC, let's expand and customize the Availability Zones (AZs) and the CIDR blocks for subnets.

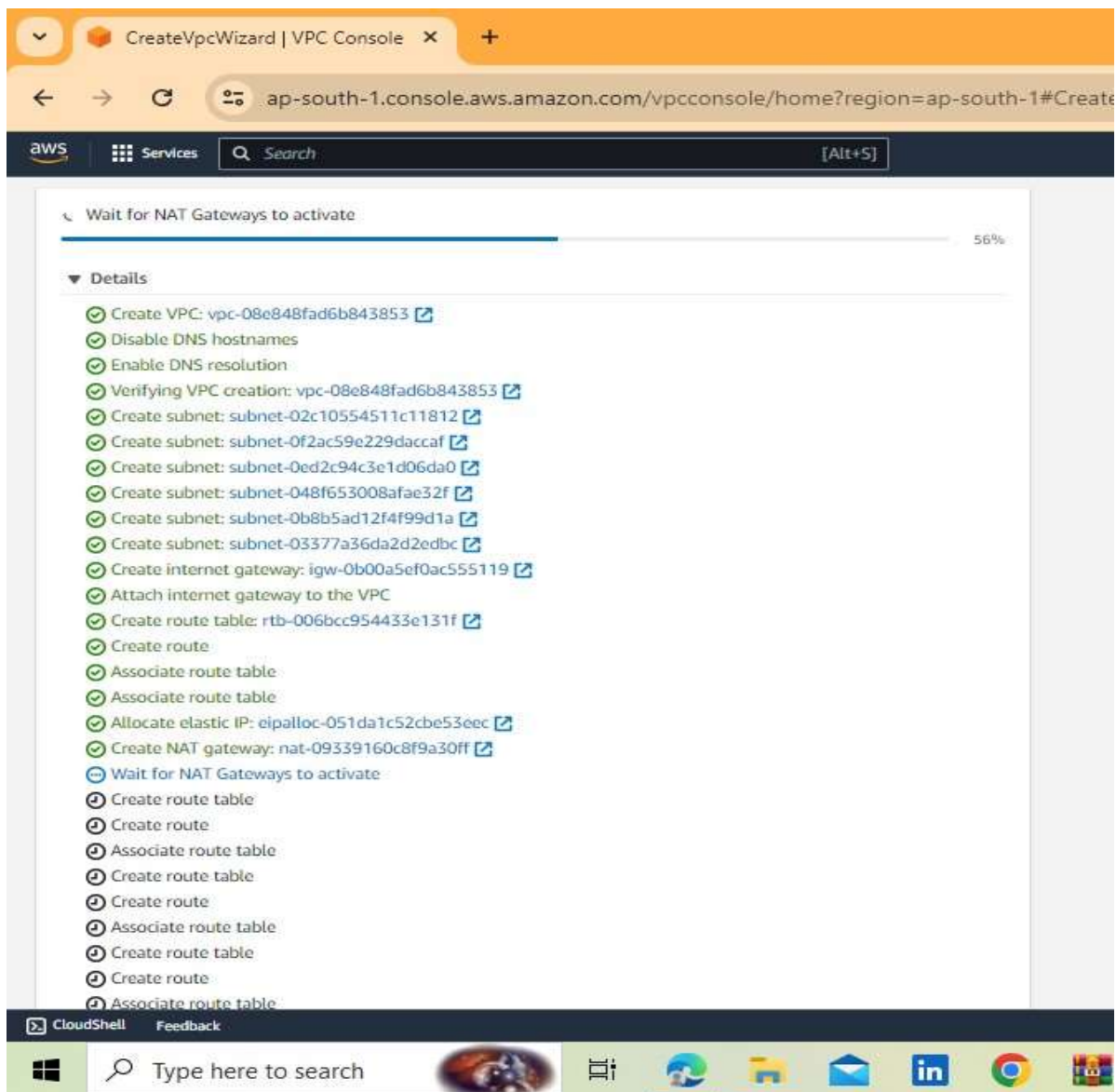
For the NAT gateway, select "in 1 AZ". Do not configure VPC endpoints. Ensure that the options "Enable DNS hostnames" and "Enable DNS resolution" remain checked.



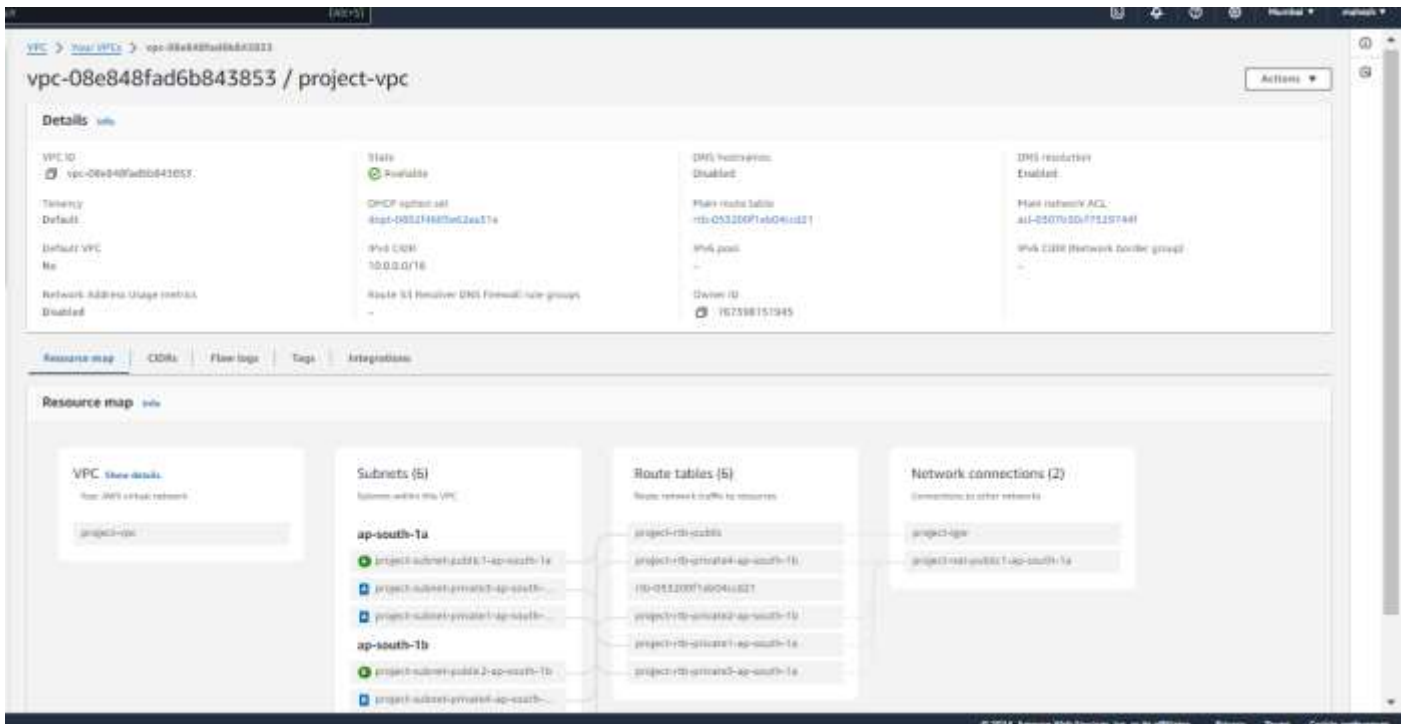
To proceed, click on the "Create VPC" button. The diagram below illustrates the route that your new VPC will follow.



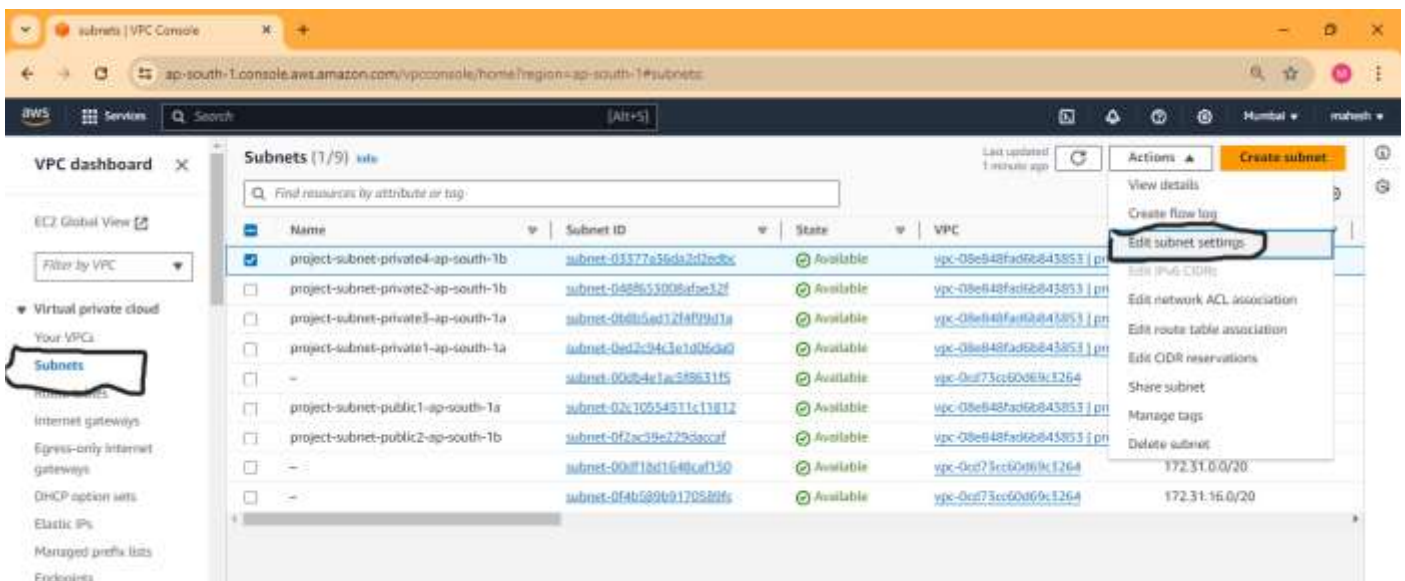
Once you click "Create VPC," you will be presented with a workflow chart that visually represents the creation process of your resources.

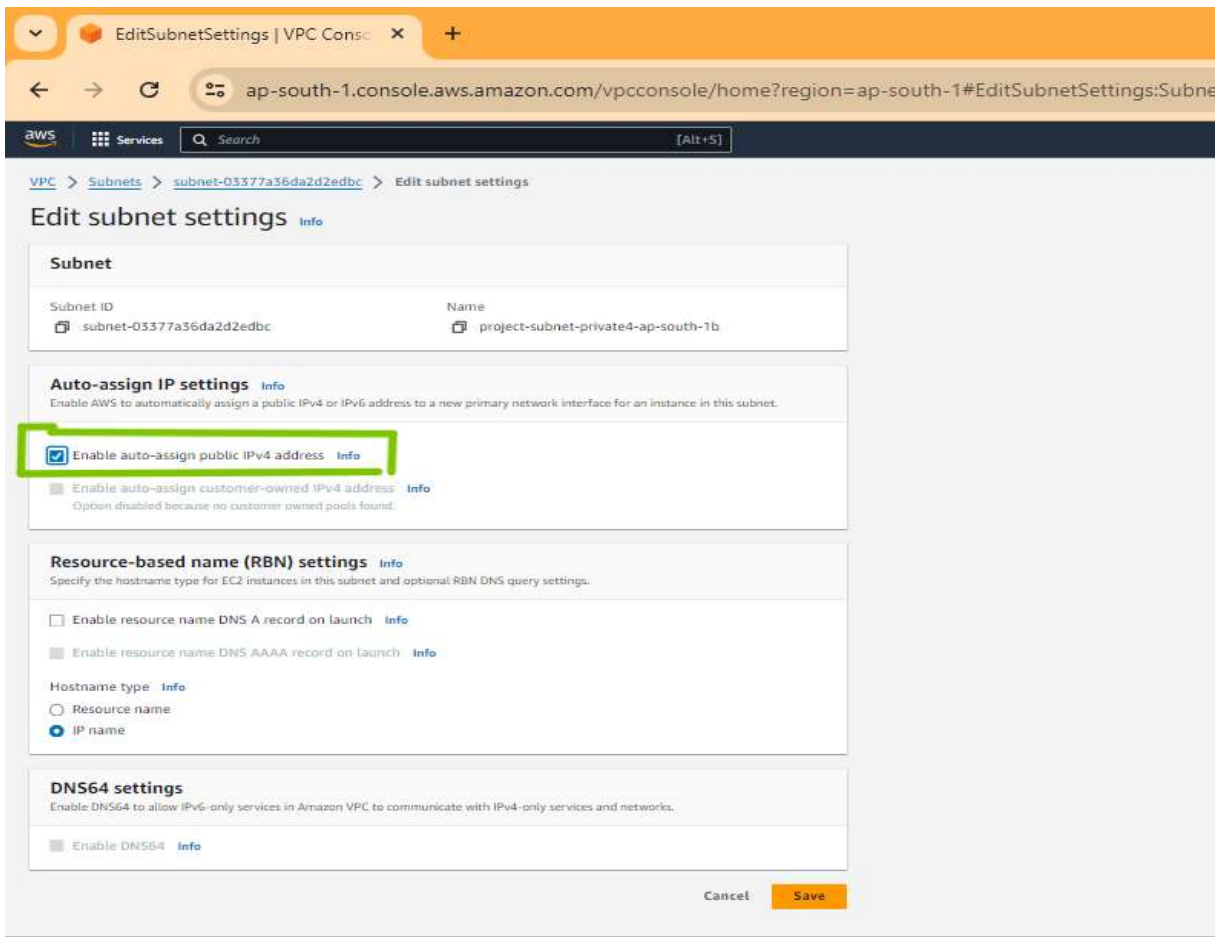


After the creation process completes, you can view your new VPC in the AWS Management Console to ensure everything is set up correctly.



Next, navigate to the Subnets tab in the VPC console. Select one of the newly created subnets. Under the "Actions" tab, expand the dropdown menu and choose "Edit subnet settings."



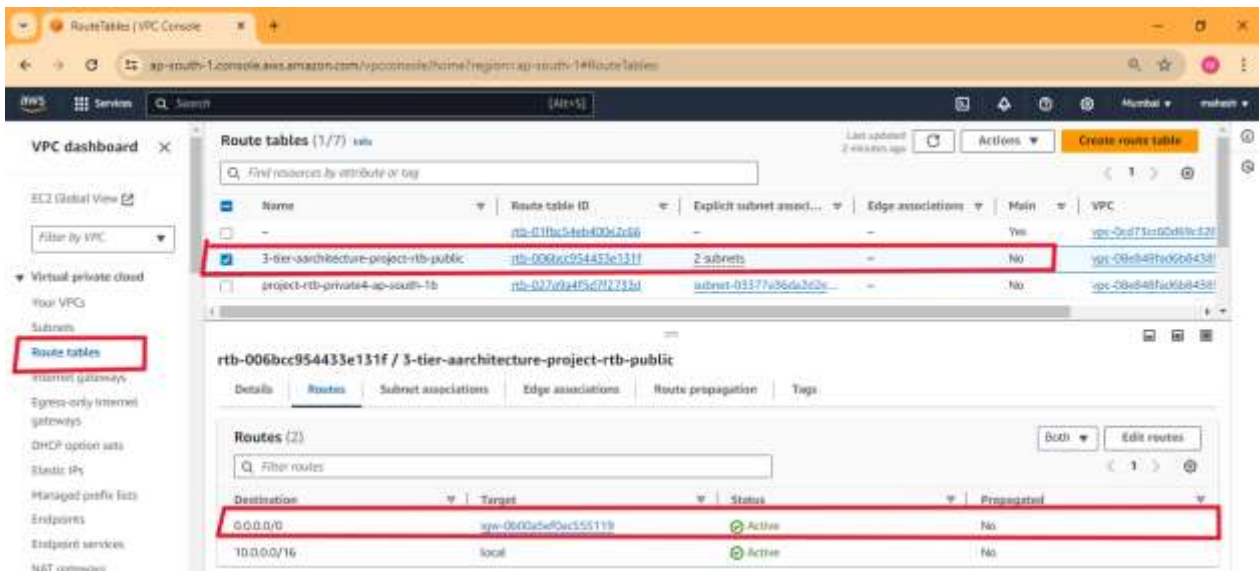


Check the box for "Enable auto-assign IPv4 address" and then click "Save." Repeat this step for all six of the newly created subnets.

Update Web Tier Public Route Table:

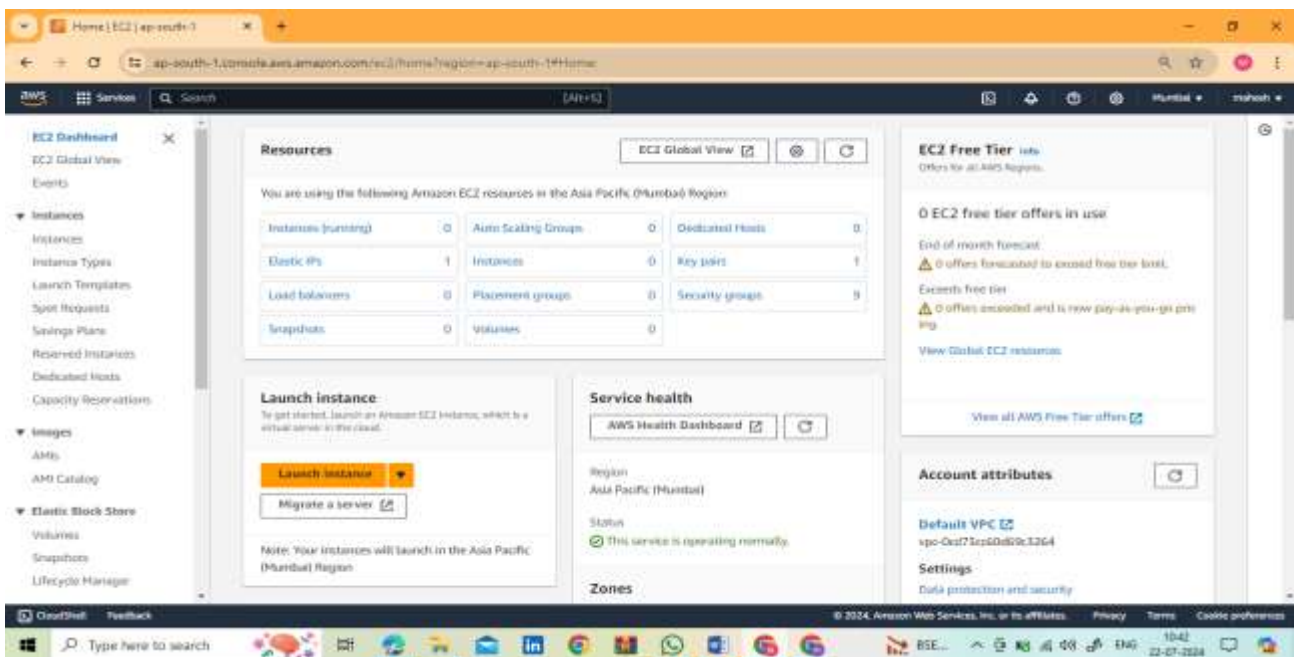
To ensure the correct subnets are associated with the appropriate route table:

1. Navigate to the Route Tables tab under the VPC dashboard.
2. Verify that the automatically created route table is correctly associated with the green-highlighted public subnets. If not, click "edit subnet associations" and select the required subnets.



Step-2: Creating a Web Server Layer

Next, we will create our first tier, representing our front-end user interface (web interface). We will create an auto-scaling group of EC2 instances that will host a custom webpage for us. Start by heading to launch an EC2 instance.



Give your instance a name and choose an AMI. I can use Amazon Linux 2023 AMI.

Amazon Linux

Free tier eligible

Verified provider

Amazon Linux 2023 AMI

ami-0ec0e125bb6c6e8ec (64-bit (x86), uefi-preferred) / ami-09b459d1498a0c8e2 (64-bit (Arm), uefi)

Amazon Linux 2023 is a modern, general purpose Linux-based OS that comes with 5 years of long term support. It is optimized for AWS and designed to provide a secure, stable and high-performance execution environment to develop and run your cloud applications.

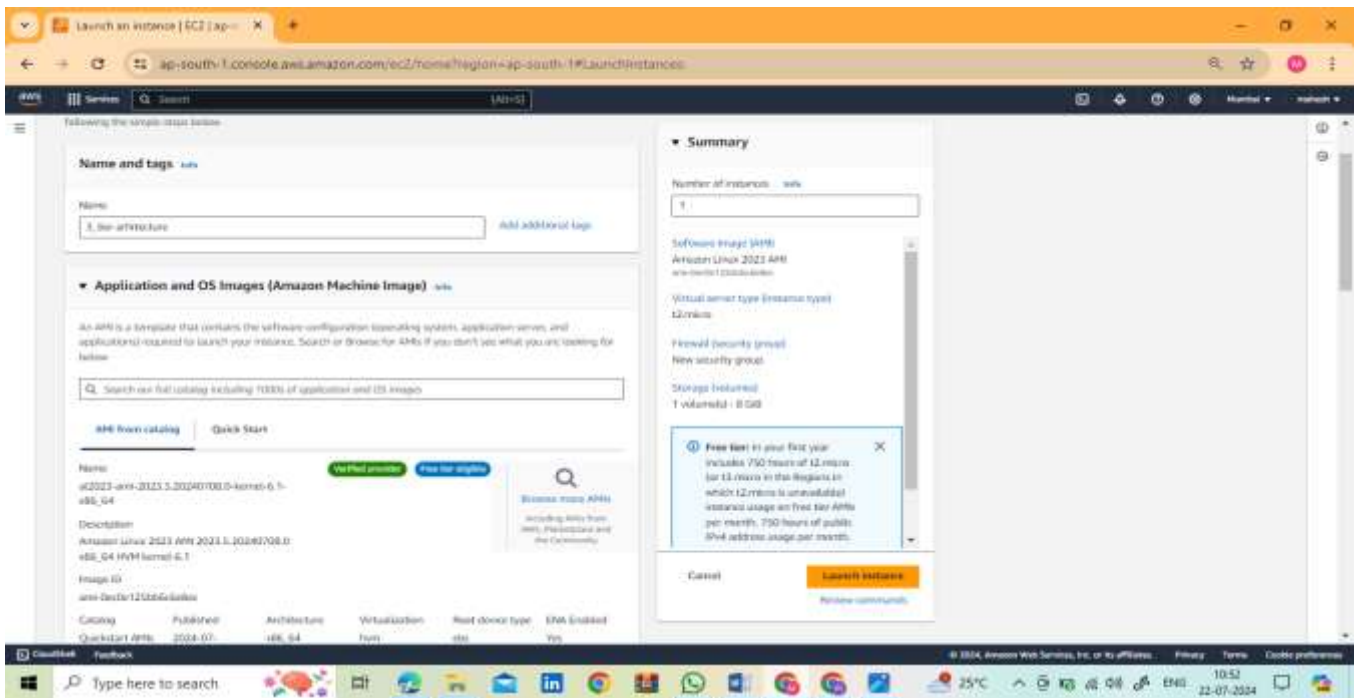
Platform: amazon Root device type: ebs Virtualization: hvm ENA enabled: Yes

Select

64-bit (x86), uefi-preferred

64-bit (Arm), uefi

Give some name for your instance, I will as 3_tier architecture



Choose the key pair you will use, and ensure that you select your new VPC and the appropriate subnet. Make sure Auto-assign IP is enabled.

Launch an instance | EC2 | ap-south-1

RouteTables | VPC Console

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#LaunchInstances

Instance type Info | Get advice

Instance type

t2.micro Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true

On-Demand Linux base pricing: 0.0124 USD per Hour

On-Demand Windows base pricing: 0.017 USD per Hour

On-Demand RHEL base pricing: 0.0268 USD per Hour

On-Demand SUSE base pricing: 0.0124 USD per Hour

Additional costs apply for AMIs with pre-installed software

All generations

Compare instance types

Key pair (login) Info

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name - required

CR-04-24

Create new key pair

Network settings Info

VPC - required Info

vpc-08e848fad6b843853 (project-vpc) 10.0.0.0/16

Subnet Info

subnet-02c10554511c11812 project-subnet-public1-ap-south-1a

VPC: vpc-08e848fad6b843853 Owner: 767398151945

Availability Zone: ap-south-1a IP addresses available: 4090 CIDR: 10.0.0.0/20

Create new subnet

Auto-assign public IP Info

Enable

Create a new security group. For inbound security group rules, add rules for SSH, HTTP, and HTTPS from any source. Although this configuration is not standard or secure, it is acceptable for the purposes of this demonstration.

Firewall (security groups) | Info

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☒ Create security group ☐ Select existing security group

Security group name - required

project-web-tier-SecurityGroup

This security group will be added to all network interfaces. The name can't be edited after the security group is created. Max length is 255 characters. Valid characters: a-z, A-Z, 0-9, spaces, and _/[]!@#\$%^&*~`|'";:.,-+=~

Description - required | Info

project-web-tier-SG

Inbound Security Group Rules

▼ Security group rule 1 (TCP: 22, 0.0.0.0/0) Remove

Type	Protocol	Port range	Source type	Source	Description - optional
ssh	TCP	22	Anywhere	0.0.0.0/0	e.g. SSH for admin desktop

▼ Security group rule 2 (TCP: 80, 0.0.0.0/0) Remove

Type	Protocol	Port range	Source type	Source	Description - optional
HTTP	TCP	80	Anywhere	0.0.0.0/0	e.g. SSH for admin desktop

▼ Security group rule 3 (TCP: 443, 0.0.0.0/0) Remove

Type	Protocol	Port range	Source type	Source	Description - optional
HTTPS	TCP	443	Anywhere	0.0.0.0/0	e.g. SSH for admin desktop

Leave the configuration storage settings unchanged. In the Advanced details section, scroll all the way to the bottom. We will use a script to launch an Apache web server when the instance starts.

Launch an instance | EC2 | ap-south-1 | RouteTables | VPC Console

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#LaunchInstances

Services | Search | [Alt+S]

Summary

Number of instances | Info

1

Software image (AMI)

Amazon Linux 2023 AMI
ami-0c0e125b1e1e1e1e1e

Virtual server type (instance type)

t2.micro

Firewall (security group)

New security group

Storage (volumes)

1 volume(s) - 8 GiB

Cancel Launch instance Review commands

User data - optional | Info

Upload a file with your user data or enter it in the field.

Choose file

```
#!/bin/bash
yum update -y
yum install -y httpd
systemctl start httpd
systemctl enable httpd

cd /var/www/html

sudo echo "<h1> My project Website</h1>index.html"
```

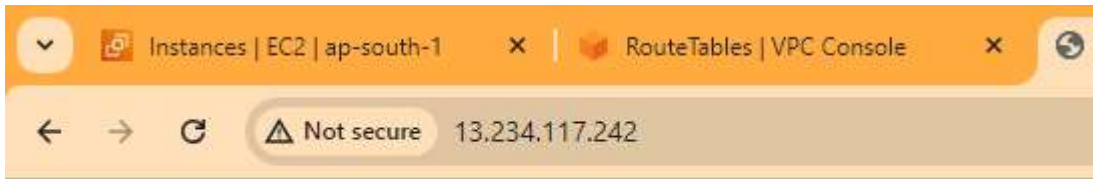
☐ User data has already been base64 encoded

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

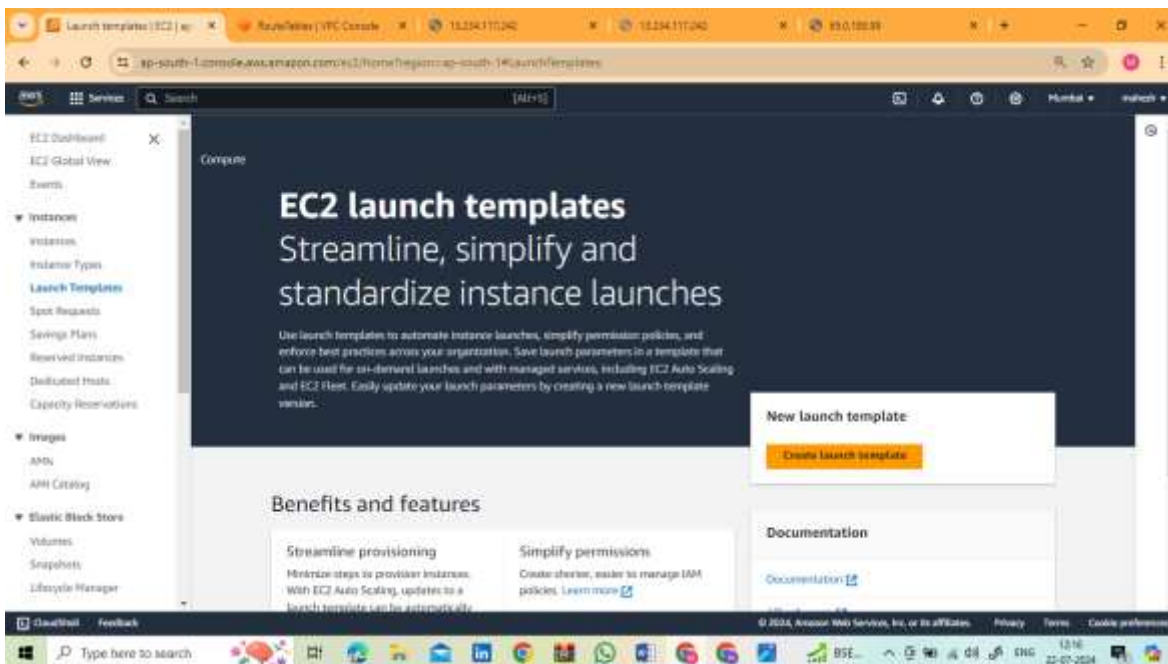
11:27 22-07-2024

Launch your new instance! Once your instance is up and running, copy the public IP address and paste it into a web browser.



It works!

For this project to work, we need to create an auto-scaling group and attach it to our EC2 instance. This will enhance our reliability and availability. Next, we are going to create a launch template. Before proceeding, we need to define the launch template, which will specify the resources to be allocated when the auto-scaling group launches on-demand instances. Under the EC2 dashboard, select "Launch Templates" and click the "Create launch template" button.



Give your template a name and select the option to "provide guidance."

Create launch template

Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched at a later time. Templates can have multiple versions.

Launch template name and description

Launch template name - *required*

Must be unique to this account. Max 120 chars. No spaces or special characters like '&', '@', '!'.

Template version description

Max 255 chars

Auto Scaling guidance [Info](#)

Select this if you intend to use this template with EC2 Auto Scaling

☒ Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

► Template tags

► Source template

Use our recently launched AMI, select the t2.micro instance type, and choose your key pair.

Recents | Quick Start

☐ Recently launched

☒ **Currently in use**

[Browse more AMIs](#)
Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

al2023-ami-2023.5.20240708.0-kernel-6.1-x86_64

ami-0ec0e125bb6c6e8ec
2024-07-05T19:47:59.000Z Architecture: x86_64 Virtualization: hvm ENA enabled: true Root device type: ebs
Root mode: user-preferred

Description

Amazon Linux 2023 AMI 2023.5.20240708.0 x86_64 HVM kernel-6.1

Architecture	AMI ID	
x86_64	ami-0ec0e125bb6c6e8ec	Verified provider

▼ **Instance type** [Info](#) [Get advice](#) [Advanced](#)

Instance type

t2.micro

Family: t2 1 vCPU 1 GiB Memory Current generations: true Free tier eligible

On-Demand Linux base pricing: 0.0126 USD per Hour

On-Demand Windows base pricing: 0.017 USD per Hour

On-Demand RHEL base pricing: 0.0288 USD per Hour

On-Demand SUSE base pricing: 0.0126 USD per Hour

Additional costs apply for AMIs with pre-installed software

☐ All generations [Compare instance types](#)

▼ **Key pair (login)** [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name

[Create new key pair](#)

For the firewall settings, select "Choose an existing security group" and ensure the security group (SG) we created for the web tier is selected. In the Advanced network configuration section, enable Auto-assign public IP.

▼ Network settings [Info](#)

Subnet [Info](#)

Don't include in launch template [Create new subnet](#)

When you specify a subnet, a network interface is automatically added to your template.

Firewall (security groups) [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☒ Select existing security group ☐ Create security group

Common security groups [Info](#)

Select security groups

project-web-tier-SecurityGroup sg-0856d3b0fc037cdf1 [X](#)
VPC: vpc-08e848fad6b843853

Compare security group rules

Security groups that you add or remove here will be added to or removed from all of your network interfaces.

▼ Advanced network configuration

Network interface 1 [Remove](#)

Device index [Info](#)

0

Network interface [Info](#)

New interface

Existing network interfaces are not recommended when creating a template for auto-scaling.

Description [Info](#)

Subnet [Info](#)

Don't include in launch template
Not applicable for EC2 Auto Scaling

Security groups [Info](#)

Select security groups

Show all selected (1)

Auto-assign public IP [Info](#)

Enable

We will leave the storage options unchanged for now. Click on the Advanced details tab, scroll down, and enter the same script we used earlier for our EC2 instance.

Click the "Create launch template" button.

Create launch template | EC2

RouteTables | VPC Console

13.234.117.242

13.234.117.242

63.0.180.09

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#CreateLaunchTemplate

Don't include in launch template

User data - optional [Info](#)

Upload a file with your user data or enter it in the field.

[Choose file](#)

```
#!/bin/bash
yum update -y
yum install -y httpd
systemctl start httpd
systemctl enable httpd

cd /var/www/html

sudo echo "<h1>My Project Website</h1>"index.html
```

☐ User data has already been base64 encoded

Summary

Software image (AMI)

Amazon Linux 2023 AMI 2023.5.2...read more
ami-0a0c125bb6c8a80c

Virtual server type (instance type)

t2.micro

Firewall (security group)

project-web-tier-SecurityGroup

Storage (volumes)

1 volume(s) - 8 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance

[Cancel](#) [Create launch template](#)

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Type here to search

26°C

ENG

12:41

22-07-2024

Navigate to the Auto Scaling tab at the bottom of the EC2 dashboard. Click “Create auto scaling group.” The launch template we just created will be used by the auto-scaling group to launch new EC2 instances when scaling up.

Name your auto-scaling group (ASG), choose the launch template you created, and then click the Next button.

The screenshot shows the AWS Management Console interface for creating an Auto Scaling group. The browser tabs at the top include 'Create Auto Scaling group', 'RouteTables | VPC Console', and two IP addresses. The address bar shows the URL: `ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#CreateAutoScalingGroup`. The console header shows the AWS logo, 'Services', a search bar, and '[Alt+S]'. The left sidebar shows a navigation menu with 'EC2' selected, and a sub-menu with 'Auto Scaling groups' and 'Create Auto Scaling group'. The main content area is titled 'Choose launch template' with a subtitle 'Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group'. The page is divided into two main sections. The top section, 'Name', contains a text input field for 'Auto Scaling group name' with the placeholder text 'Enter a name to identify the group.' and the value 'My-project-web-server-auto-scaling-group'. Below the input field is a note: 'Must be unique to this account in the current Region and no more than 255 characters.' The bottom section, 'Launch template', contains a dropdown menu for 'Launch template' with the value 'Project-web-tier-template' and a 'Create a launch template' link. Below this is a 'Version' dropdown menu with the value 'Default (1)' and a 'Create a launch template version' link. A blue box highlights the 'Name' section, and another blue box highlights the 'Launch template' section.

Under Network, ensure you select the VPC you created earlier. Additionally, under Availability Zones and Subnets, select the public subnets that were created; your selections may vary.

Click the Next button.

Network Info

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC
Choose the VPC that defines the virtual network for your Auto Scaling group.

vpc-08e848fad6b843853 (project-vpc)
10.0.0.0/16

[Create a VPC](#)

Availability Zones and subnets
Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC.

Select Availability Zones and subnets

ap-south-1b | subnet-0f2ac59e229daccac (project-subnet-public2-ap-south-1b)
10.0.16.0/20

ap-south-1a | subnet-02c10554511c11812 (project-subnet-public1-ap-south-1a)
10.0.0.0/20

[Create a subnet](#)

Cancel Skip to review Previous **Next**

Now we have the option to allocate a load balancer for our ASG. A load balancer will distribute incoming traffic across multiple servers, which helps improve availability and performance.

Select "Attach to a new load balancer" and choose "Application Load Balancer." Assign a name to your load balancer, and specify it as "Internet facing" since it is intended for our web tier.

Configure advanced options - *optional* info

Integrate your Auto Scaling group with other services to distribute network traffic across multiple servers using a load balancer or to establish service-to-service communications using VPC Lattice. You can also set options that give you more control over health check replacements and monitoring.

Load balancing info

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

☐ No load balancer

Traffic to your Auto Scaling group will not be fronted by a load balancer.

☐ Attach to an existing load balancer

Choose from your existing load balancers.

☒ Attach to a new load balancer

Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to a new load balancer

Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type

Load balancer type cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, visit the [Load Balancing console](#).

☒ Application Load Balancer

HTTP, HTTPS

☐ Network Load Balancer

TCP, UDP, TLS

Load balancer name

Names can be changed after the load balancer is created.

My-project web server auto scaling group-1

Load balancer scheme

Scheme cannot be changed after the load balancer is created.

☐ Internal

☒ Internet-facing

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC

vpc-08e848fad6b843653

project-vpc

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS.

ap-south-1b

subnet-0f2ac59e229daccaf

ap-south-1a

subnet-02c10554511c11812

Your VPC and the two public subnets should already be selected. In the "Listeners and routing" section, select "Create a target group," ensuring it is set to port 80 for HTTP traffic.

Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) after your load balancer is created.

Protocol

HTTP

Port

80

Default routing (forward to)

Create a target group

New target group name

An instance target group with default settings will be created.

My-project web server auto scaling group-1

Taas - *optional*

Ensure that the service for Leave No VPC Lattice is activated. Click to enable Elastic Load Balancing health checks.

VPC Lattice integration options [Info](#)

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ No VPC Lattice service

VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

[Create new VPC Lattice service](#)

☐ Attach to VPC Lattice service

Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

☒ Always enabled

Additional health check types - optional [Info](#)

☒ Turn on Elastic Load Balancing health checks **Recommended**

Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

EC2 Auto Scaling will start to detect and act on health checks performed by Elastic Load Balancing. To avoid unexpected terminations, first verify the settings of these health checks in the [Load Balancer console](#).

☐ Turn on VPC Lattice health checks

VPC Lattice can monitor whether instances are available to handle requests. If it considers a target as failed a health check, EC2 Auto Scaling replaces it after its next periodic check.

Health check grace period [Info](#)

This time period delays the first health check until your instances finish initializing. It doesn't prevent an instance from terminating when placed into a non-running state.

seconds

Verify that "Enable Group Metrics Collection within Cloud Watch" is checked.

Additional settings

Monitoring [Info](#)

☒ Enable group metrics collection within CloudWatch

Default instance warmup [Info](#)

The amount of time that CloudWatch metrics for new instances do not contribute to the group's aggregated instance metrics, as their usage data is not reliable yet.

☐ Enable default instance warmup

Cancel

Skip to review

Previous

Next

Next, we configure the group size and scaling policy for our Auto Scaling Group (ASG). To ensure reliability and performance, set the desired capacity and minimum capacity to 2, and the maximum capacity to 4.

Configure group size and scaling - *optional* [Info](#)

Define your group's desired capacity and scaling limits. You can optionally add automatic scaling to adjust the size of your group.

Group size [Info](#)

Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.

Desired capacity type

Choose the unit of measurement for the desired capacity value. vCPUs and Memory(GiB) are only supported for mixed instances groups configured with a set of instance attributes.

Units (number of instances) ▼

Desired capacity

Specify your group size.

2

Scaling [Info](#)

You can resize your Auto Scaling group manually or automatically to meet changes in demand.

Scaling limits

Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity

2

Equal or less than desired capacity

Max desired capacity

4

Equal or greater than desired capacity

Automatic scaling - *optional*

Choose whether to use a target tracking policy. [Info](#)

You can set up other metric-based scaling policies and scheduled scaling after creating your Auto Scaling group.

☐ No scaling policies

Your Auto Scaling group will remain at its initial size and will not dynamically resize to meet demand.

☒ Target tracking scaling policy

Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Scaling policy name

Target Tracking Policy

Metric type [Info](#)

Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.

Average CPU utilization ▼

Target value

50

Instance warmup [Info](#)

300

seconds

In the scaling policies section, select "Target tracking scaling policy." Set the metric type to "Average CPU Utilization" with a target value of 50.

Then, click the "Next" button.

On the next screen, you have the option to add notifications through SNS topics. However, I have skipped this step for now. Click the "Next" button.

Add notifications - *optional* [Info](#)

Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.

Add notification

Cancel

Skip to review

Previous

Next

Add tags - optional Info

Add tags to help you search, filter, and track your Auto Scaling group across AWS. You can also choose to automatically add these tags to instances when they are launched.

ⓘ You can optionally choose to add tags to instances (and their attached EBS volumes) by specifying tags in your launch template. We recommend caution, however, because the tag values for instances from your launch template will be overridden if there are any duplicate keys specified for the Auto Scaling group.

Tags (0)

Add tag
50 remaining

Cancel
Previous
Next

Review your settings on the next page, and at the bottom, click the "Create Auto Scaling Group" button. You should see a green banner indicating your success. Once the ASG has finished updating capacity, navigate to your EC2 dashboard to verify that the new instance has been created.

Note: In my previous examples, my names were not accepted for the auto scaling group. Ensure that you follow the appropriate naming conventions.

EC2 > Auto Scaling groups

Auto Scaling groups (1/1) Info

< 1 > ⚙

<input checked="" type="checkbox"/>	Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max
<input checked="" type="checkbox"/>	proj-web-svr-auto-scale-grp	Project-web-tier-template Version Defac	2	-	2	2	4

As you can see in the ec2 console it will show 2 more desired instaness, the Auto Scaling Group (ASG) is performing as expected.

Successfully initiated starting of i-0dc5a854a74b87854

Notifications 🔔 0 ⚠ 0 🔄 2 📄 0 📄 0

Instances (1/3) Info

All states ▼

Instance state: running
Clear filters

<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 D
<input type="checkbox"/>		i-0d7ac5ef7170b05eb	Running	t2.micro	2/2 checks passed	View alarms +	ap-south-1b	-
<input checked="" type="checkbox"/>	3_tier architect...	i-0dc5a854a74b87854	Running	t2.micro	2/2 checks passed	View alarms +	ap-south-1a	-
<input type="checkbox"/>		i-0c6f840fddaa5a0	Running	t2.micro	2/2 checks passed	View alarms +	ap-south-1a	-

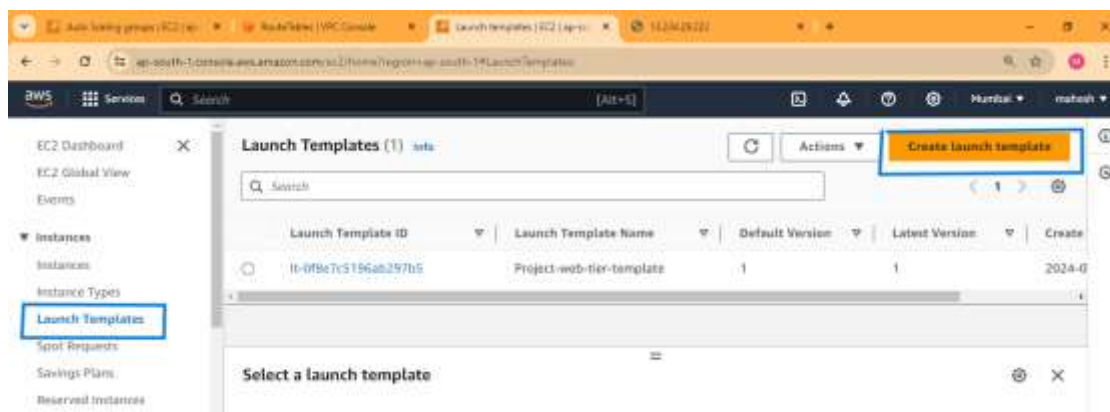
Before we proceed, it's a good idea to take a moment to connect to the instances that were created. As you can see, everything has been successful!



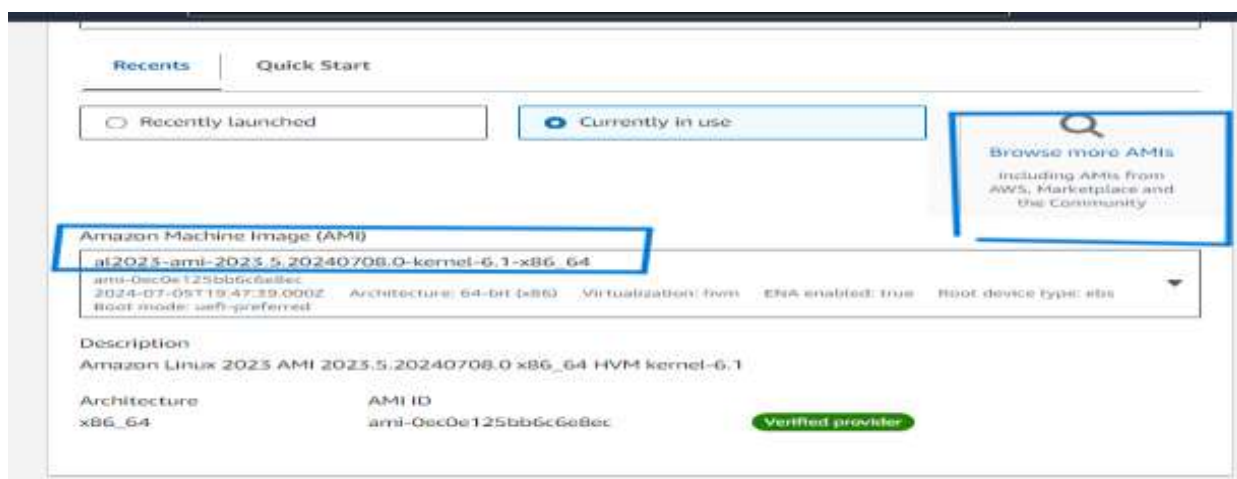
It works!

Step-3 : Creating an Application Tier

Next, we will create the back-end of our 3-tier architecture. Instead of starting with an EC2 instance, I'll begin by navigating to the Launch Templates tab under the EC2 dashboard.



Name your new template and select the "Guidance" tab again.



Select "Browse more AMIs," then choose Amazon Linux 2 for your AMI. Select the `t2.micro` instance type and also choose your key pair.

▼ **Instance type** [Info](#) | [Get advice](#) Advanced

Instance type

t2.micro Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true

On-Demand Linux base pricing: 0.0124 USD per Hour

On-Demand Windows base pricing: 0.017 USD per Hour

On-Demand RHEL base pricing: 0.0268 USD per Hour

On-Demand SUSE base pricing: 0.0124 USD per Hour

Additional costs apply for AMIs with pre-installed software

☐ All generations

[Compare instance types](#)

▼ **Key pair (login)** [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name

CR-04-24

[Create new key pair](#)

Under the network settings, we want to limit access to the application tier for security purposes. Ensure there is no public access to the application layer or the data tier. We will create a new security group. Select our VPC; I now realize that a better name could have been chosen for this part!

Name your new security group and select the VPC that we created at the beginning.

▼ **Network settings** [Info](#)

Subnet [Info](#)

Don't include in launch template

When you specify a subnet, a network interface is automatically added to your template.

[Create new subnet](#)

Firewall (security groups) [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☐ Select existing security group

☒ Create security group

Security group name - required

application-tier-sg

This security group will be added to all network interfaces. The name can't be edited after the security group is created. Max length is 255 characters. Valid characters: a-z, A-Z, 0-9, spaces, and _-:/()#,@[]+=&:()!\$*

Description - required [Info](#)

allows ssh access to application tier

VPC [Info](#)

vpc-08e848fad6b843853 (project-vpc)

10.0.0.0/16

[Add security group rule](#)

No security group rules are currently included in this template. Add a new rule to include it in the launch template.

We will create three security group rules:

1. For SSH Access: Use "My IP" as the source.
2. For Custom TCP: Use the security group from our web tier (tier-1) as the source.
3. For ICMP (IPv4): Set the source type to "Anywhere" to allow ping requests from the public internet. This will help us test if traffic is routed properly to the application tier.

The screenshot shows the AWS IAM console interface for configuring security group rules. Three rules are listed:

- Security group rule 1 (TCP, 0, sg-0856d3b0fc037cdf1)**: Type is Custom TCP, Protocol is TCP, Port range is 0. Source type is Custom, and the source field is set to sg-0856d3b0fc037cdf1 (highlighted with a blue box). Description is optional, set to "e.g. SSH for admin desktop".
- Security group rule 2 (TCP, 22, 0.0.0.0/0)**: Type is ssh, Protocol is TCP, Port range is 22. Source type is Anywhere, and the source field is set to 0.0.0.0/0 (highlighted with a blue box). Description is optional, set to "e.g. SSH for admin desktop".
- Security group rule 3 (ICMP, All, 0.0.0.0/0)**: Type is All ICMP - IPv4, Protocol is ICMP, Port range is All. Source type is Anywhere, and the source field is set to 0.0.0.0/0 (highlighted with a blue box). Description is optional, set to "e.g. SSH for admin desktop".

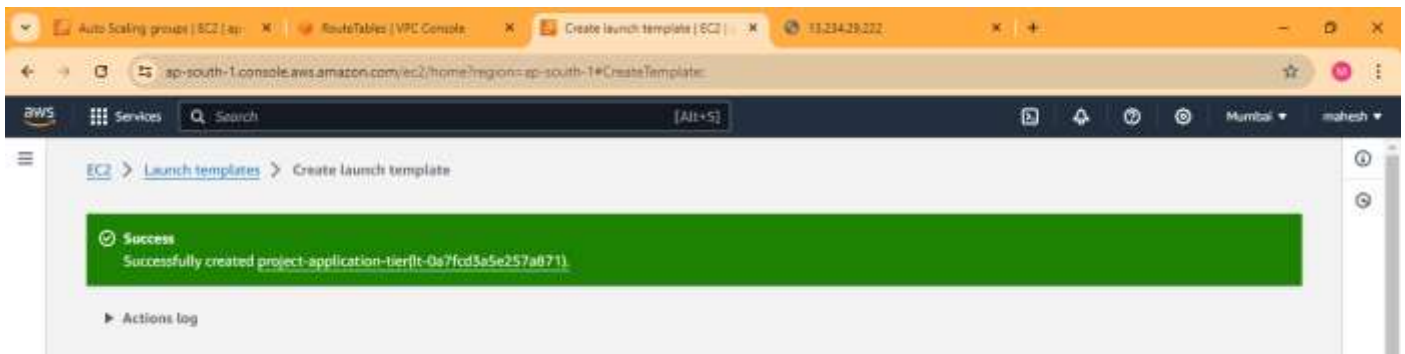
We will leave the storage volumes as they are and proceed to the bottom of the "Advanced details" section to enter our script. After entering the script, click "Next."

```
#!/bin/bash
yum update -y httpd
yum install -y httpd
systemctl start httpd
systemctl enable httpd

cd/var/www/html

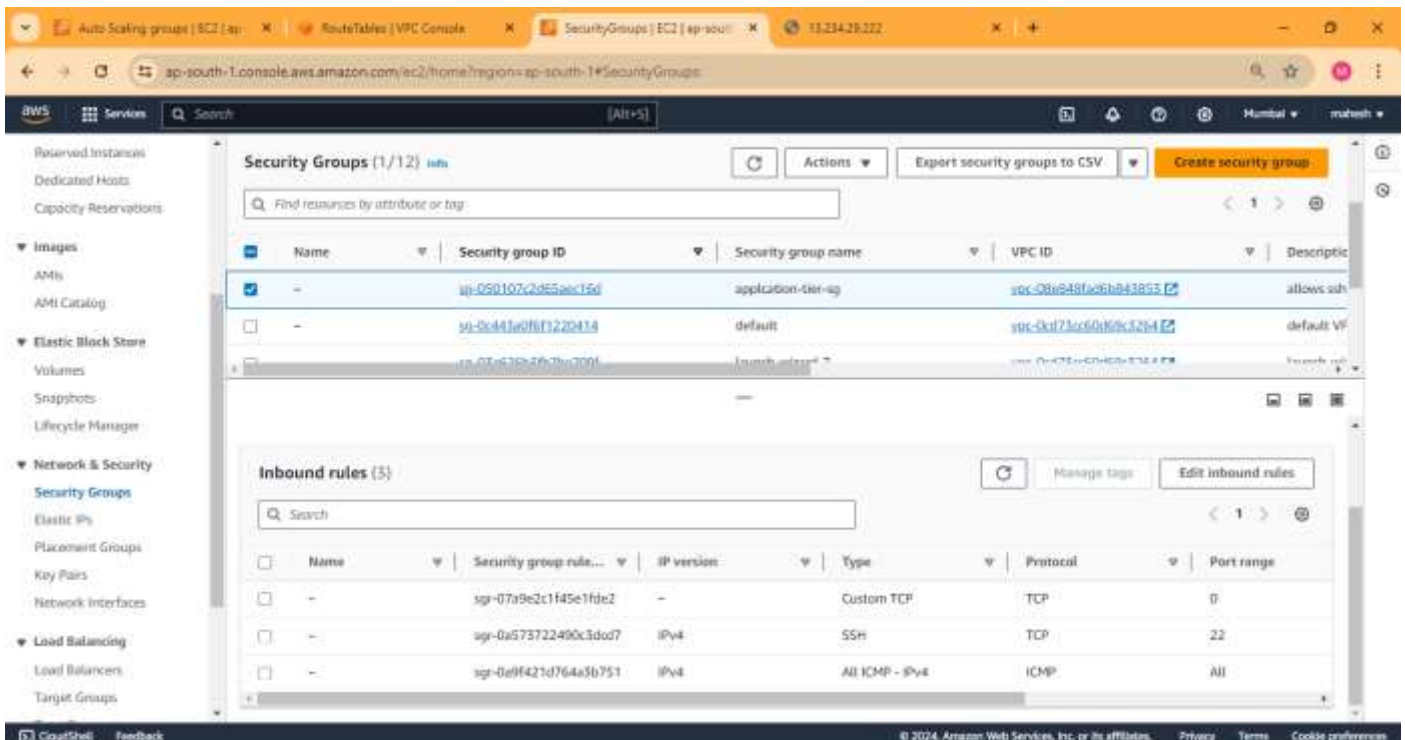
sudo echo "<h1>my project website</h1>"index.html
```

☐ User data has already been base64 encoded



Starting from scratch can be a learning experience, but modifying an existing template would indeed save time in the future. It's all part of the learning process, and finding efficient ways to manage your resources is valuable!

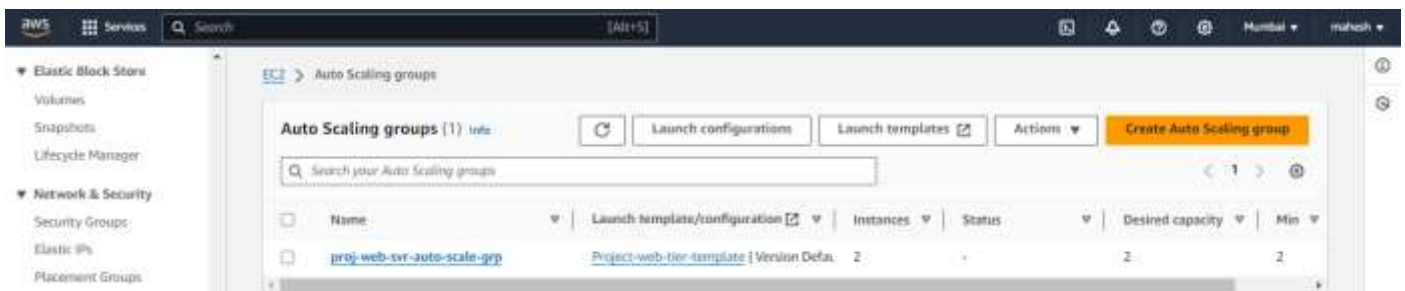
I went back to the security group and updated my inbound rules.



Once this was fixed, I went back and recreated the application layer template using the Application Tier SG1 that I just altered. I just double checked the security group rules and used the same settings for everything else above.

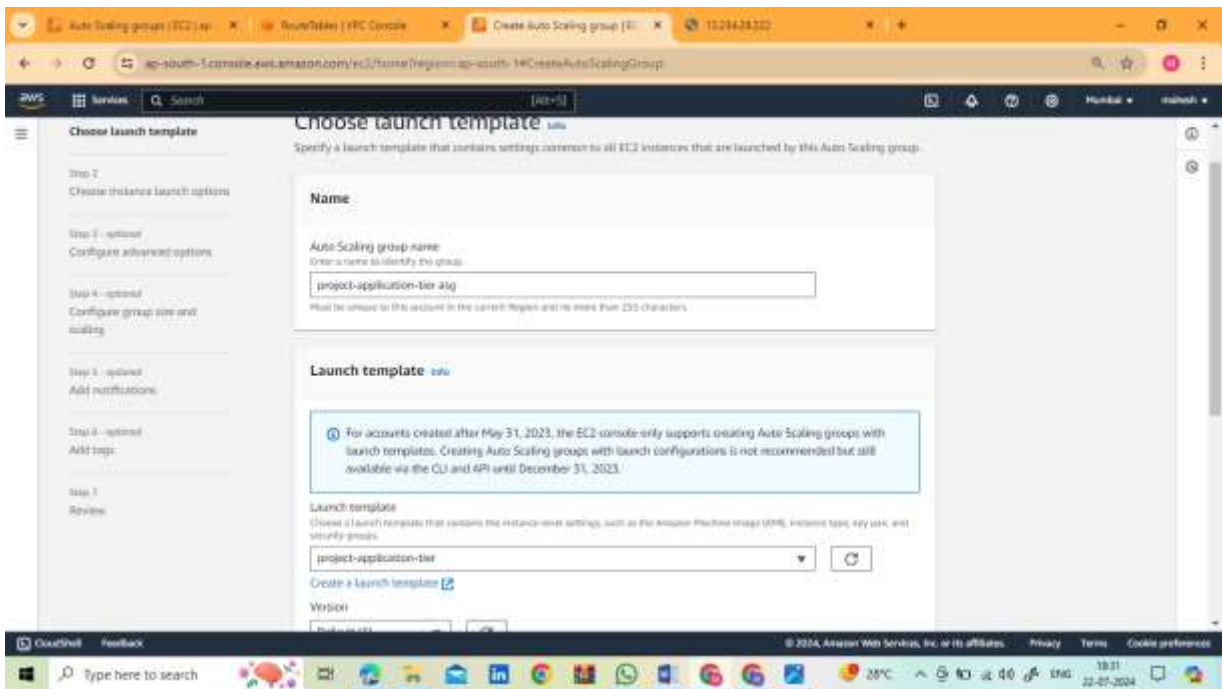
Application Tier Auto-Scaling Group:

Okay, now we're ready to create our Auto Scaling Group for the application layer. Under the EC2 dashboard, go to "Create an Auto Scaling Group."

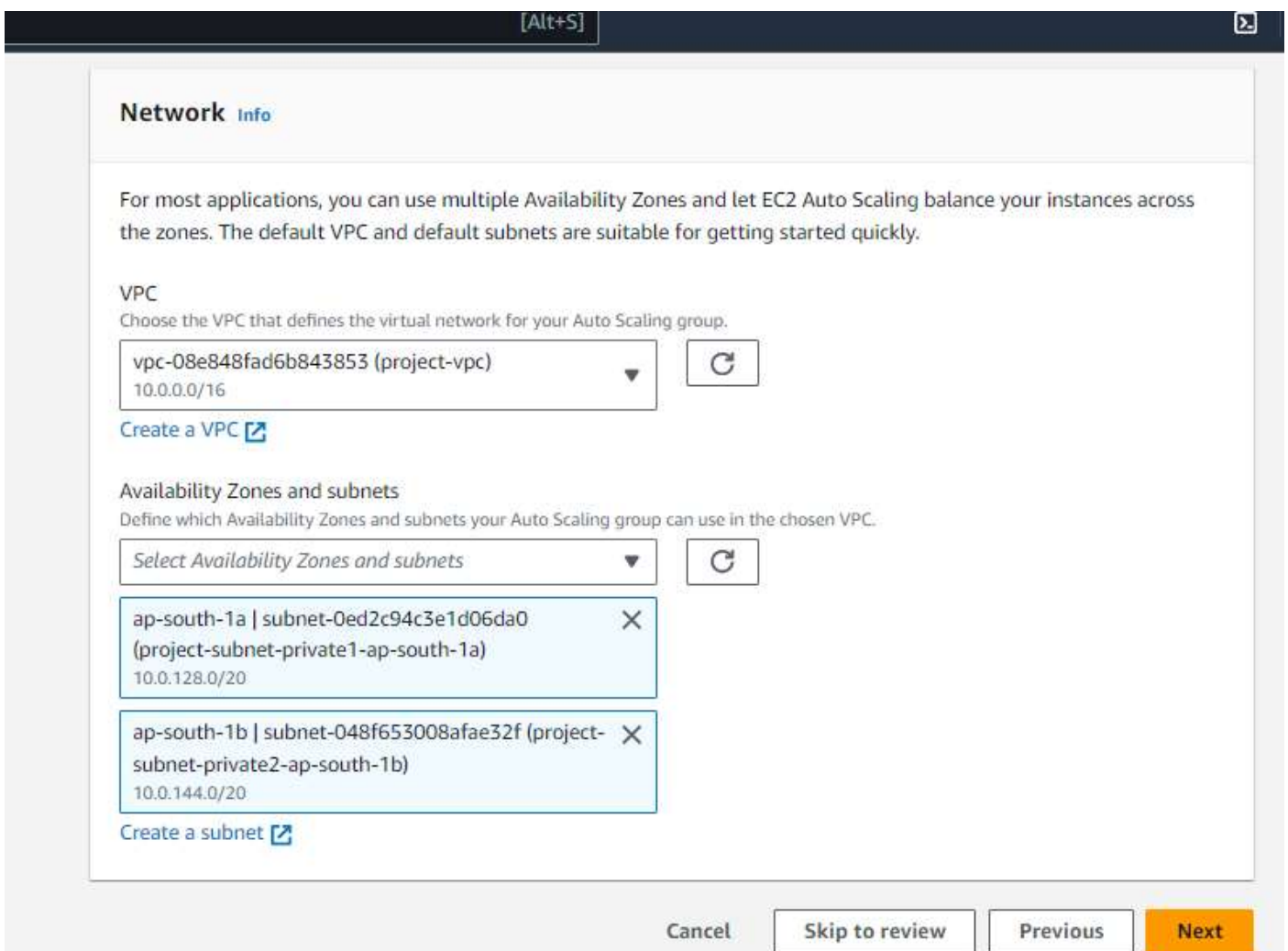


Name your new Auto Scaling Group and select the appropriate launch template, then click the "Next" button.

Choose the correct VPC and select 2 private subnets, then click the "Next" button.



We are given the option to attach a load balancer, and we want to do this. Select an application load balancer, name it, and configure it as an internal load balancer. Double-check that the VPC and subnets are correct. My settings are accurate.



RouteTables | VPC Console

Create Auto Scaling group | EC2

13.234.29.222

sofe.aws.amazon.com/ec2/home?region=ap-south-1#CreateAutoScalingGroup:

[Alt+F5]

Attach to a new load balancer

Define a new load balancer to create for attachment for this Auto Scaling group.

Load balancer type

Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, visit the [Load Balancing console](#).

☒ Application Load Balancer
HTTP, HTTPS

☐ Network Load Balancer
TCP, UDP, TLS

Load balancer name

Name cannot be changed after the load balancer is created.

project-application-tier-asg-1

Load balancer scheme

Scheme cannot be changed after the load balancer is created.

☒ Internal

☐ Internet-facing

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group; you can select different subnets and add subnets from additional Availability Zones.

VPC

vpc-04b6402af2b643053 project-vpc

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS resolution.

ap-south-1a subnet-04b165300bafac32f

ap-south-1b subnet-0ed2c94c3c1d066da0

Under "Listeners and routing," create a new target group and set the port to 80 once again.

Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) after your load balancer is created.

Protocol Port Default routing (forward to)

HTTP 80 Create a target group

New target group name

An instance target group with default settings will be created.

project-application-tier-asg-1

Tags - optional

Consider adding tags to your load balancer. Tags enable you to categorize your AWS resources so you can more easily manage them.

Below, I have chosen to turn on health checks and enable group metrics within Cloud-Watch.

VPC Lattice integration options Info

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ No VPC Lattice service

VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

[Create new VPC Lattice service](#)

☐ Attach to VPC Lattice service

Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

☒ Always enabled

Additional health check types - optional Info

☒ Turn on Elastic Load Balancing health checks Recommended

Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

☒ EC2 Auto Scaling will start to detect and act on health checks performed by Elastic Load Balancing. To avoid unexpected terminations, first verify the settings of these health checks in the [Load Balancer console](#)

☐ Turn on VPC Lattice health checks

VPC Lattice can monitor whether instances are available to handle requests. If it considers a target as failed a health check, EC2 Auto Scaling replaces it after its next periodic check.

Health check grace period Info

This time period delays the first health check until your instances finish initializing. It doesn't prevent an instance from terminating when placed into a non-running state.

300 seconds

Additional settings

Monitoring Info

☒ Enable group metrics collection within CloudWatch

Default instance warmup Info

The amount of time that CloudWatch metrics for new instances do not contribute to the group's aggregated instance metrics, as their usage data is not reliable yet.

☐ Enable default instance warmup

Cancel

Skip to review

Previous

Next

On the next screen, set your desired capacity, minimum capacity, and maximum capacity:

- Desired capacity: 2
- Minimum capacity: 2
- Maximum capacity: 4

Configure group size and scaling - optional Info

Define your group's desired capacity and scaling limits. You can optionally add automatic scaling to adjust the size of your group.

Group size Info

Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.

Desired capacity type

Choose the unit of measurement for the desired capacity value. vCPUs and Memory(GiB) are only supported for mixed instances groups configured with a set of instance attributes.

Units (number of instances)

Desired capacity

Specify your group size

2

Scaling [Info](#)
You can resize your Auto Scaling group manually or automatically to meet changes in demand.

Scaling limits
Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity <input type="text" value="2"/> Equal or less than desired capacity	Max desired capacity <input type="text" value="4"/> Equal or greater than desired capacity
--	---

Automatic scaling - optional
Choose whether to use a target tracking policy [Info](#)
You can set up other metric-based scaling policies and scheduled scaling [after creating your Auto Scaling group](#).

☐ **No scaling policies**
Your Auto Scaling group will remain at its initial size and will not dynamically resize to meet demand.

☒ **Target tracking scaling policy**
Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Scaling policy name

Metric type [Info](#)
Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.

Target value

Instance warmup [Info](#)
 seconds

Then, I selected target tracking with a CPU utilization target of 50%.

Click the "Next" button. Add notifications if desired, then add tags. Review your new Auto Scaling Group (ASG) settings and create it.

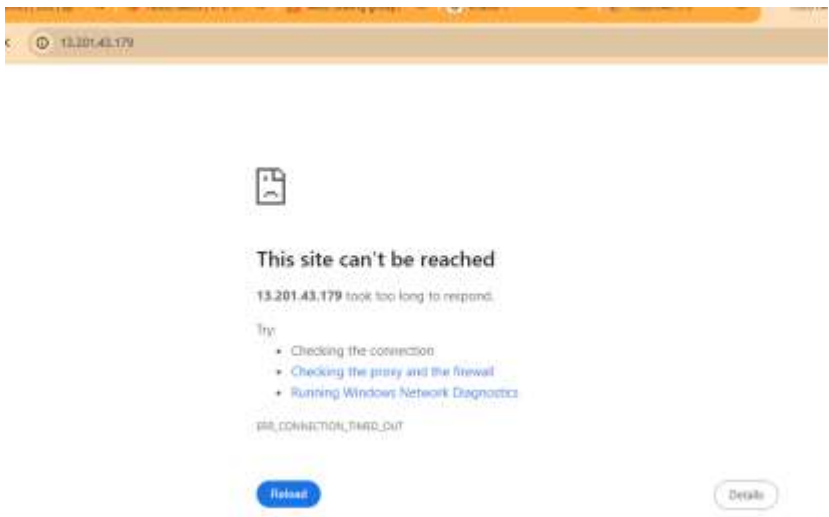
As you can see below, my new application layer ASG is updating the capacity.

project-application-tier-asg, 1 Scaling policy, 1 Load balancer, 1 Target group, 1 Listener created successfully. 1 new target group has been attached to ASG. Group metrics collection is enabled.

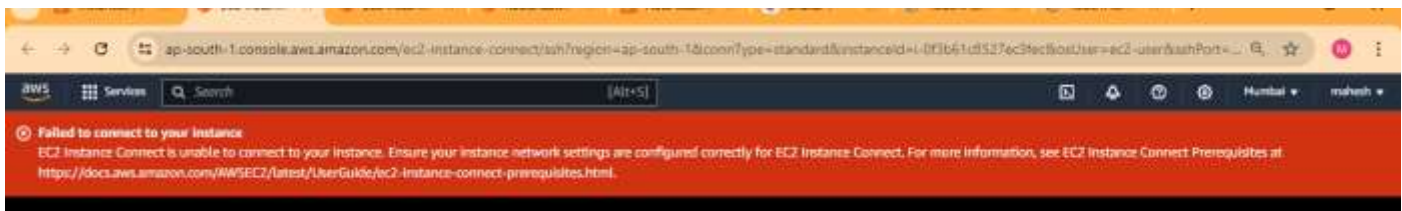
Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Avail...
project-application-tier-asg	project-application-tier Version Default	0	Updating capacity...	2	1	4	ap-out...
proj-web-tier-asg-scale-grp	Project-web-tier-termdat...	2		2	1	4	ap-out...

Once the new EC2 instances are created and running, we will attempt to SSH into them. If we set it up correctly, we should not be able to establish a connection.

When I tried to SSH into the application tier EC2 instance, the connection timed out, which is exactly the expected outcome.

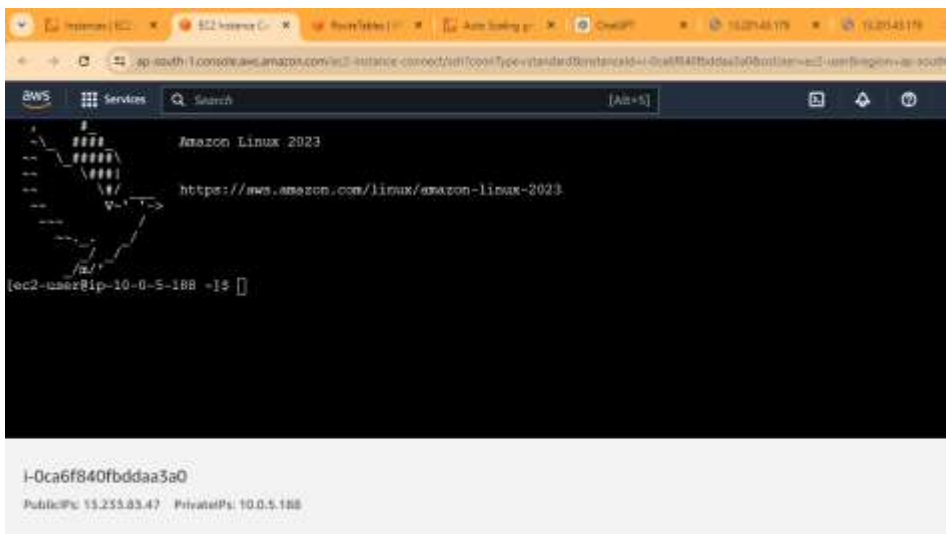


Try to connect your newly created instance using Git Bash or the AWS console. If it does not connect, it shows a timeout error.



I also tried to connect using EC2 Connect, and this failed as well.

Next, we need to check if our tier-1 servers interact with our tier-2 servers. To test this, log in to your tier-1 EC2 instances via SSH and run a ping command to the private IP address of our tier-2 servers. Below, you can see a successful ping.



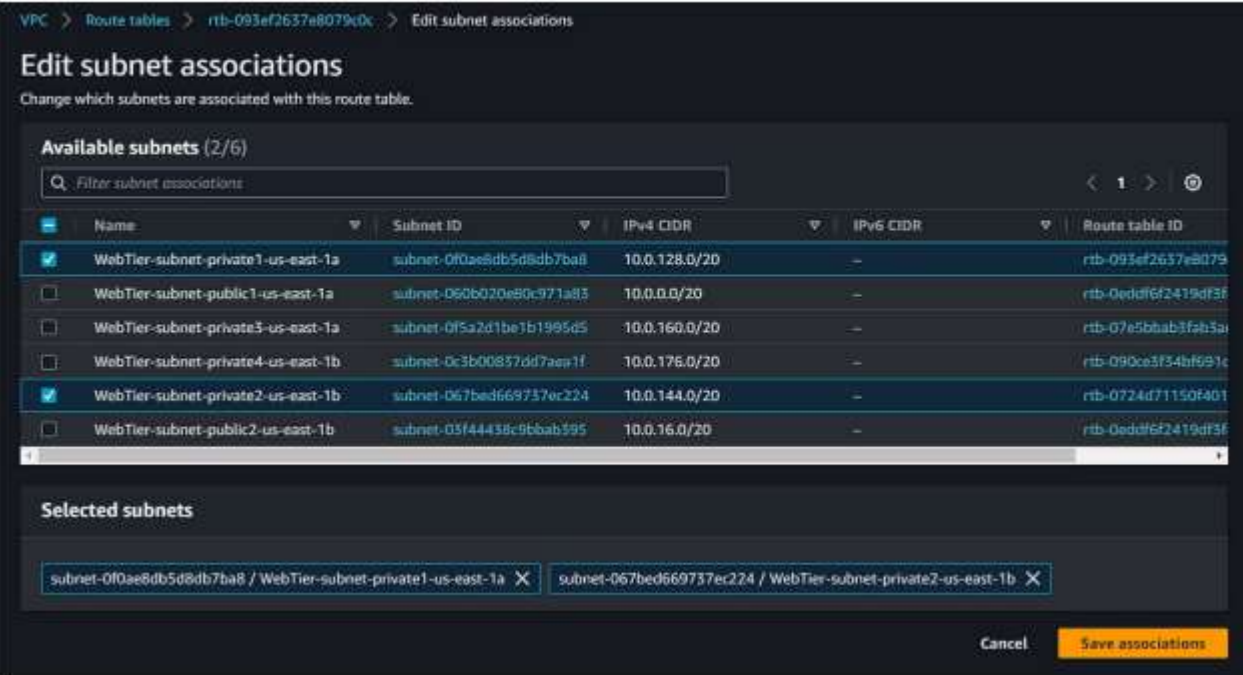
Update the application (Tier-2) Route table:

Head back to the VPC dashboard, select "Route Tables," and choose one of the route tables that was automatically created when we set up our VPC. I only have one subnet associated with this table, so click on "Edit Subnet Associations."

The screenshot displays the AWS Management Console interface for the 'Route Tables' section. The left sidebar shows the 'VPC dashboard' with a search bar and a list of VPC resources. The main content area shows a list of route tables. The 'project-rtb-private1-ap-south-1a' route table is selected, and its details are shown. The 'Subnet associations' tab is active, displaying a table of explicit subnet associations. The table has columns for Name, Subnet ID, IPv4 CIDR, and IPv6 CIDR. One association is listed for 'project-subnet-private1-ap-south-1a' with subnet ID 'subnet-0ed2c94c3e1d06da0' and IPv4 CIDR '10.0.128.0/20'. The 'Edit subnet associations' button is highlighted in the top right corner of the association table.

Name	Subnet ID	IPv4 CIDR	IPv6 CIDR
project-subnet-private1-ap-south-1a	subnet-0ed2c94c3e1d06da0	10.0.128.0/20	-

Add another subnet that is private.

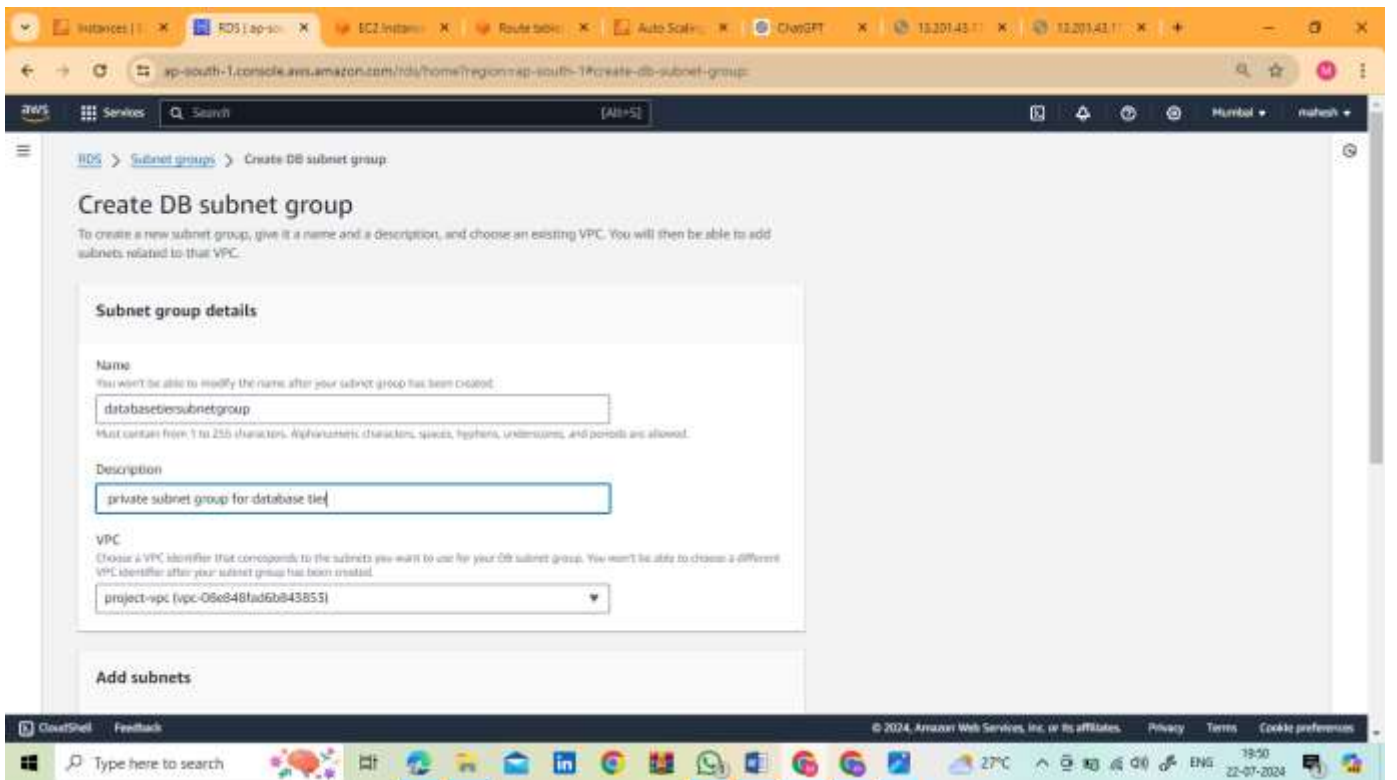


Part 4: Creating a Database Tier:

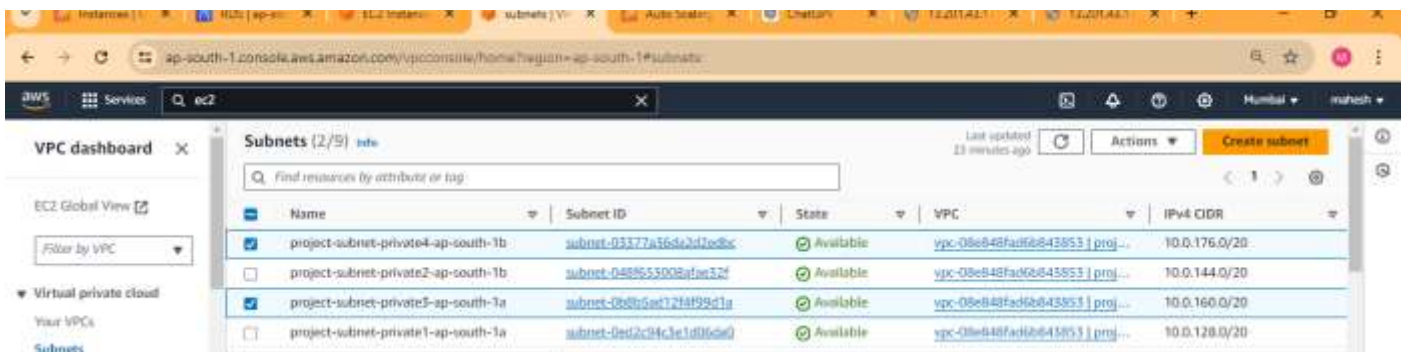
Almost there! We have created and tested 2 out of the 3 tiers successfully. We are now going to build our database tier. For this exercise, we will use a MySQL RDS database, though AWS offers several types of databases.

Create a DB Subnet Group:

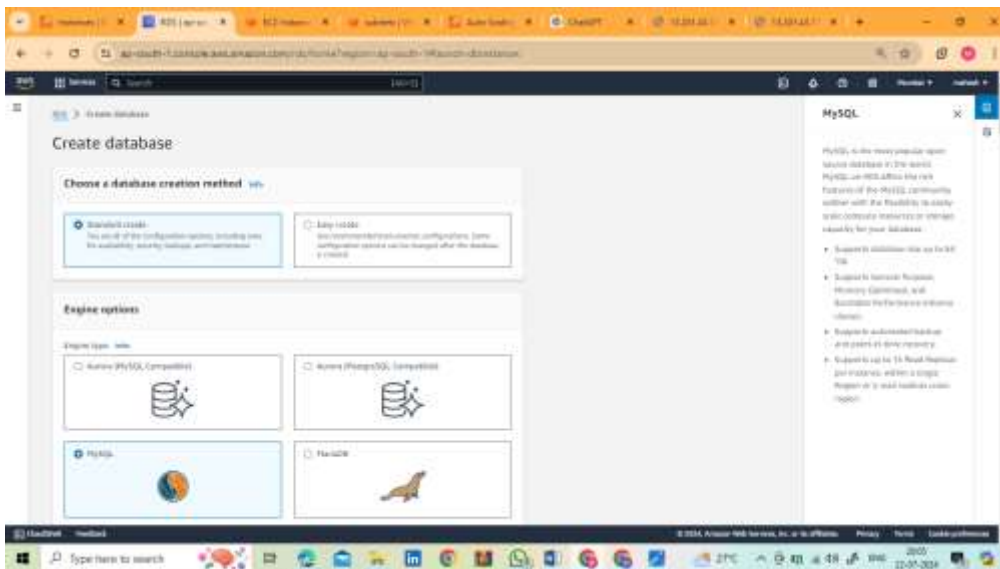
We will begin by creating a subnet group. Navigate to the RDS console, click on "Subnet Groups" in the left-side menu, and then click the orange "Create DB Subnet Group" button.



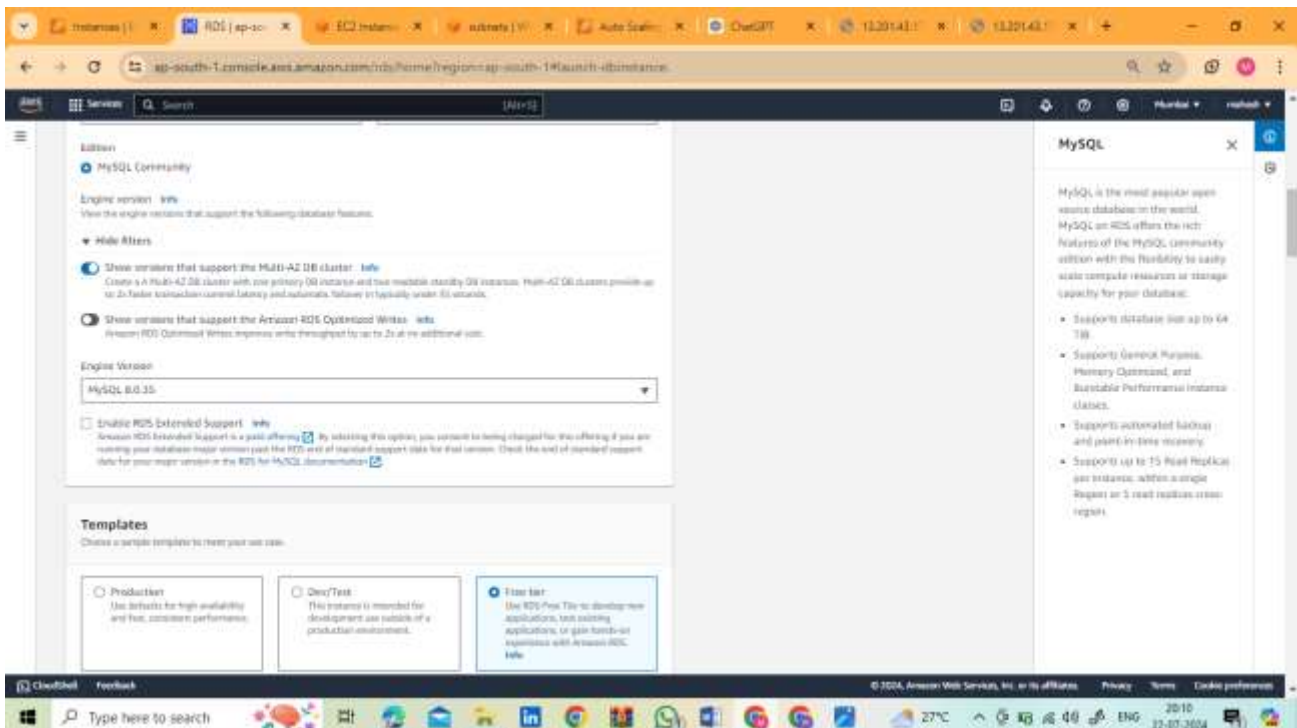
For the next part, we need to know the availability zones for the last two subnets that were automatically created. Head back to the VPC console. Under "Subnets," locate the last two subnets you have, ensuring you do not select the private subnets already used in tier 2.



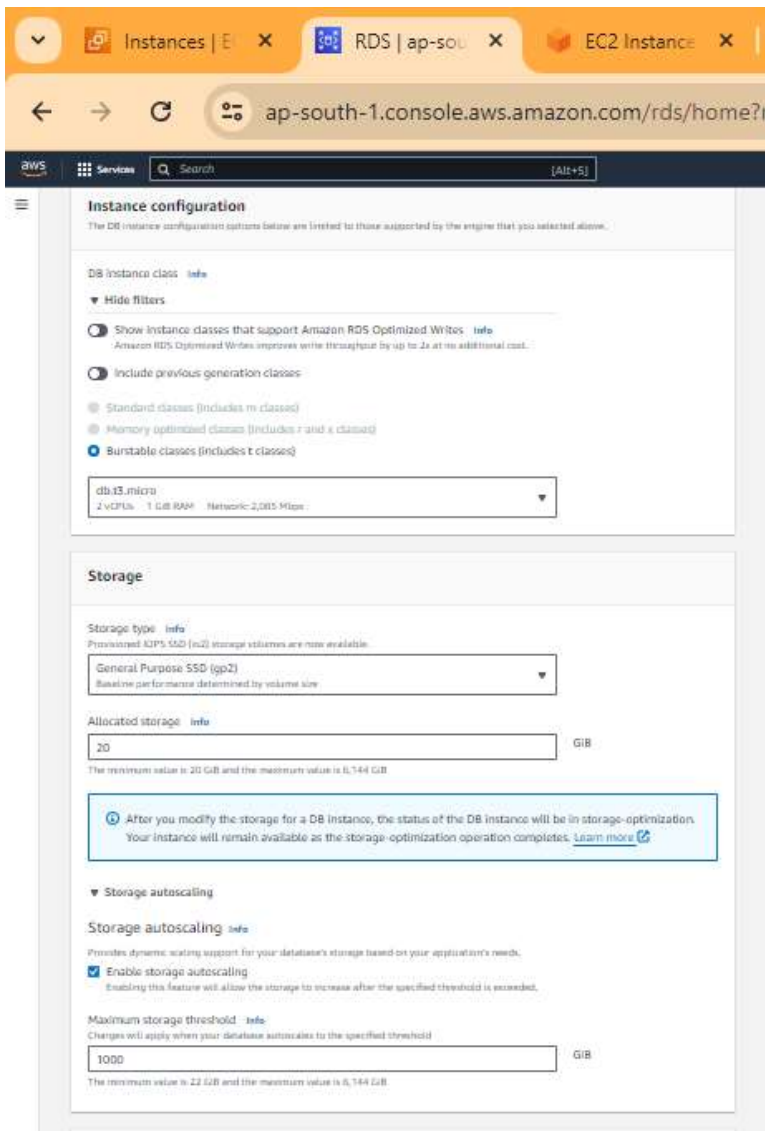
Back at the RDS console, select the availability zones that you are going to use.



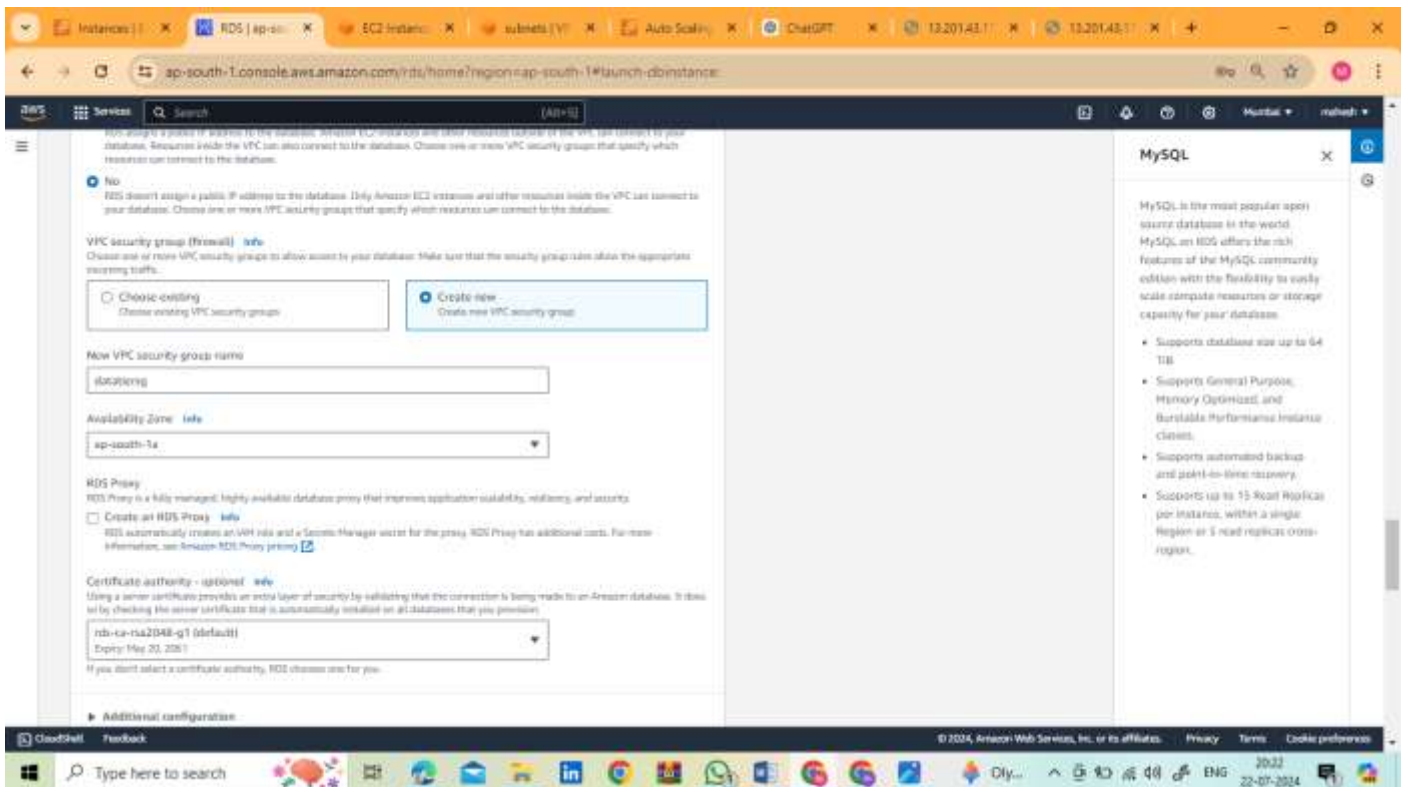
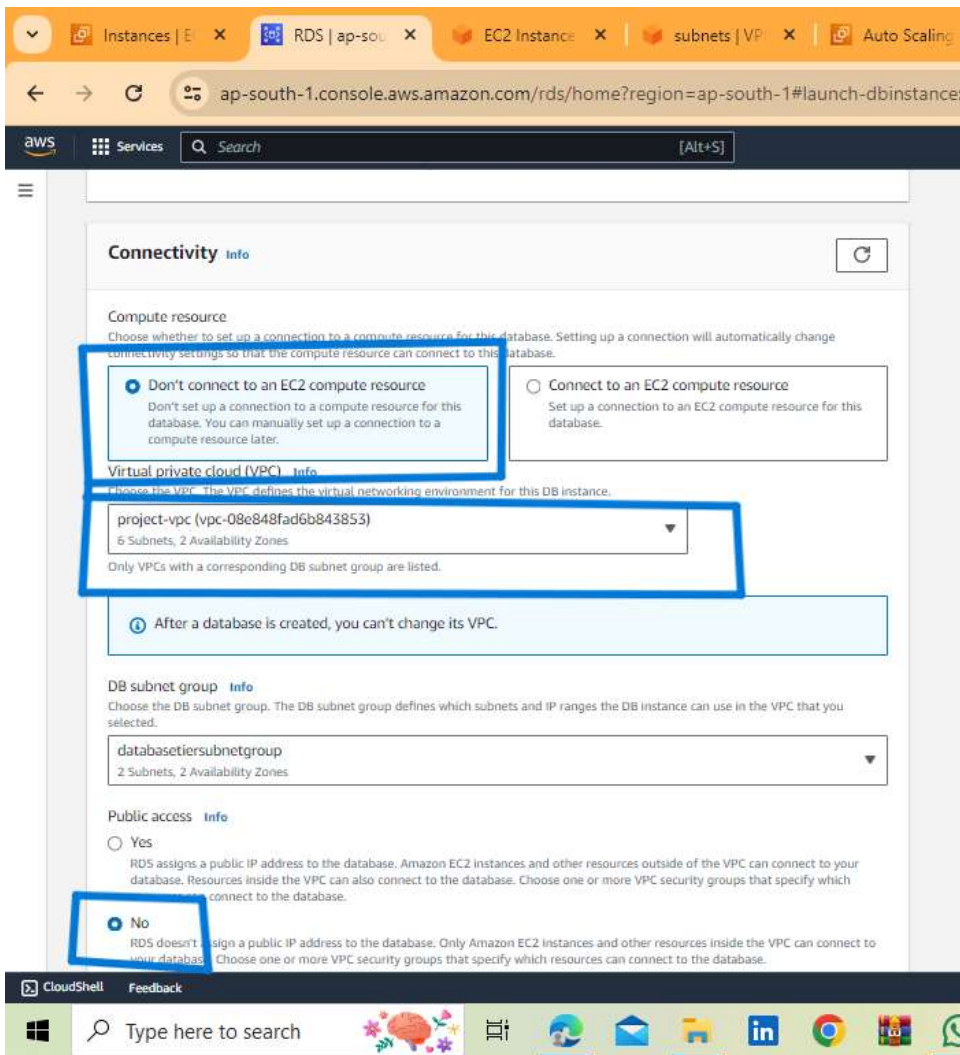
Next, you can choose a Multi-AZ deployment with three database instances: one primary instance and two read-only standby instances. This setup provides a very reliable system, but we do not need this at the moment.



There are also availability and durability options; however, none are available with the free tier. We do not need them either. Under 'Settings,' name your DB and create a master username and password. These credentials should be different from your AWS account login, as they are specific to the database you are creating.



We are going to set up our network manually, so choose not to connect to an EC2 resource. Select the appropriate VPC; the subnet group you created earlier should be listed as default. Choose 'Create new VPC security group (firewall)'.



In 'Database authentication,' I left the default option checked.

Database authentication

Database authentication options [Info](#)

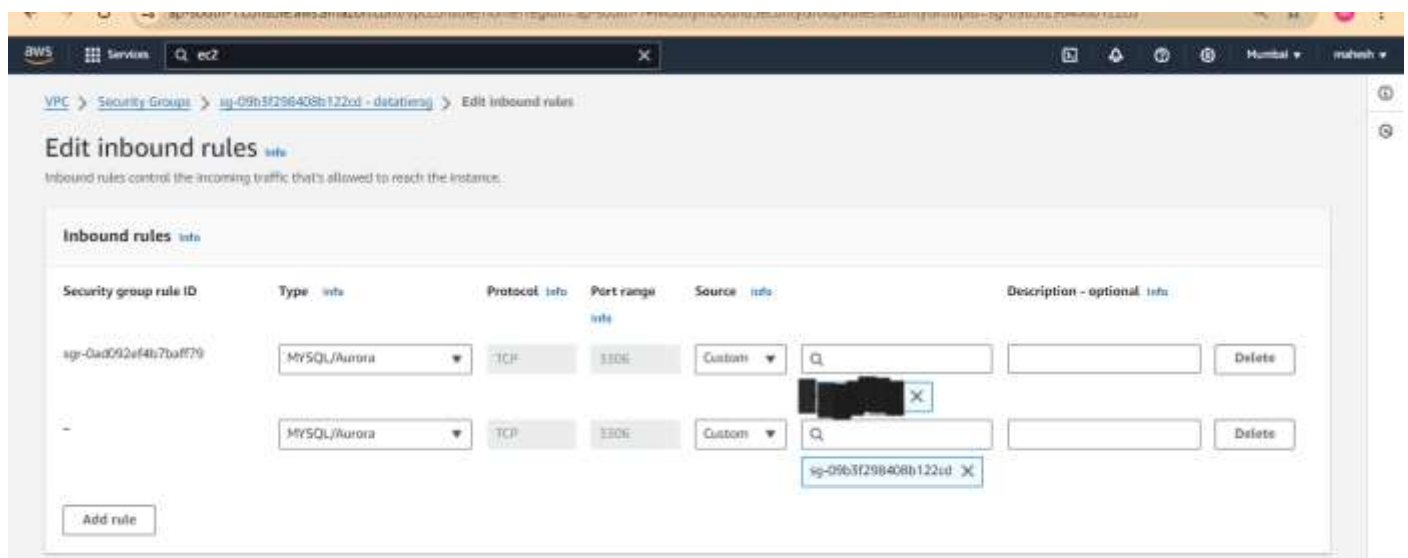
- ☒ Password authentication
Authenticates using database passwords.
- ☐ Password and IAM database authentication
Authenticates using the database password and user credentials through AWS IAM users and roles.
- ☐ Password and Kerberos authentication
Choose a directory in which you want to allow authorized users to authenticate with this DB instance using Kerberos Authentication.

Click the 'Create database' button.

Update the Database Tier Security Group:

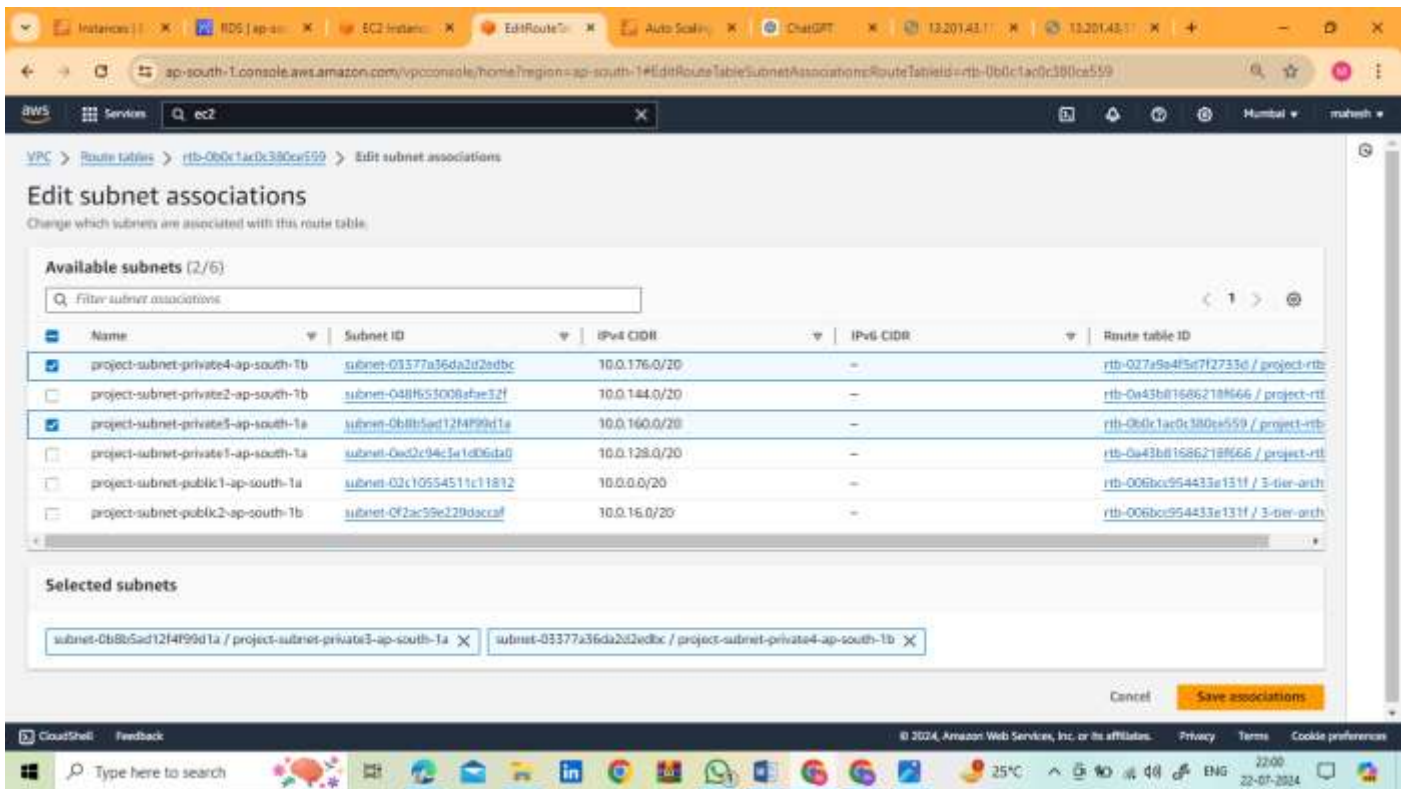
Navigate to the VPC console, select 'Security groups' from the left-side menu, and then find the database tier security group you just created. Select the security group you just created. You need to edit the inbound rules; by default, the database SG has an inbound rule to allow MySQL/Aurora traffic on port 3306 from your IP address. Delete this rule.

Create a new rule for MySQL/Aurora on port 3306. For the 'Source,' select 'Custom' and add your security group for your application layer (tier-2 SG).



UpdateTier3PrivateRouteTables:

In the last step for our database tier, we need to ensure that the route table associated with our database's private subnets lists both subnets in the subnet associations. If not, add the other subnet and save.



Our three-tier architecture is complete! We have already tested our web and application layers, but we are going to go a step further here.

Part5: Testing

We can't directly SSH to the database, but we can use an SSH forwarding agent to achieve this. You need to add your access key pair file to your keychain. To do this, first make sure you are on your local host (use the command `exit` to get out of any EC2 instance you're connected to). Then use the following command:

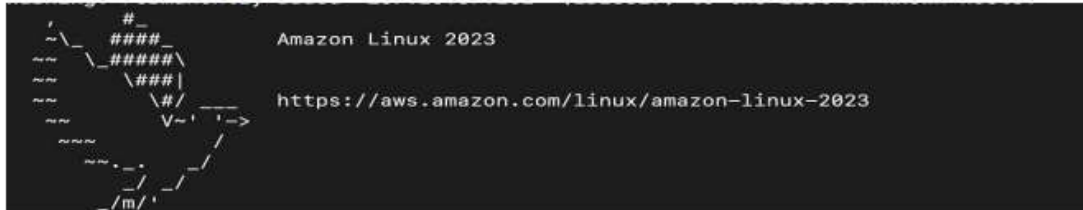
```
```bash
ssh-add -K <keypair.pem>
```
```

```
Identity added: LUIT_Project1.pem (LUIT_Project1.pem)
aaronbachman@Aarons-MacBook-Pro Keys %
```

Now that your key pair file is added to your keychain, the SSH agent will scan through all of the keys associated with the keychain and find your matching key.

Now reconnect to the web tier EC2; however, this time use `-A` to specify you want to use the SSH agent.

```
ssh-Aec2-user@<public-ip-address>
```



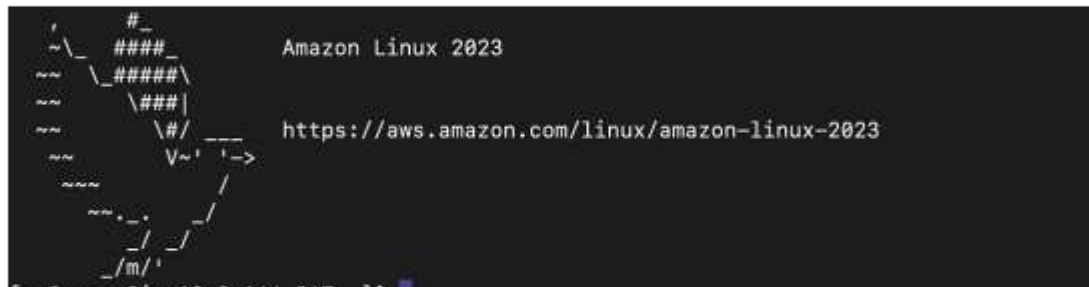
Once you are logged back into your tier-1 EC2, use the following command to check if the SSH agent forwarded the private key.

```
ssh-add-l
```



Our keypair has been forwarded to our public instance. Go copy your tier-2 application layer private IP address and copy it into the next command.

```
ssh-Aec2-user@<private-ip-address>
```



We have now SSH'd from your public tier 1 web instance into your private tier 2 application instance!

Testing Connectivity to the Database Tier

There are a few ways you can connect to your RDS database from your application tier. One way is to install MySQL on your private tier 2 instance to access your database. We are going to utilize this method. While logged into your application tier instance, use this command:

```
sudo dnf install mariadb105-server
```

```
Installed:
mariadb-connector-c-3.1.13-1.amzn2023.0.3.x86_64
mariadb-connector-c-config-3.1.13-1.amzn2023.0.3.noarch
mariadb105-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-backup-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-common-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-cracklib-password-check-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-errmsg-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-gssapi-server-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-server-3:10.5.16-1.amzn2023.0.7.x86_64
mariadb105-server-utils-3:10.5.16-1.amzn2023.0.7.x86_64
mysql-selinux-1.0.4-2.amzn2023.0.3.noarch
perl-B-1.80-477.amzn2023.0.3.x86_64
perl-DBD-MariaDB-1.22-1.amzn2023.0.4.x86_64
perl-DBI-1.643-7.amzn2023.0.3.x86_64
perl-Data-Dumper-2.174-460.amzn2023.0.2.x86_64
perl-File-Copy-2.34-477.amzn2023.0.3.noarch
perl-FileHandle-2.03-477.amzn2023.0.3.noarch
perl-Math-BigInt-1:1.9998.18-458.amzn2023.0.2.noarch
perl-Math-Complex-1.59-477.amzn2023.0.3.noarch
perl-Sys-Hostname-1.23-477.amzn2023.0.3.x86_64
perl-base-2.27-477.amzn2023.0.3.noarch

Complete!
```

This command installs the MariaDB package, which is used to read MySQL. Once installed, you should be able to use the following command to log into your RDS MySQL database. You will need your RDS endpoint, user name, and password. To find your RDS database endpoint, navigate to the database you created and find the endpoint under Connectivity & Security.

`mysql-h<rds-database-endpoint>-P3306-u<username>-p`

```
[ec2-user@ip-10-0-10-10 ~]$ mysql -h datatierdb.cmytqkhcetki.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 8.0.32 Source distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MySQL [(none)]>
```

We have now successfully connected to our MySQL database from the application tier. We have connectivity with all of our tiers!