Cognizant Technology Solutions

A Project Report on

"EMPLOYEE PERFORMANCE TRACKING"

Prepared By

Ankita Shivane	2320390
Bhavesh Uprikar	2320803
Harshit Yadav	2320382
Mimansa Yadav	2320956
Nandhini T	2320539
Rajanya Kamilya	2320828
Shambhavi Shukla	2320529

REQUIREMENTS

Employee Performance Tracking:

• Import employee performance data into Amazon Redshift.
• Track employee performance metrics such as sales targets, customer satisfaction ratings, and productivity.
• Identify top-performing employees by analyzing performance metrics and comparing them to predefined targets or benchmarks.
• Identify areas for improvement by analyzing performance gaps and providing targeted training or support.
• Visualize employee performance data using charts and graphs (e.g., bar charts, radar charts) to identify trends and patterns.

INDEX

TOPIC

PURPOSE:

- **Data Analysis:** Perform complex analytics on large datasets stored in S3 using the powerful querying capabilities of Redshift.

- **Business Intelligence (BI):** Create reports, dashboards, and visualizations from data stored in S3 by loading it into Redshift for BI purposes.

- **Data Warehousing:** Store and organize large volumes of structured and semi-structured data from S3 into Redshift for centralized data warehousing.

- **Data Integration:** Integrate data from multiple sources stored in S3 into a single data warehouse in Redshift for comprehensive analysis and reporting.

- **Data Migration:** Migrate data from on-premises data warehouses or other cloud storage solutions to Redshift for better scalability, performance, and cost-effectiveness.

- **ETL (Extract, Transform, Load):** Extract data from S3, transform it as needed, and load it into Redshift for further analysis and reporting.

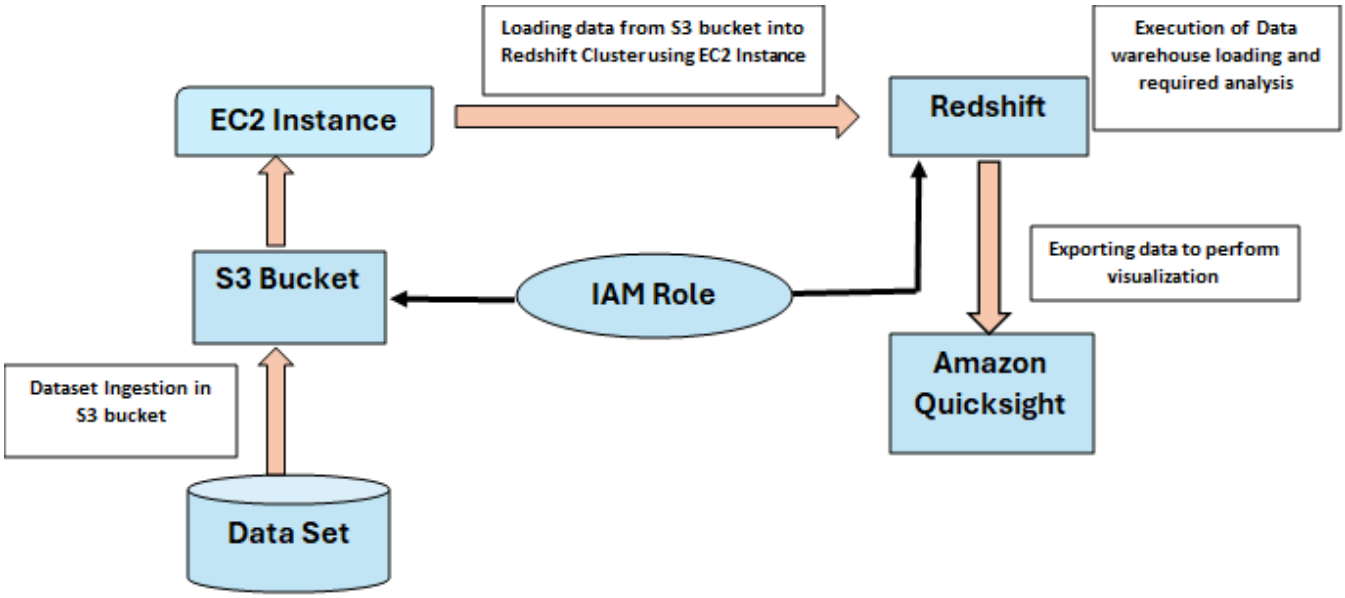
- **Real-time Analytics:** Continuously load streaming data from S3 into Redshift to perform real-time analytics and gain insights into rapidly changing data.

- **Cost Optimization:** Optimize costs by storing raw or historical data in S3's cost-effective storage and transferring only relevant data into Redshift for analysis.

- **Data Archiving:** Archive historical data from S3 into Redshift for long-term storage and analysis while keeping it easily accessible for future reference.

- **Data Governance and Security**: Centralize sensitive data stored in S3 into Redshift for better governance, access control, and security management, ensuring compliance with regulatory requirements.

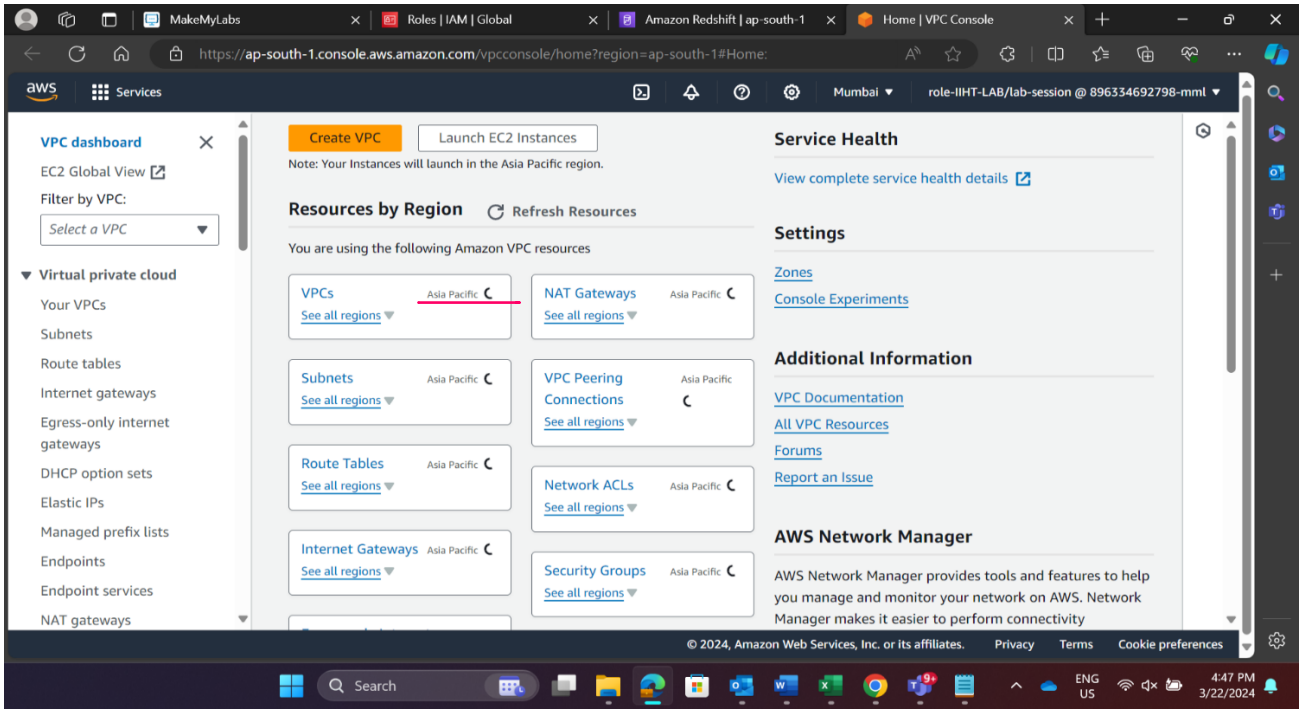
PICTORIAL FLOWCHART

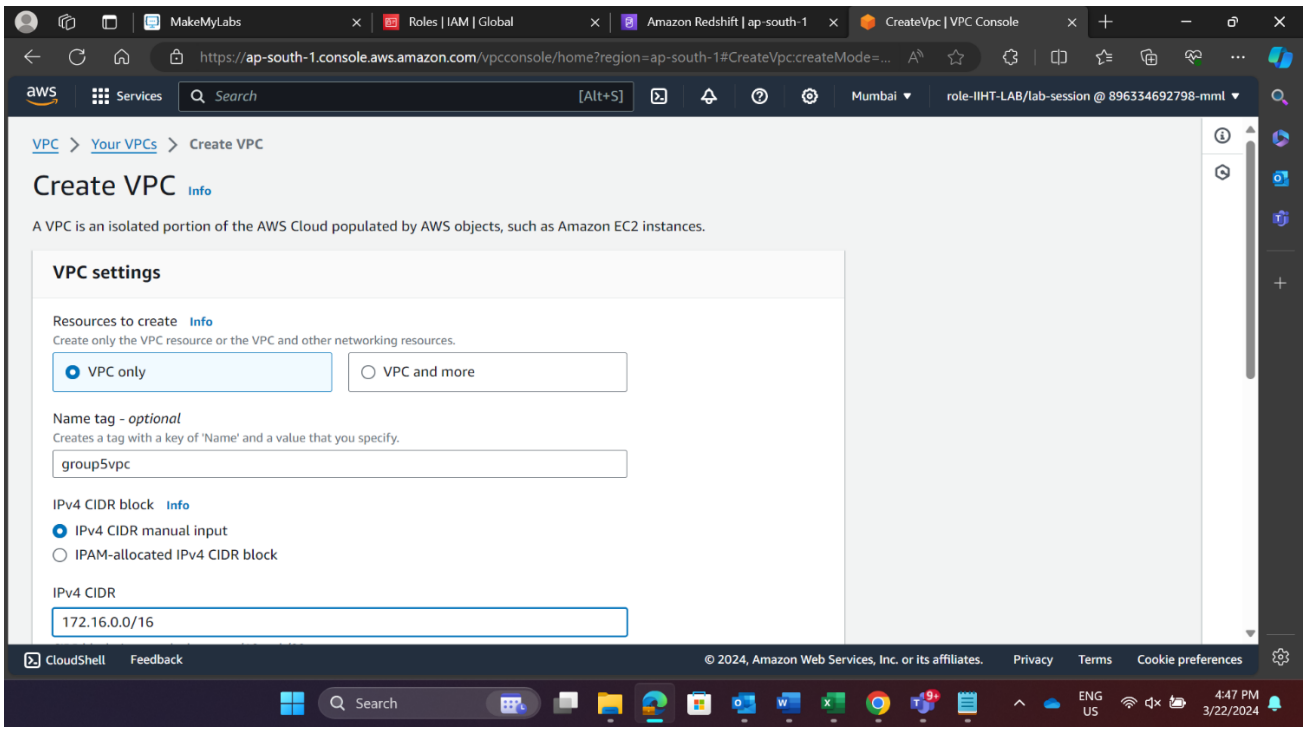


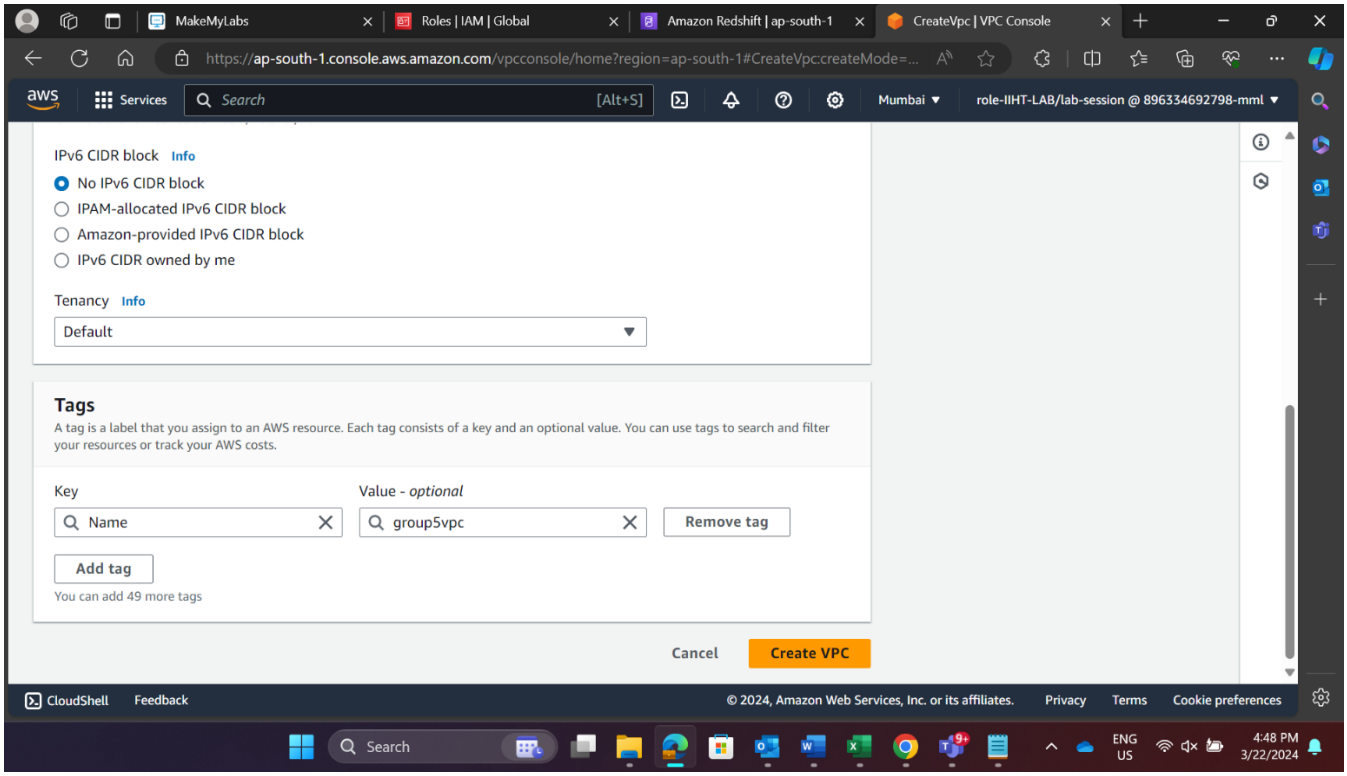
DESIGN AND ARCHITECTURE

VPC CREATION

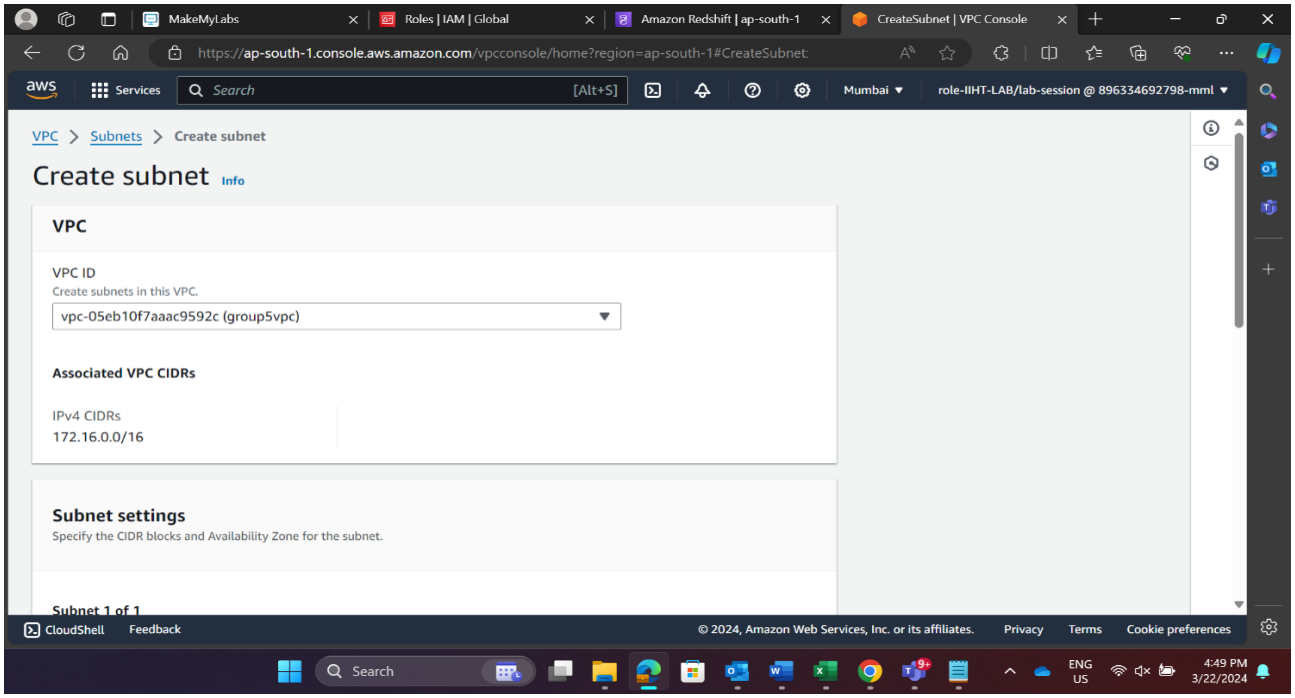
Step 1: Initiate the VPC creation process, provide the necessary details for your VPC, including the VPC name, CIDR block (the range of IP addresses for your VPC).

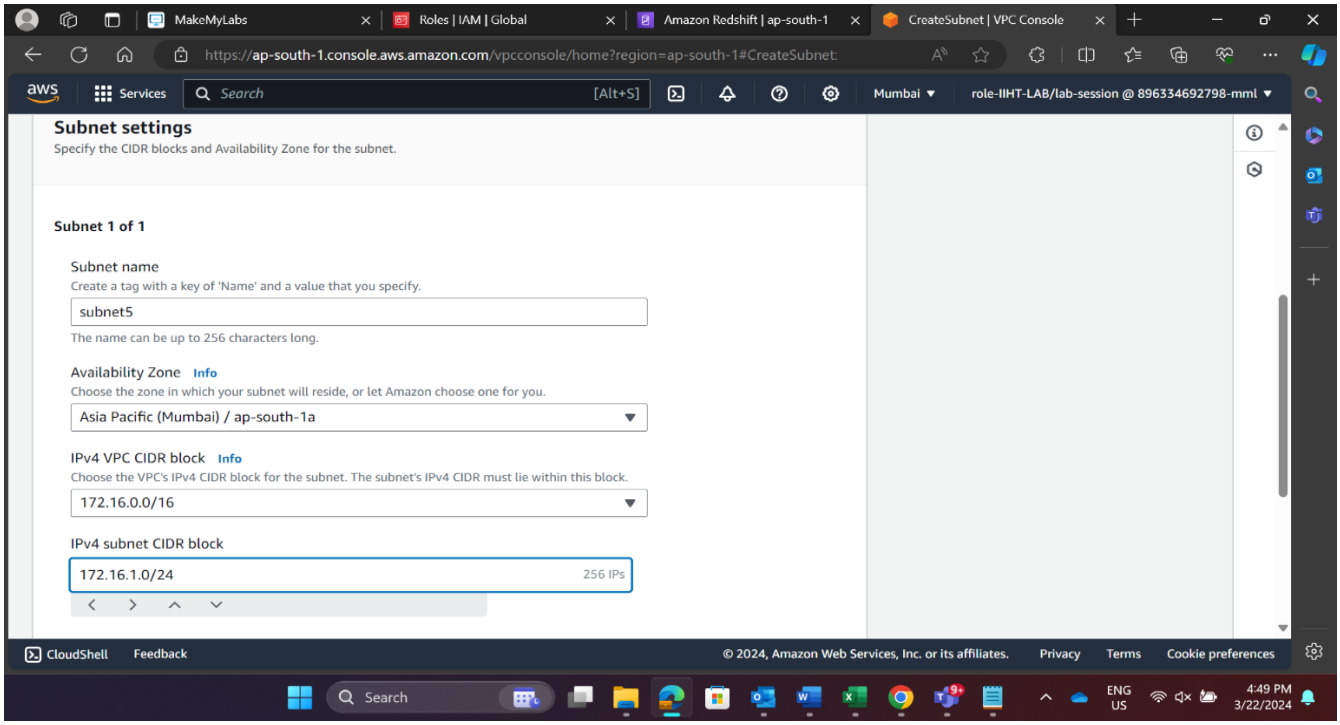


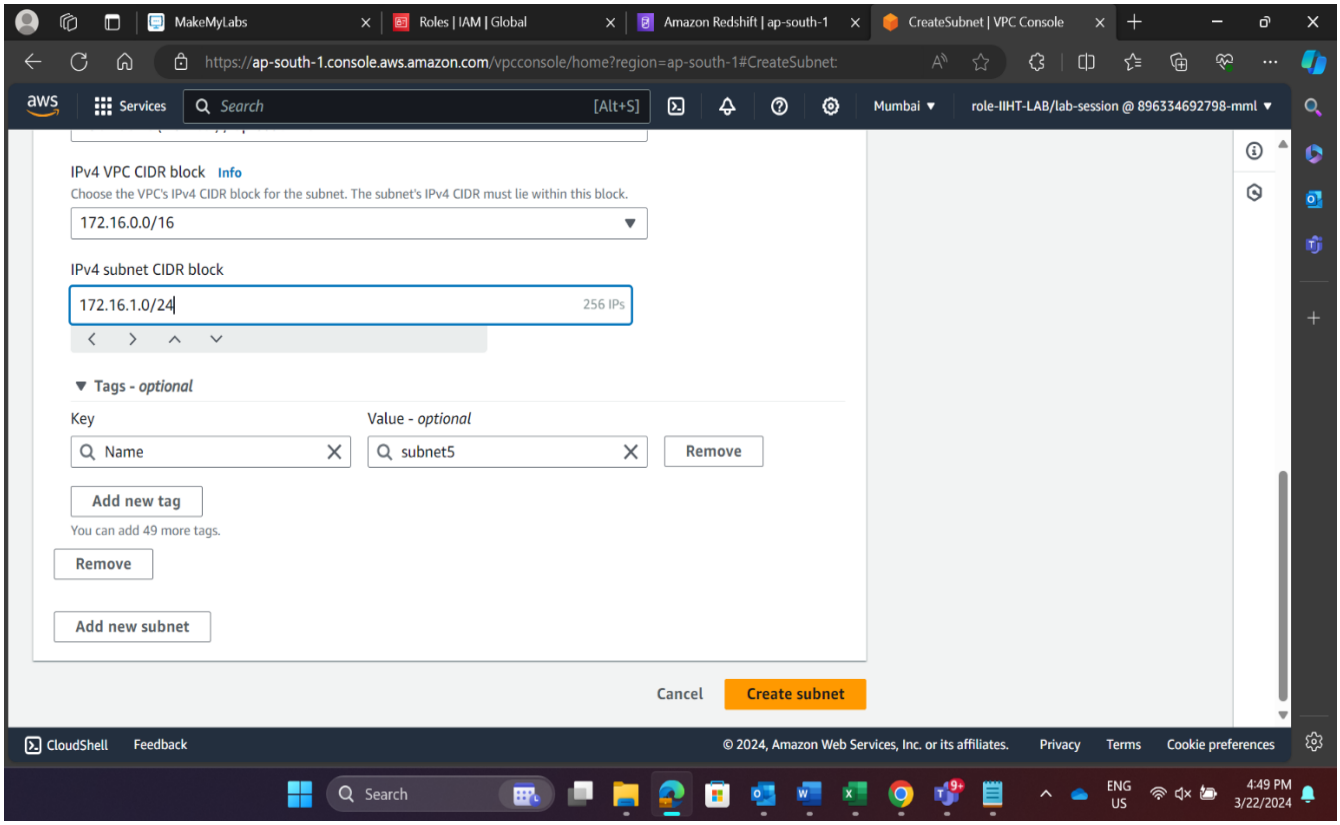


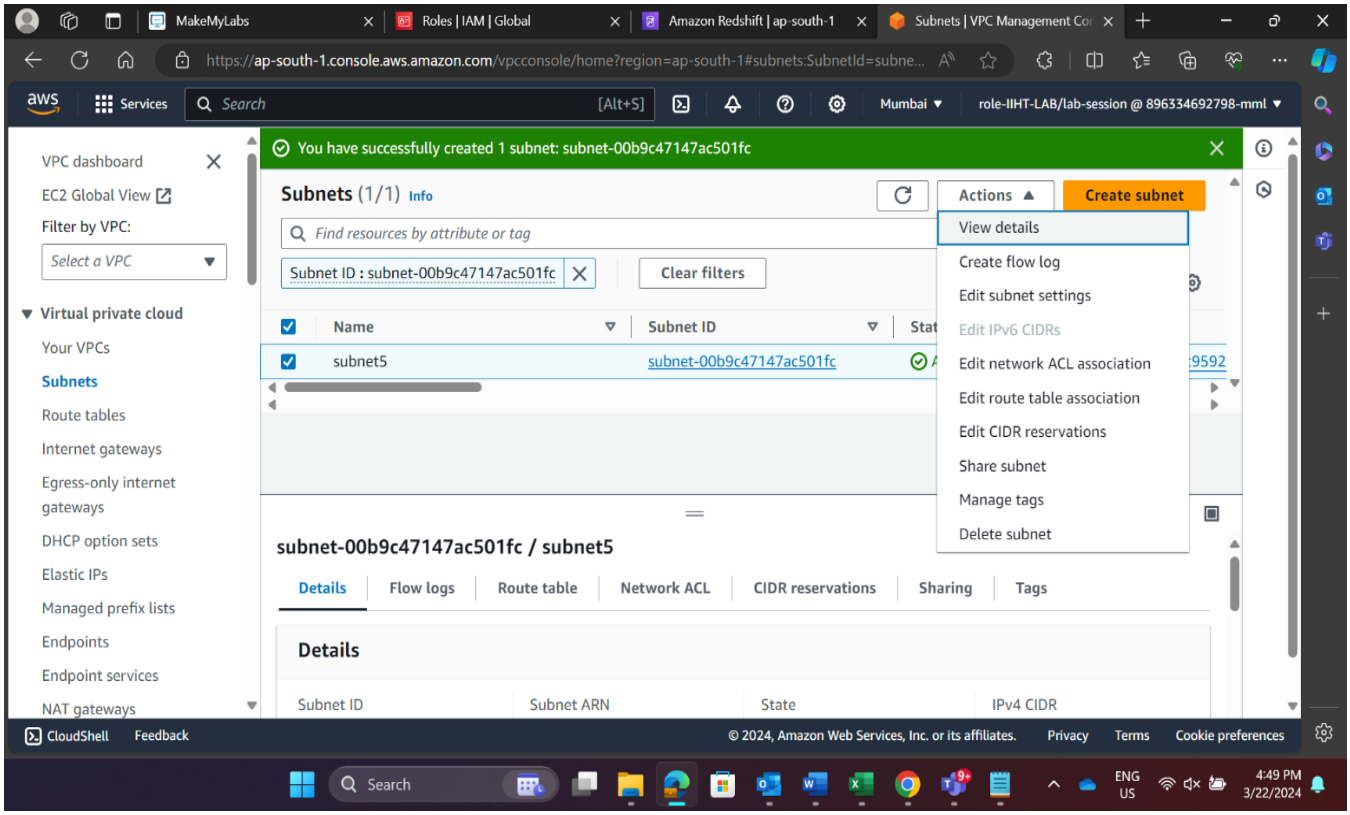


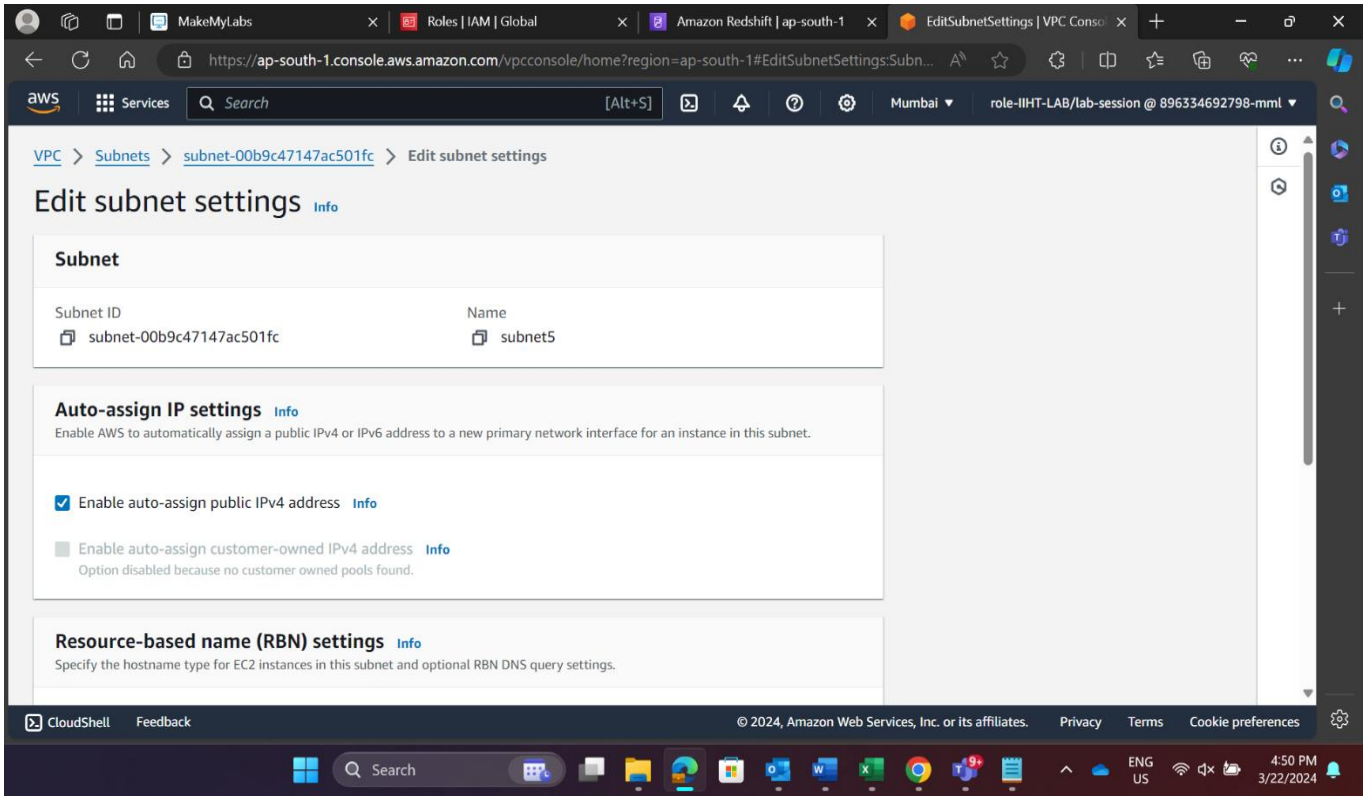
Step 2: Define the subnets within your VPC and specify the CIDR block for subnet.

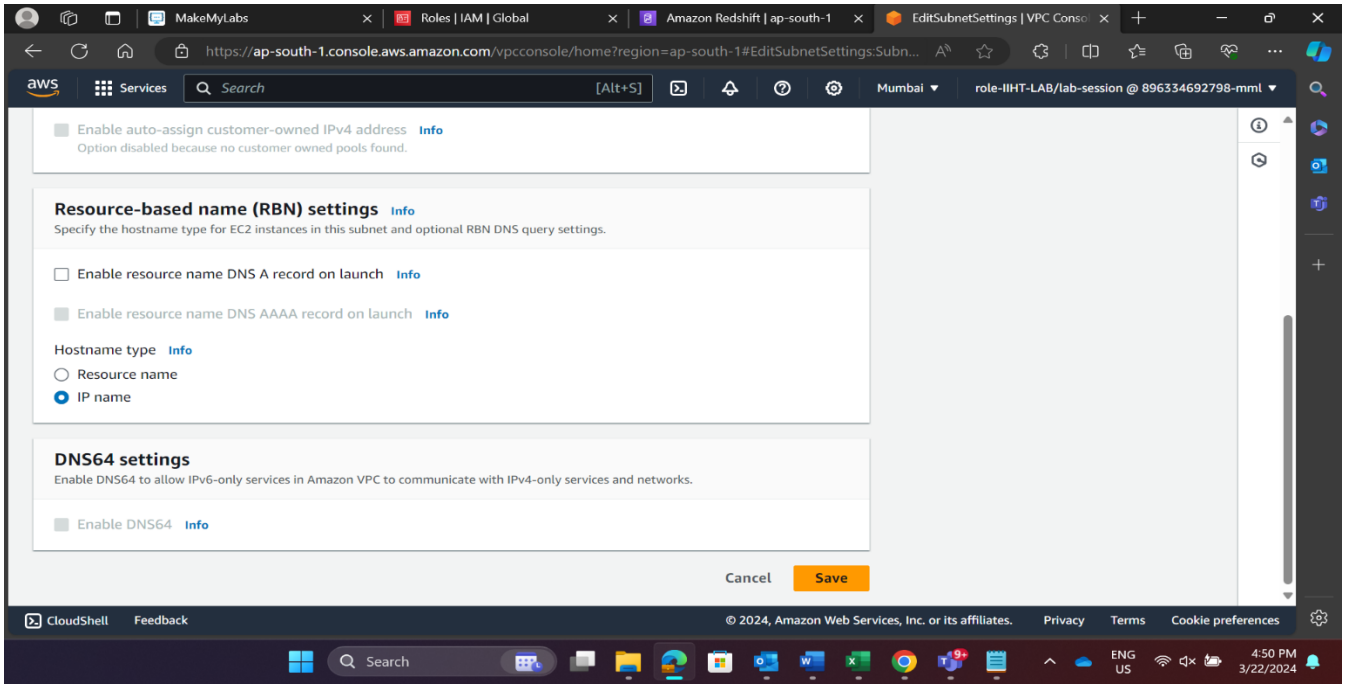



MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
CreateSubnet | VPC Console
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#CreateSubnet:
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
Subnet settings
Specify the CIDR blocks and Availability Zone for the subnet.
Subnet 1 of 1
Subnet name
Create a tag with a key of 'Name' and a value that you specify.
subnet5
The name can be up to 256 characters long.
Availability Zone Info
Choose the zone in which your subnet will reside, or let Amazon choose one for you.
Asia Pacific (Mumbai) / ap-south-1a
IPv4 VPC CIDR block Info
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.
172.16.0.0/16
IPv4 subnet CIDR block
172.16.1.0/24
256 IPs
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:49 PM
3/22/2024

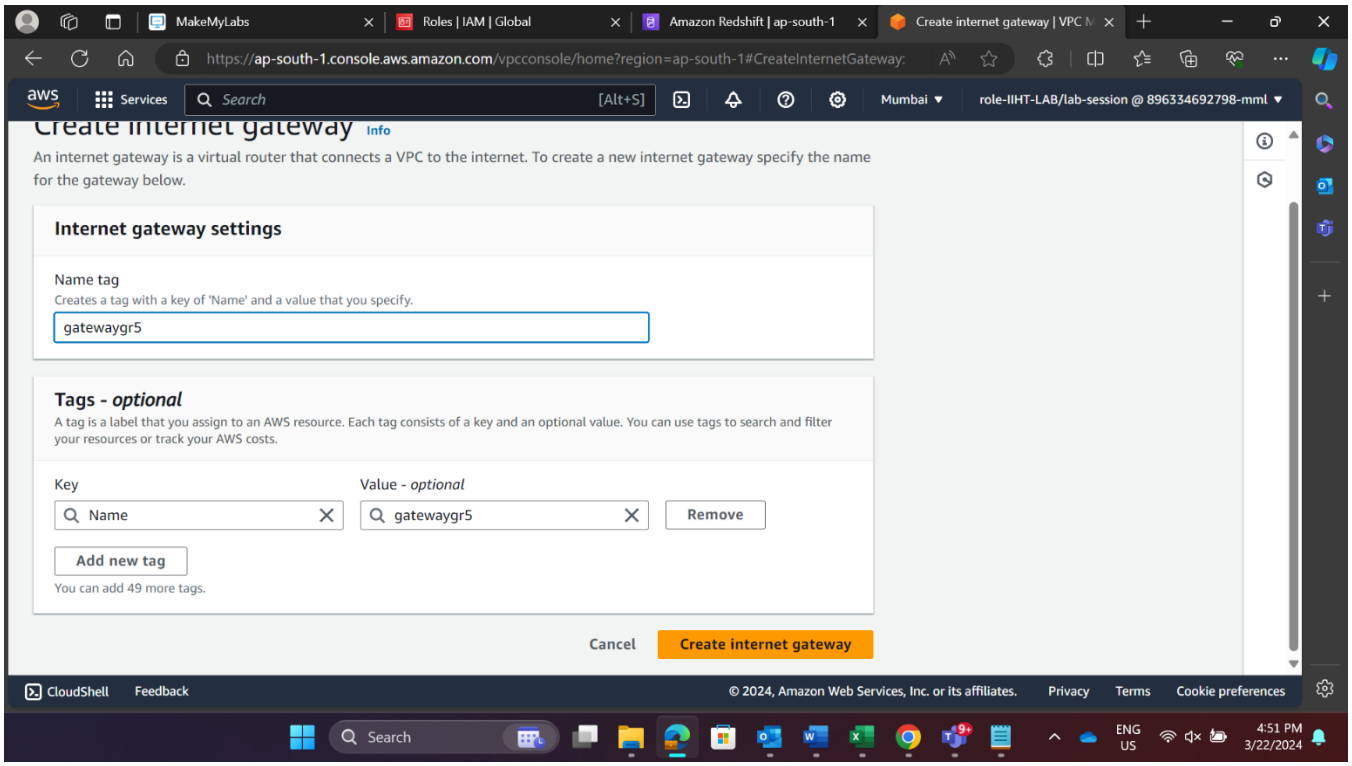

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
CreateSubnet | VPC Console
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#CreateSubnet:
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
IPv4 VPC CIDR block Info
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.
172.16.0.0/16
IPv4 subnet CIDR block
172.16.1.0/24
256 IPs
Tags - optional
Key
Value - optional
Name
subnet5
Remove
Add new tag
You can add 49 more tags.
Remove
Add new subnet
Cancel
Create subnet
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:49 PM
3/22/2024

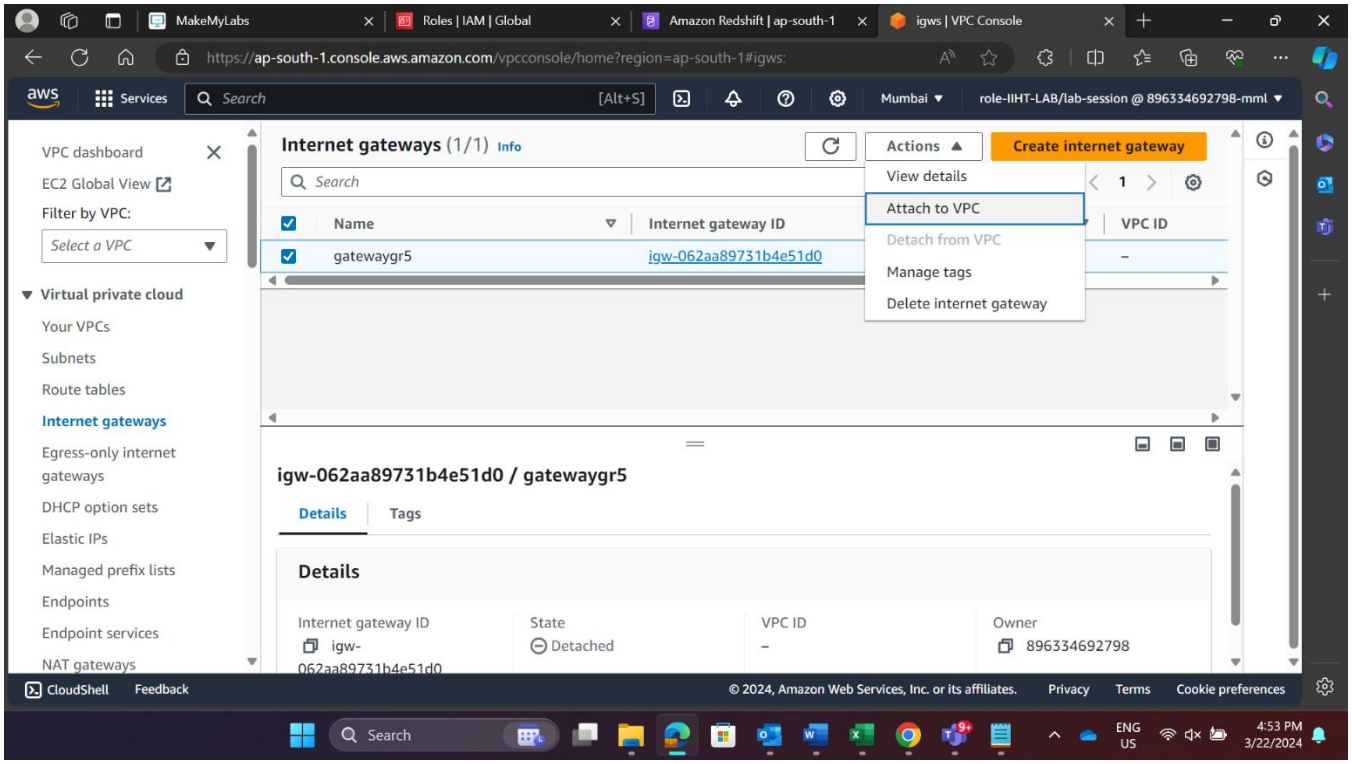

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
Subnets | VPC Management Co
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#subnets:SubnetId=subne...
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
VPC dashboard
EC2 Global View
Filter by VPC:
Select a VPC
Virtual private cloud
Your VPCs
Subnets
Route tables
Internet gateways
Egress-only internet gateways
DHCP option sets
Elastic IPs
Managed prefix lists
Endpoints
Endpoint services
NAT gateways
You have successfully created 1 subnet: subnet-00b9c47147ac501fc
Subnets (1/1) Info
Actions
Create subnet
Find resources by attribute or tag
Subnet ID : subnet-00b9c47147ac501fc
Clear filters
Name
Subnet ID
subnet5
subnet-00b9c47147ac501fc
View details
Create flow log
Edit subnet settings
Edit IPv6 CIDRs
Edit network ACL association
Edit route table association
Edit CIDR reservations
Share subnet
Manage tags
Delete subnet
subnet-00b9c47147ac501fc / subnet5
Details
Flow logs
Route table
Network ACL
CIDR reservations
Sharing
Tags
Details
Subnet ID
Subnet ARN
State
IPv4 CIDR
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG US
4:49 PM
3/22/2024

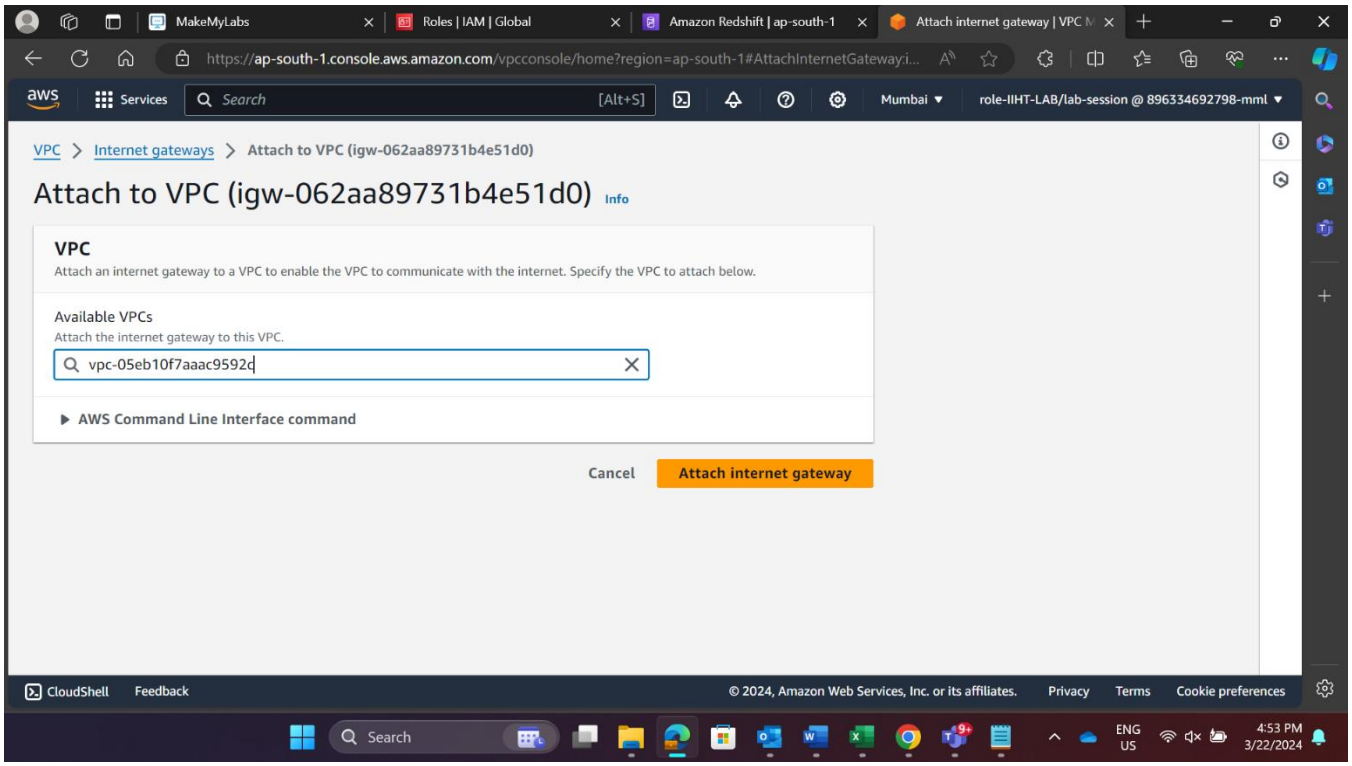

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
EditSubnetSettings | VPC Conso
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#EditSubnetSettings:Subn...
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
VPC > Subnets > subnet-00b9c47147ac501fc > Edit subnet settings
Edit subnet settings Info
Subnet
Subnet ID
subnet-00b9c47147ac501fc
Name
subnet5
Auto-assign IP settings Info
Enable AWS to automatically assign a public IPv4 or IPv6 address to a new primary network interface for an instance in this subnet.
Enable auto-assign public IPv4 address Info
Enable auto-assign customer-owned IPv4 address Info
Option disabled because no customer owned pools found.
Resource-based name (RBN) settings Info
Specify the hostname type for EC2 instances in this subnet and optional RBN DNS query settings.
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG US
4:50 PM
3/22/2024



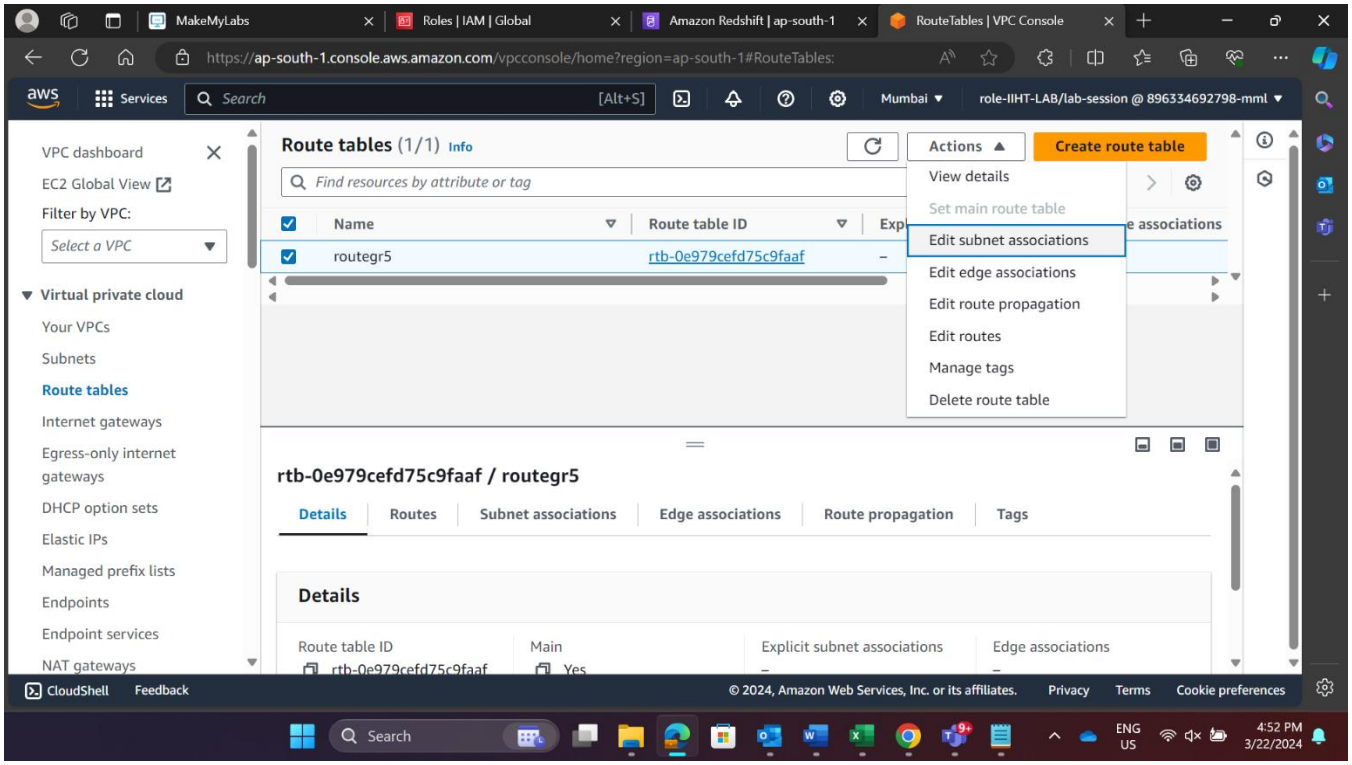
Step 3: Configure Internet Gateway: create a new internet gateway, give it a name, and attach it to your VPC.

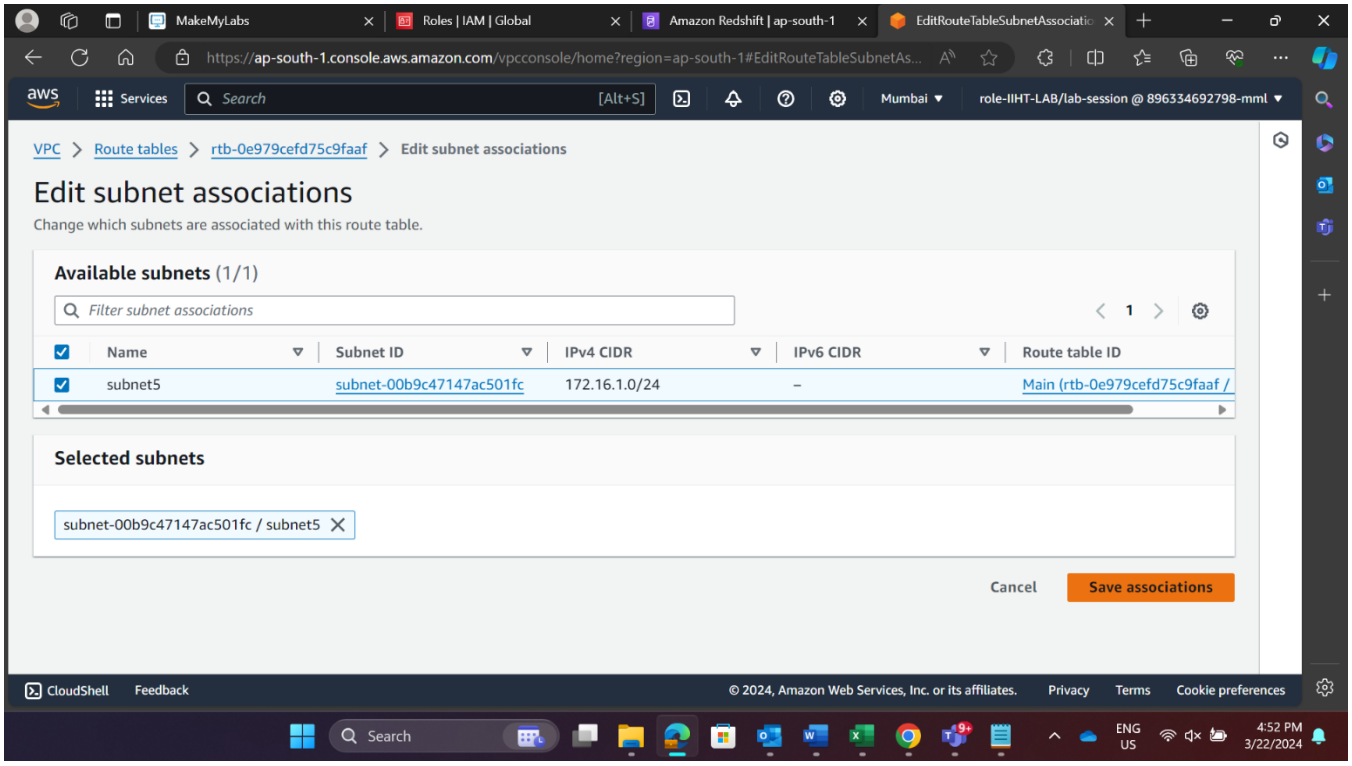


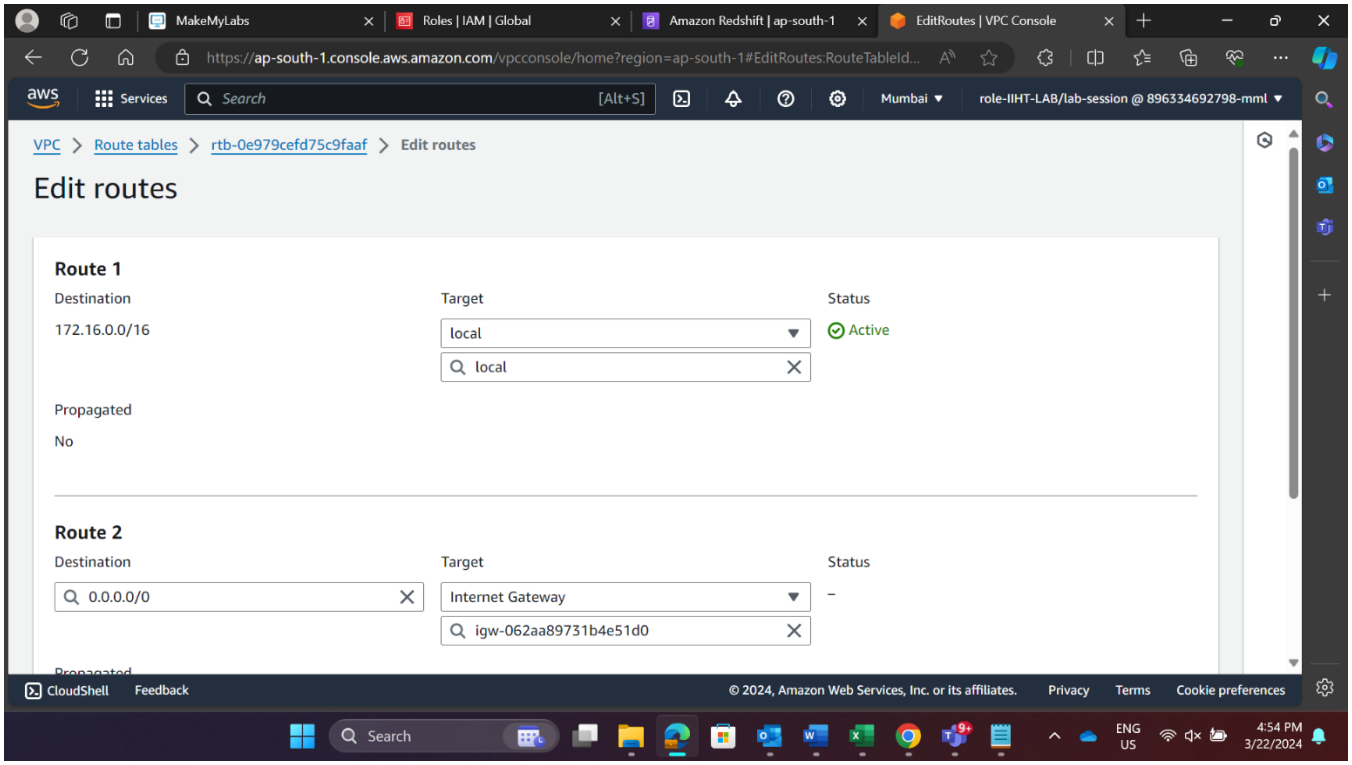

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
igws | VPC Console
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#igws:
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
VPC dashboard
EC2 Global View
Filter by VPC:
Select a VPC
Virtual private cloud
Your VPCs
Subnets
Route tables
Internet gateways
Egress-only internet gateways
DHCP option sets
Elastic IPs
Managed prefix lists
Endpoints
Endpoint services
NAT gateways
Internet gateways (1/1) Info
Actions
Create internet gateway
View details
Attach to VPC
Detach from VPC
Manage tags
Delete internet gateway
Name
Internet gateway ID
VPC ID
gatewaygr5
igw-062aa89731b4e51d0
igw-062aa89731b4e51d0 / gatewaygr5
Details
Tags
Internet gateway ID
State
Detached
VPC ID
Owner
896334692798
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
4:53 PM
3/22/2024

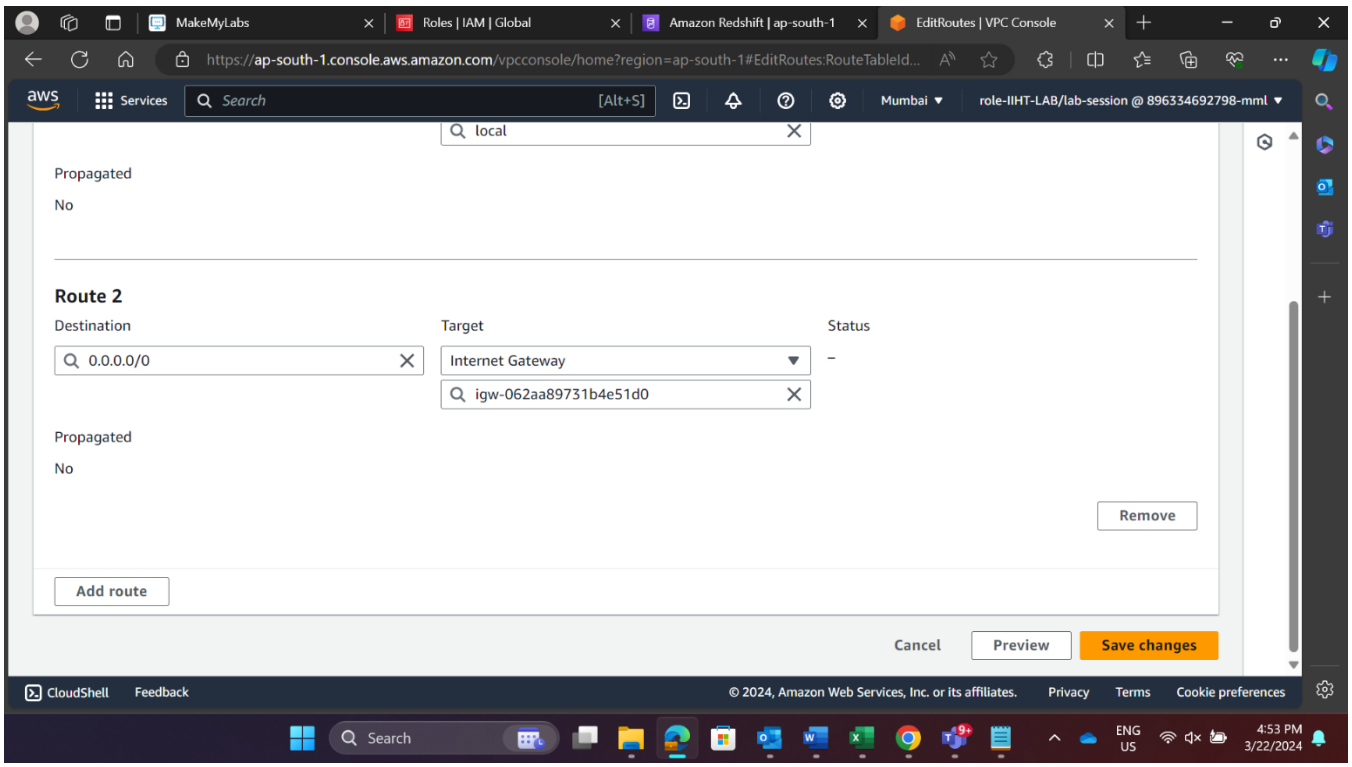

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
Attach internet gateway | VPC
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#AttachInternetGateway:i...
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
VPC > Internet gateways > Attach to VPC (igw-062aa89731b4e51d0)
Attach to VPC (igw-062aa89731b4e51d0) Info
VPC
Attach an internet gateway to a VPC to enable the VPC to communicate with the internet. Specify the VPC to attach below.
Available VPCs
Attach the internet gateway to this VPC.
vpc-05eb10f7aaac9592d
AWS Command Line Interface command
Cancel
Attach internet gateway
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
4:53 PM
3/22/2024

Step 4: Configure Route Table: create a route table, give it a name, and attach it to internet gateway.



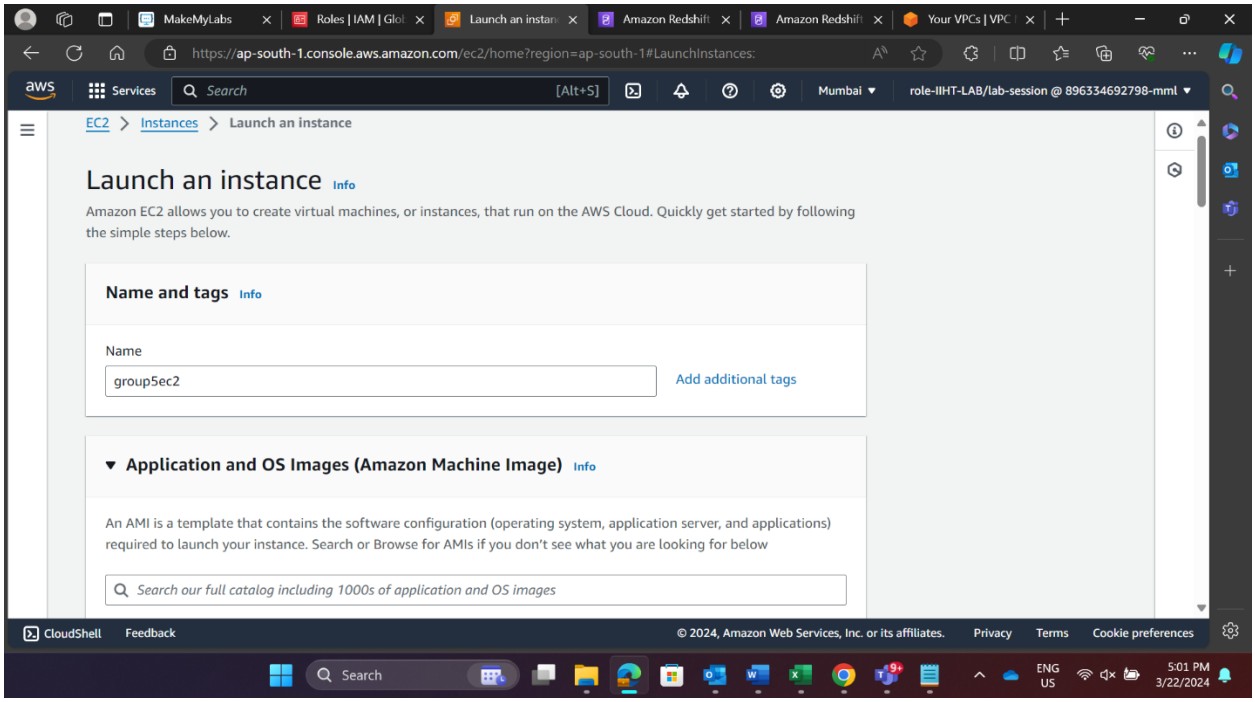



MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
EditRoutes | VPC Console
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#EditRoutes:RouteTableId...
aws
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
VPC > Route tables > rtb-0e979cefd75c9faaf > Edit routes
Edit routes
Route 1
Destination
Target
Status
172.16.0.0/16
local
Active
local
Propagated
No
Route 2
Destination
Target
Status
0.0.0.0/0
Internet Gateway
igw-062aa89731b4e51d0
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:54 PM
3/22/2024

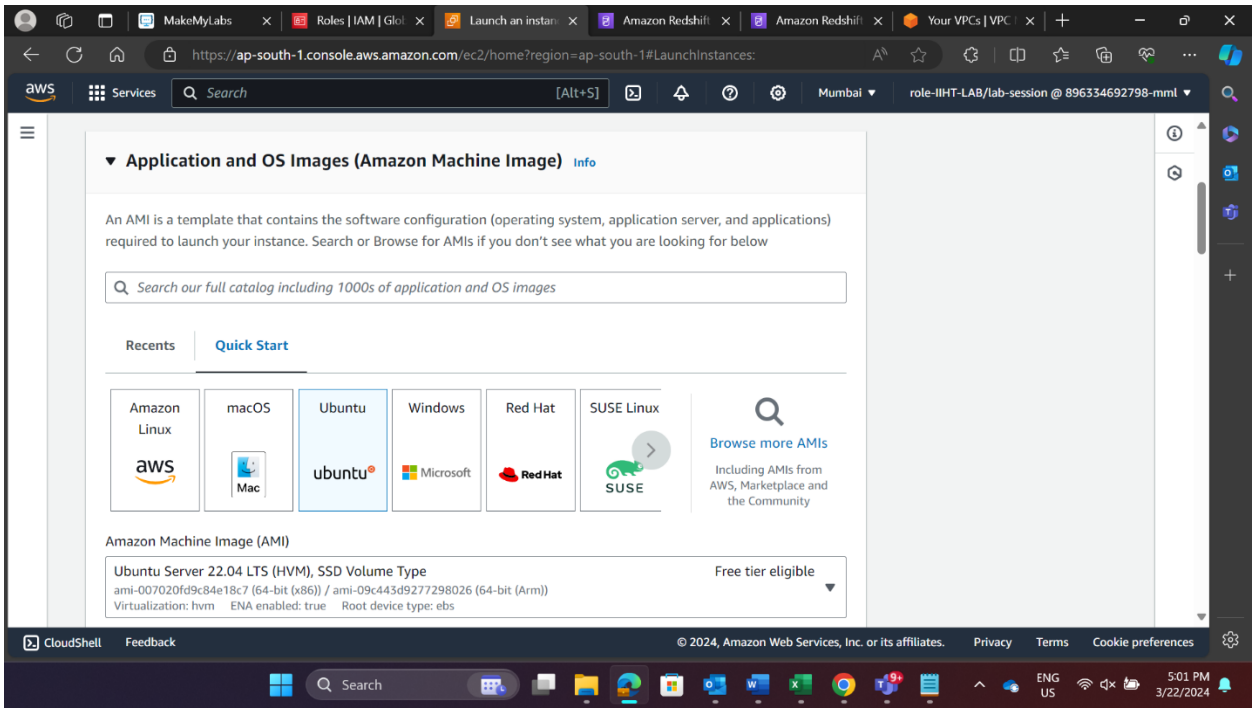

MakeMyLabs
Roles | IAM | Global
Amazon Redshift | ap-south-1
EditRoutes | VPC Console
https://ap-south-1.console.aws.amazon.com/vpcconsole/home?region=ap-south-1#EditRoutes:RouteTableId...
aws
Services
Search
[Alt+S]
Mumbai
role-IIHT-LAB/lab-session @ 896334692798-mml
local
Propagated
No
Route 2
Destination
Target
Status
0.0.0.0/0
Internet Gateway
igw-062aa89731b4e51d0
Propagated
No
Remove
Add route
Cancel
Preview
Save changes
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:53 PM
3/22/2024

EC2 INSTANCE

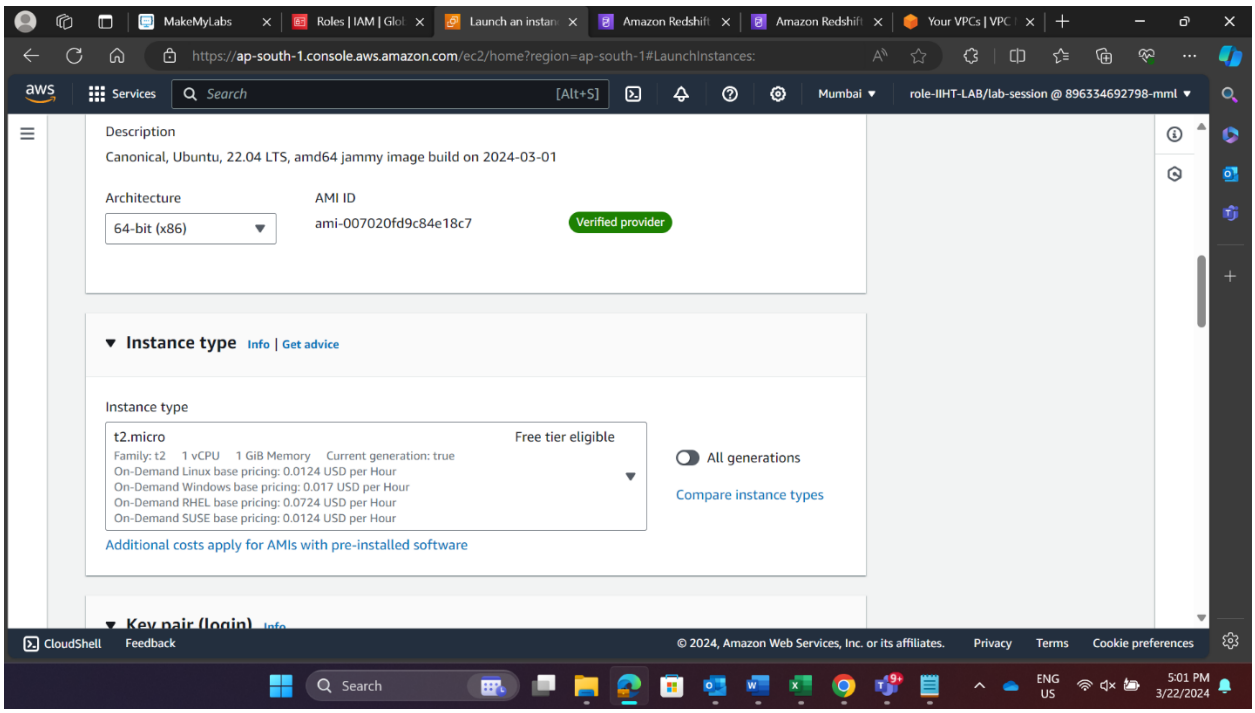
Step 1: On the EC2 dashboard, click the **"Launch Instance"** button. This will start the instance creation wizard.



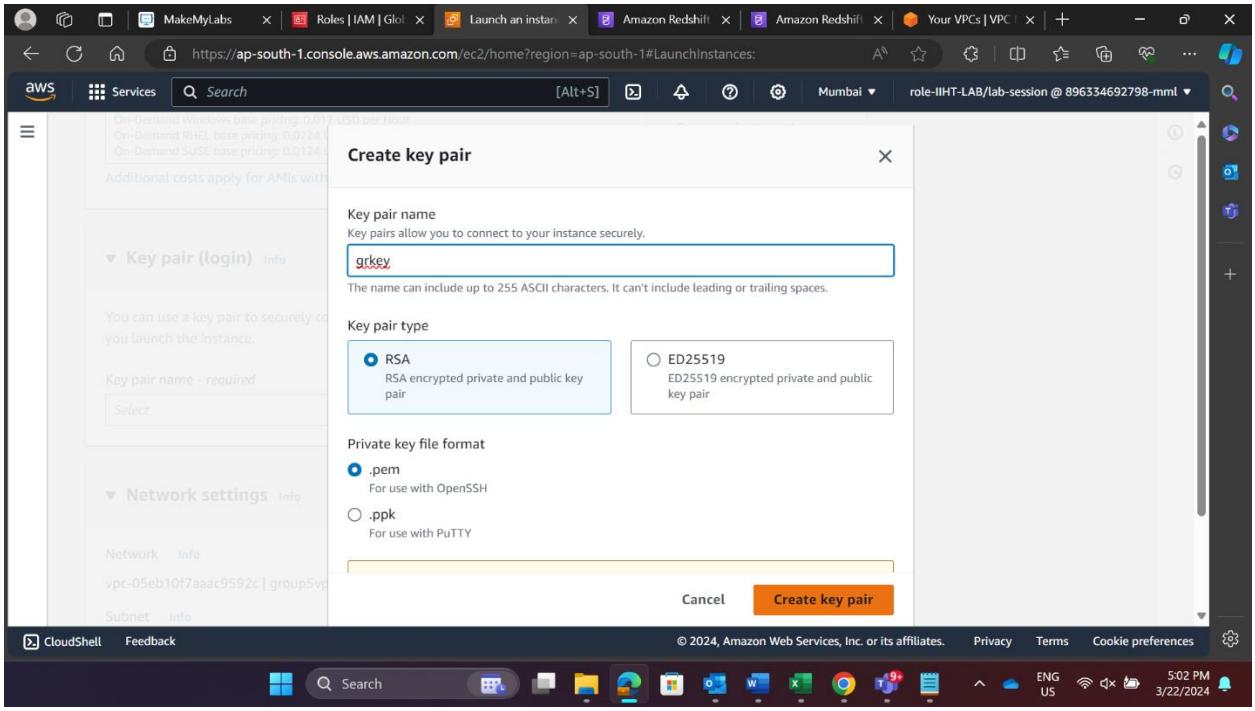
Step 2: Choose an Amazon Machine Image (AMI). It is a template for instance's operating system.



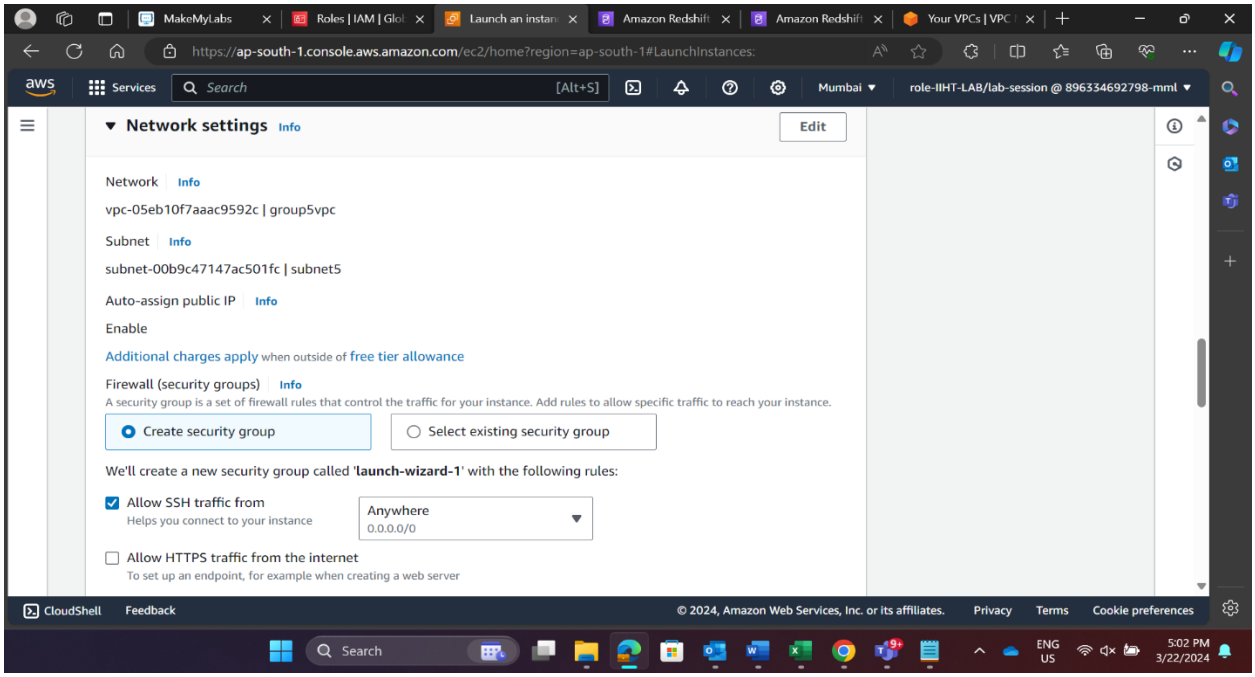
Step 3: In this step, select the type of instance based on your business requirements.



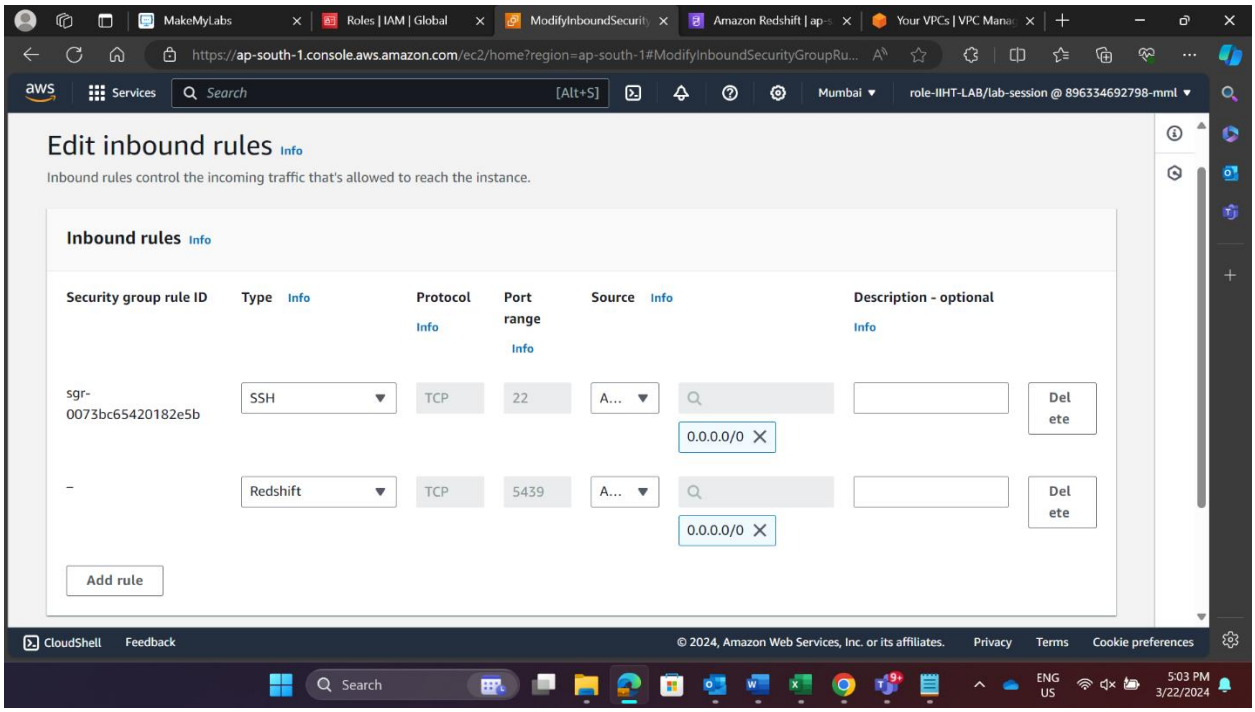
Step 4: If you don't have an existing key pair, create a new one. Download the private key file (.pem) and keep it secure.



Step 5: Set up networking details like VPC, subnet, and security groups.

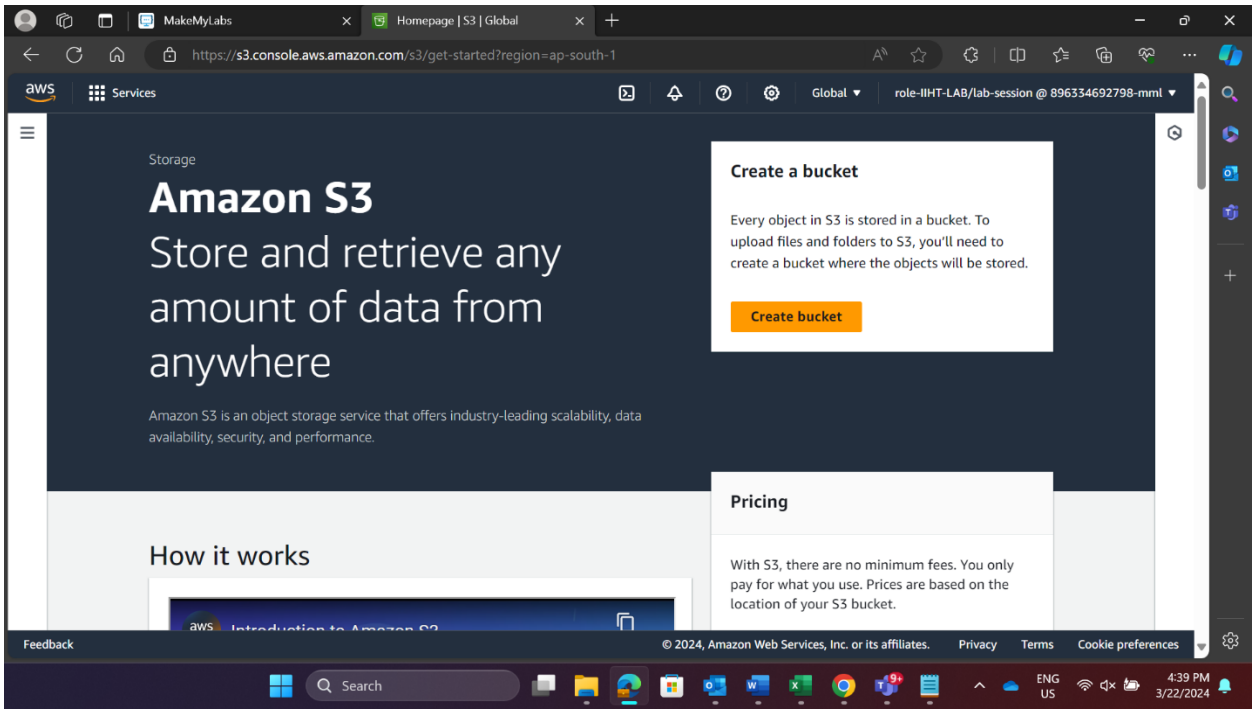


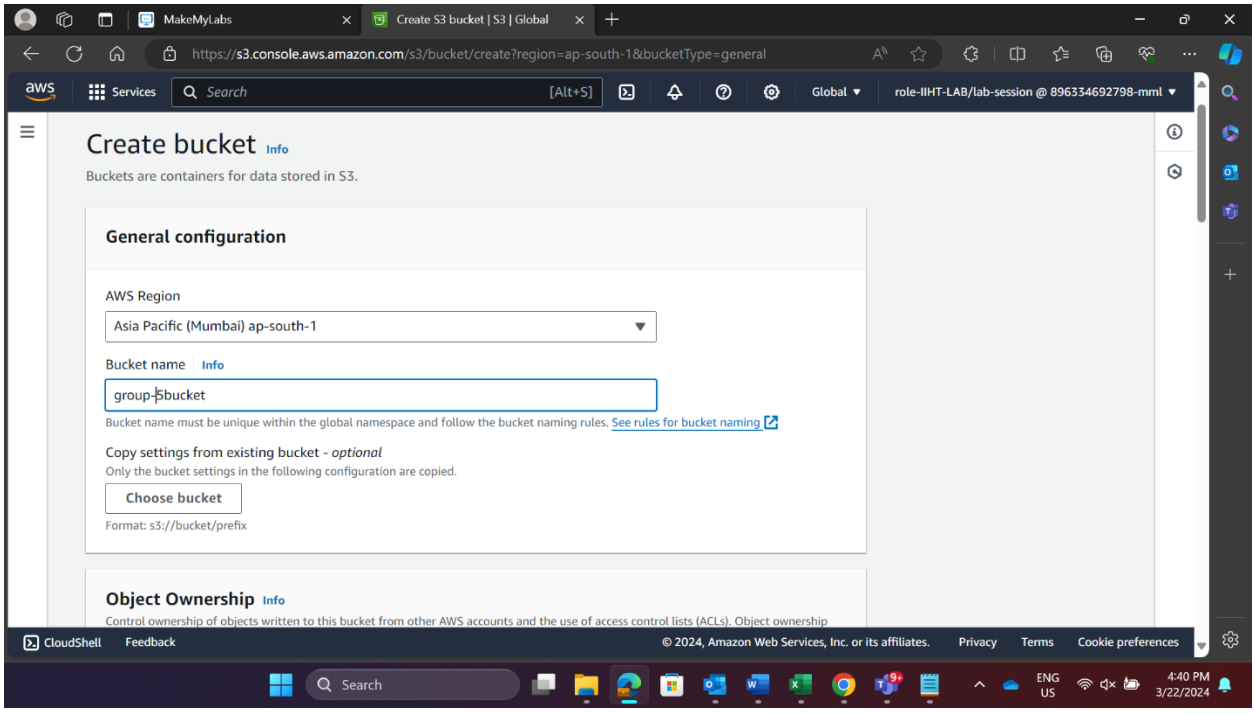
Step 6: Add rules for SSH and Redshift in inbound rule for allowing the traffic to redshift cluster.



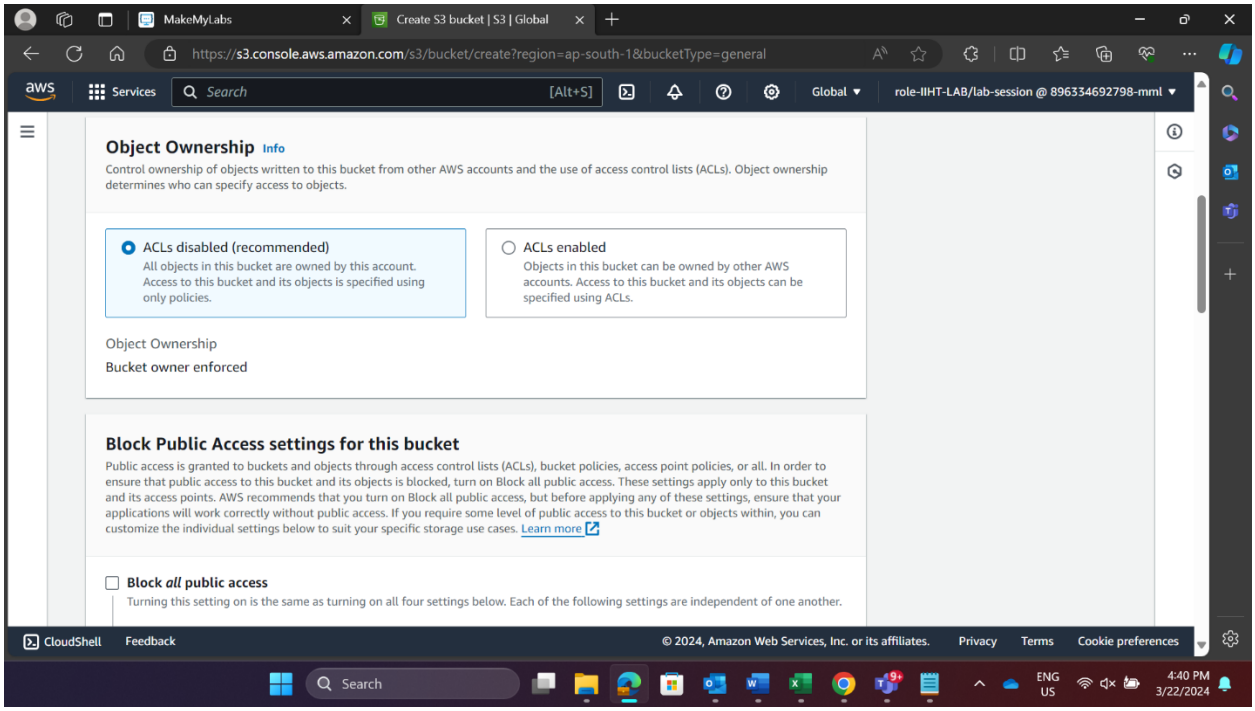
S3 BUCKET

Step 1: Create bucket by giving unique name and select the AWS region where you want to create bucket. Configure bucket permissions whether to keep the bucket private or public. After creating bucket add objects.

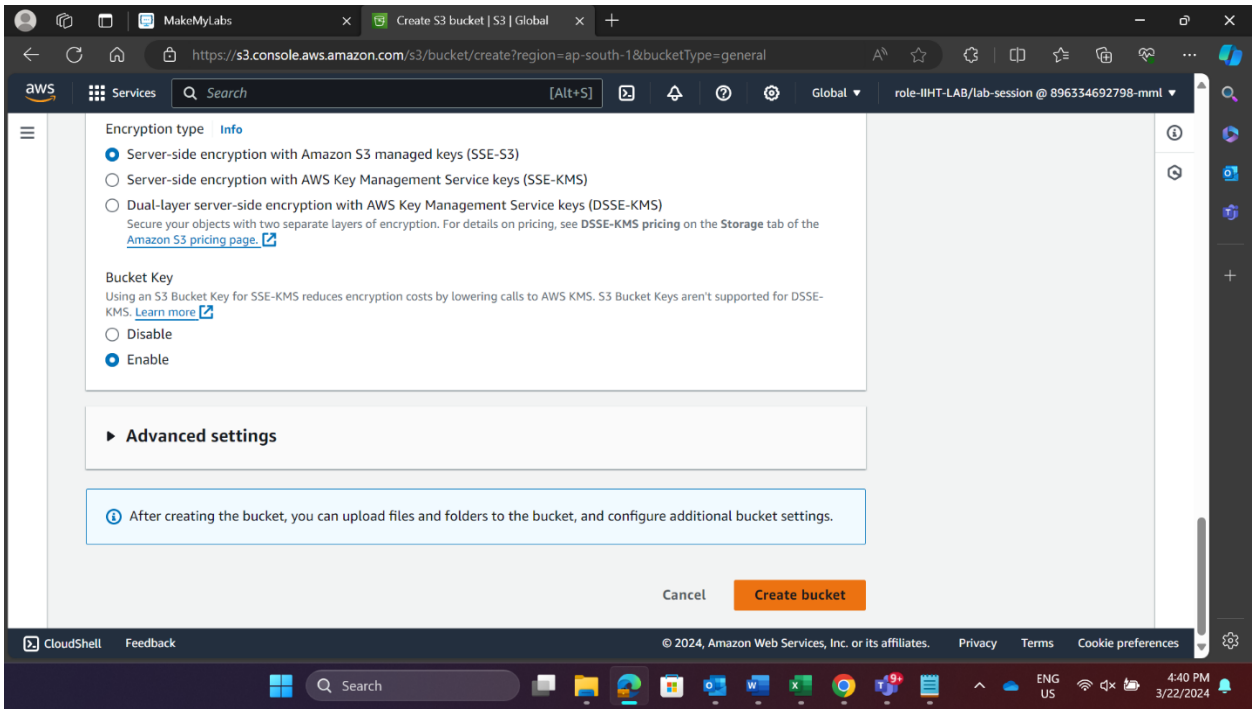




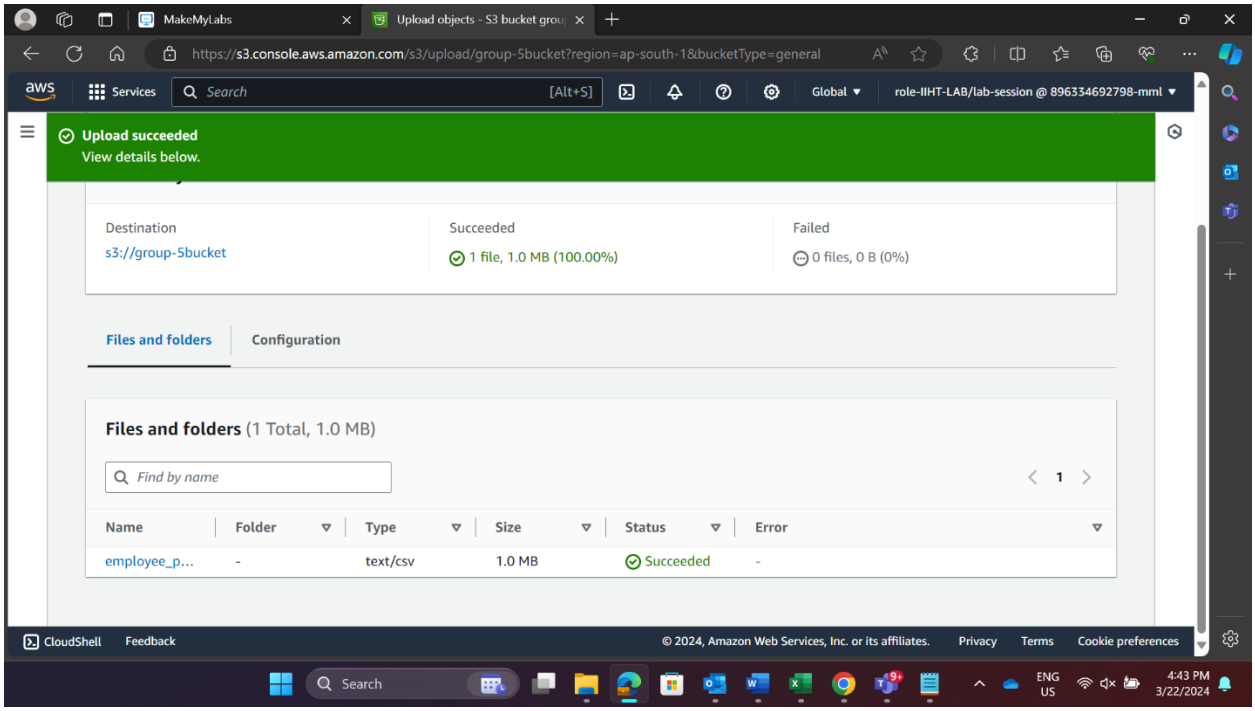
Step 2: Configure whether **ACLs (Access Control Lists) are enabled** for the bucket. You can make the bucket accessible to all or restrict access based on your requirements.



Step 3: Selecting SSE with Amazon S3 managed keys and Enabling bucket key.

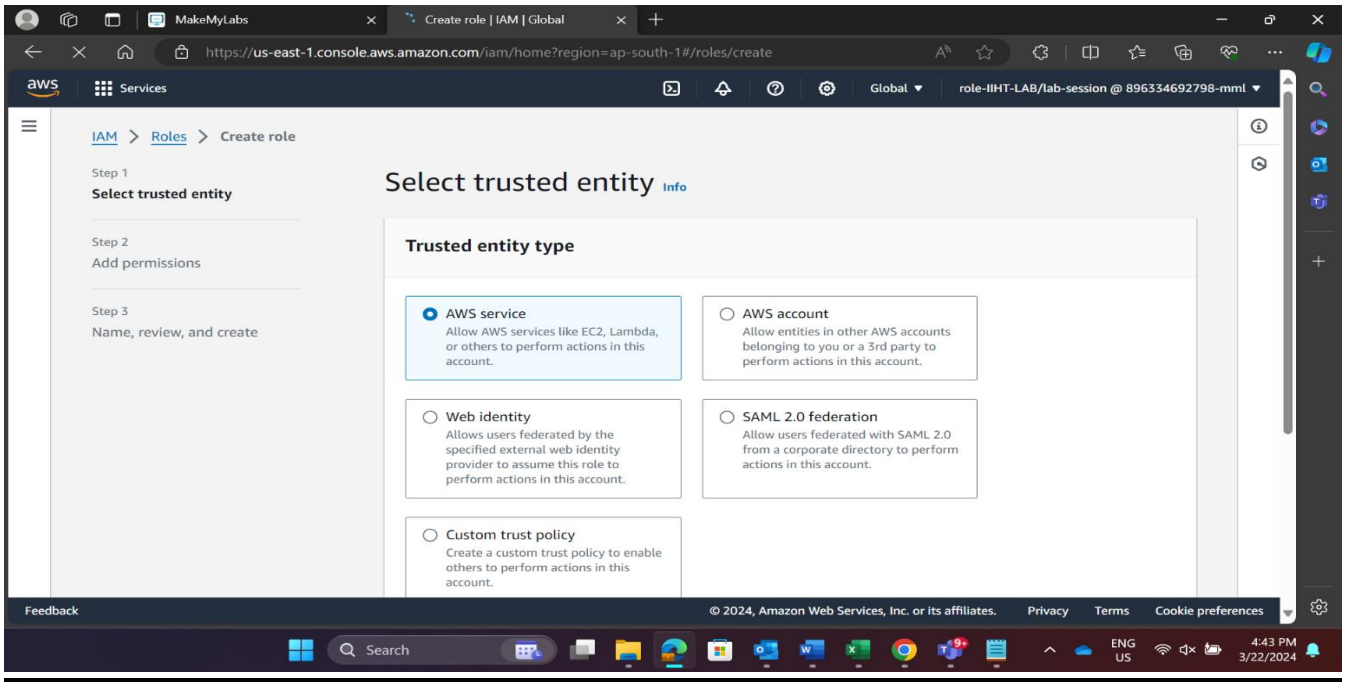


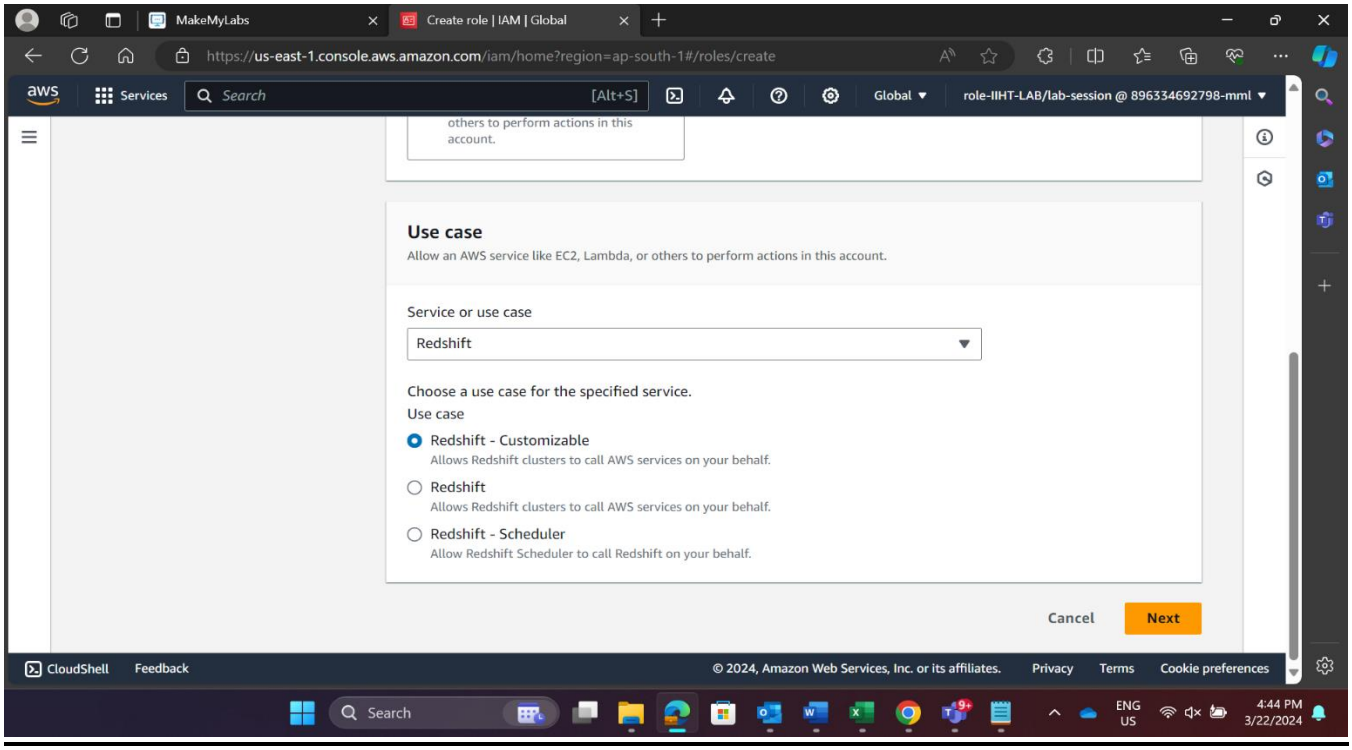
Step 4: After creating bucket, upload the file in bucket.

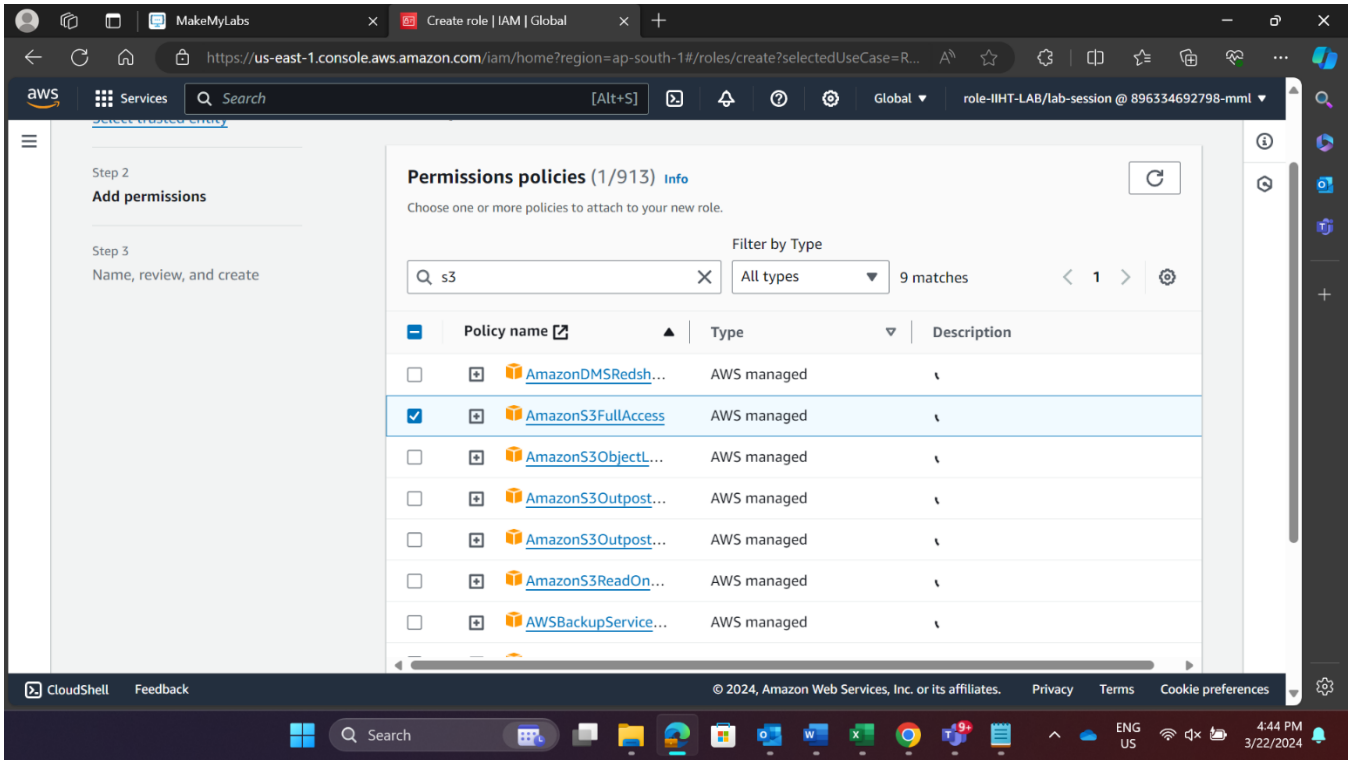


IAM ROLE ASSOCIATION

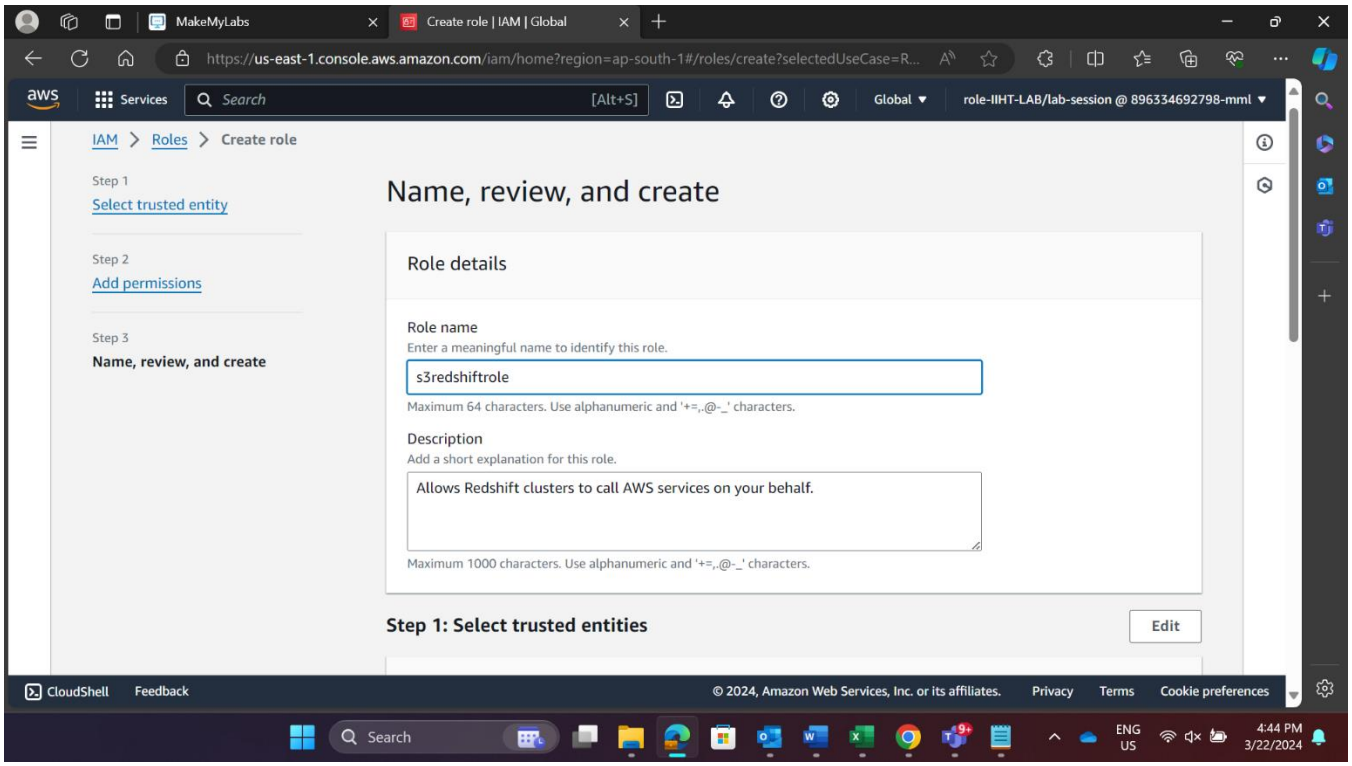
Step 1: Navigate the IAM console and create a role, choose redshift that will assume the role and then choose S3 as the service that will use this role.

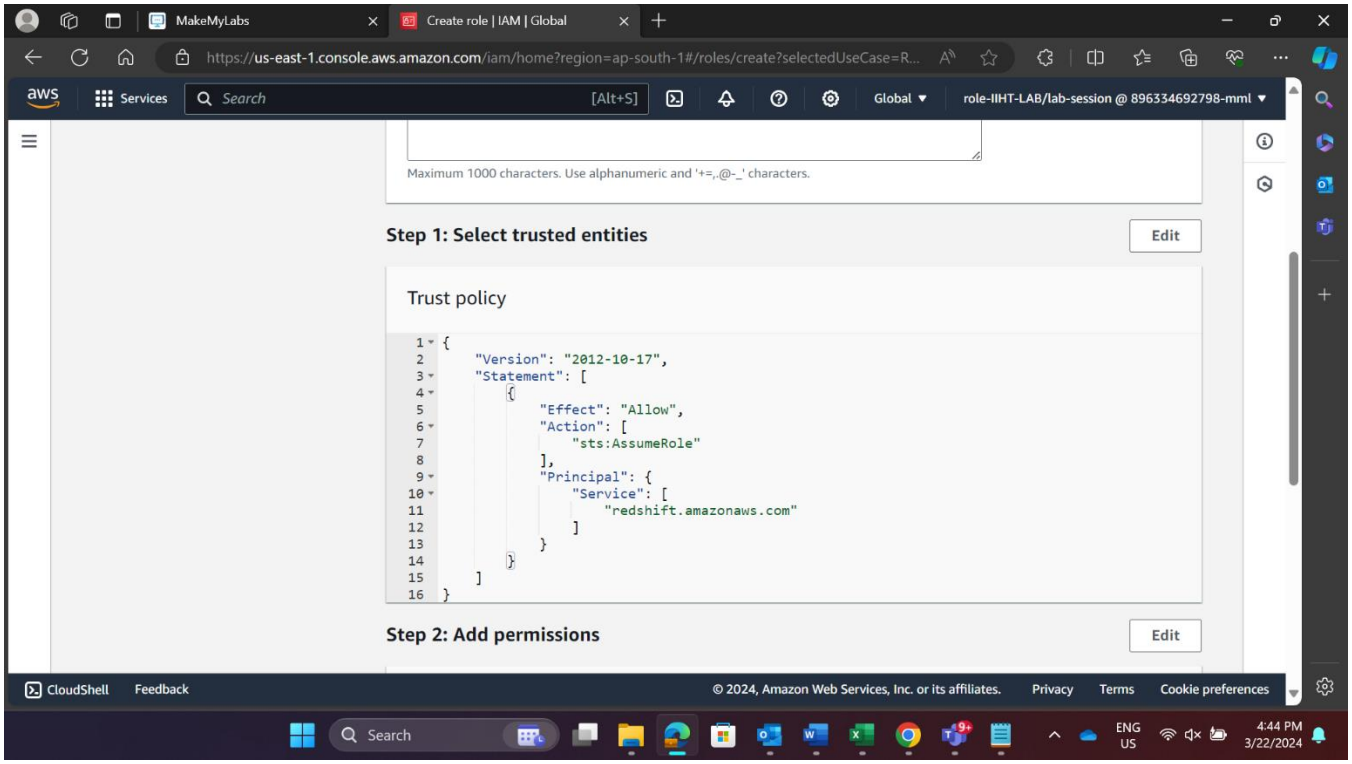


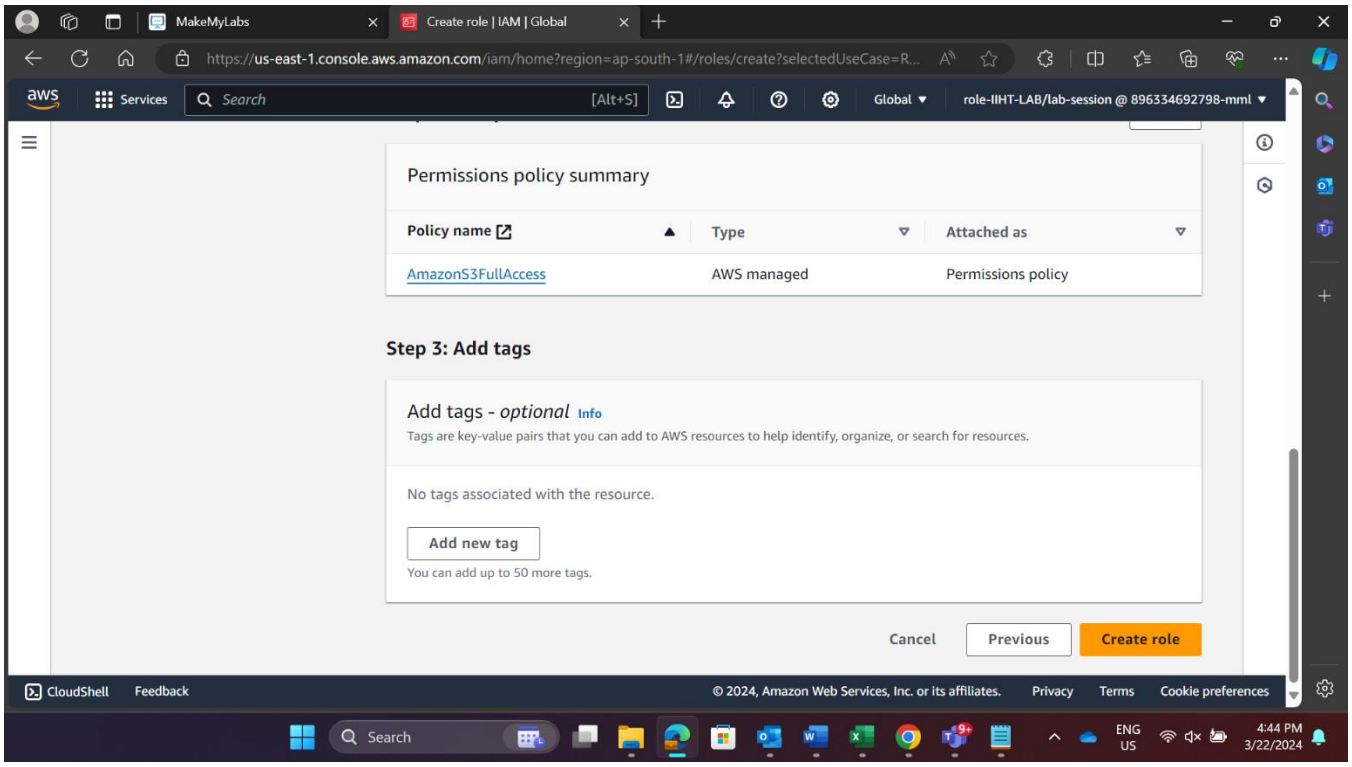




Step 2: Provide the details and finally create the role.

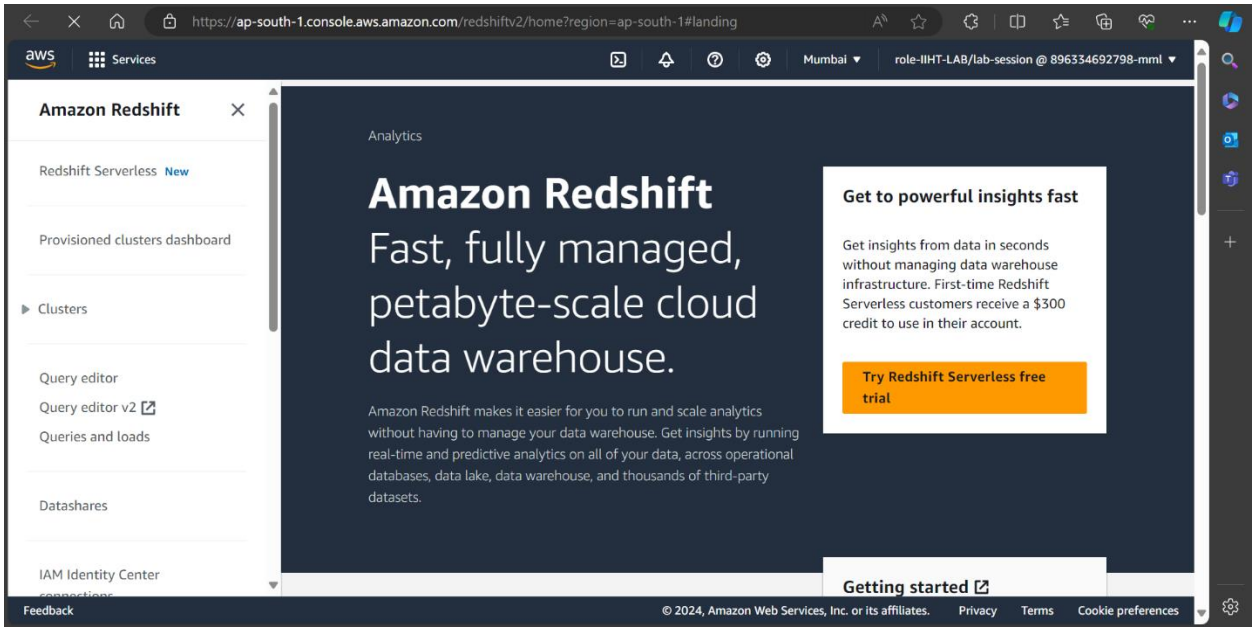



MakeMyLabs
Create role | IAM | Global
https://us-east-1.console.aws.amazon.com/iam/home?region=ap-south-1#/roles/create?selectedUseCaseR...
Services
Search
[Alt+S]
Global
role-IIHT-LAB/lab-session @ 896334692798-mml
Maximum 1000 characters. Use alphanumeric and '+=,.@-_' characters.
Step 1: Select trusted entities
Edit
Trust policy
{
"Version": "2012-10-17",
"Statement": [
{
"Effect": "Allow",
"Action": [
"sts:AssumeRole"
],
"Principal": {
"Service": [
"redshift.amazonaws.com"
]
}
}
]
}
Step 2: Add permissions
Edit
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:44 PM
3/22/2024

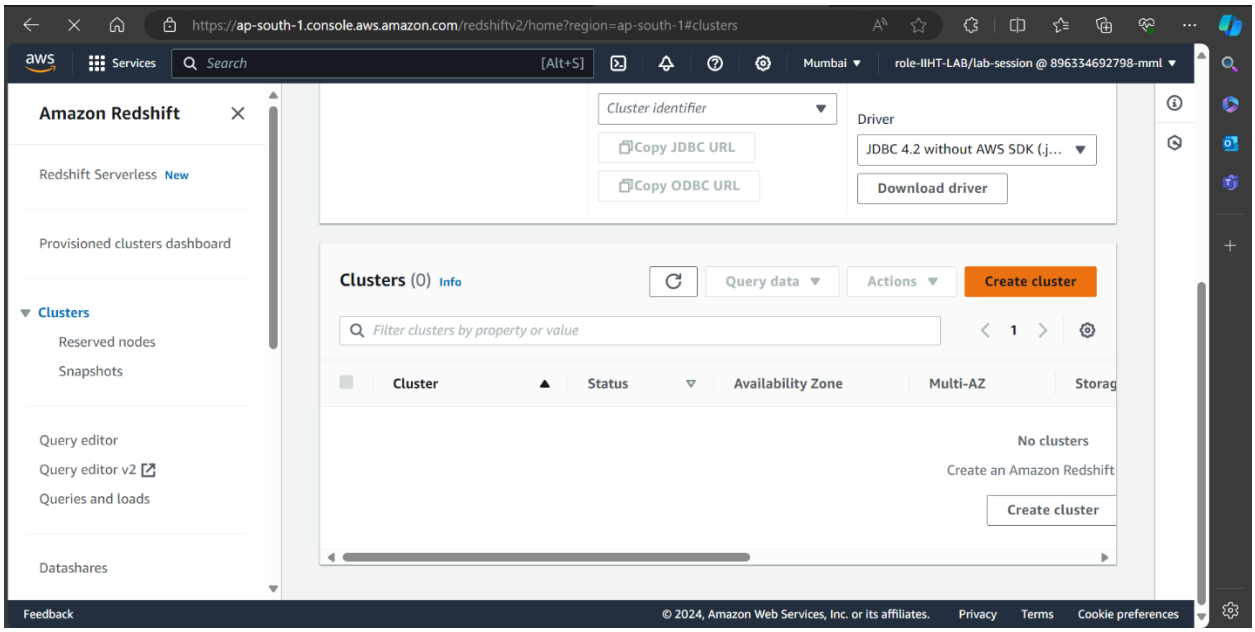

MakeMyLabs
Create role | IAM | Global
https://us-east-1.console.aws.amazon.com/iam/home?region=ap-south-1#/roles/create?selectedUseCaseR...
Services
Search
[Alt+S]
Global
role-IIHT-LAB/lab-session @ 896334692798-mml
Permissions policy summary
Policy name
Type
Attached as
AmazonS3FullAccess
AWS managed
Permissions policy
Step 3: Add tags
Add tags - optional Info
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.
No tags associated with the resource.
Add new tag
You can add up to 50 more tags.
Cancel
Previous
Create role
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
4:44 PM
3/22/2024

REDSHIFT CREATION

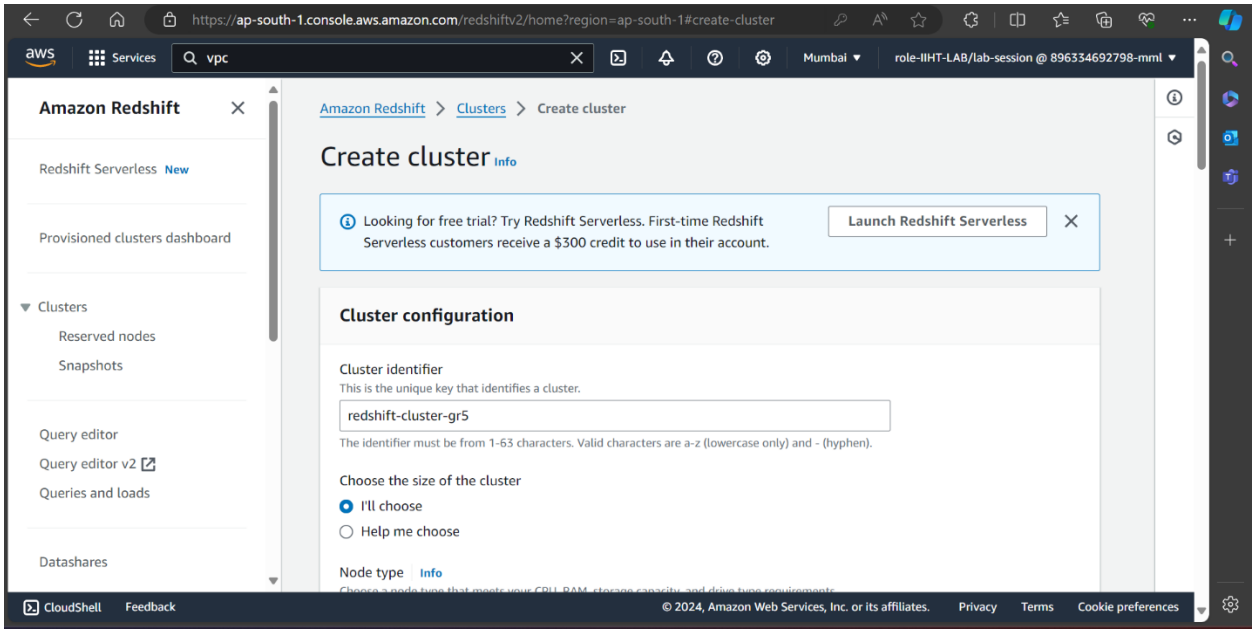
Step1: Sign in to the AWS Management Console and open the Amazon Redshift console and create the cluster. At upper right, choose the AWS Region where you want to create the cluster.



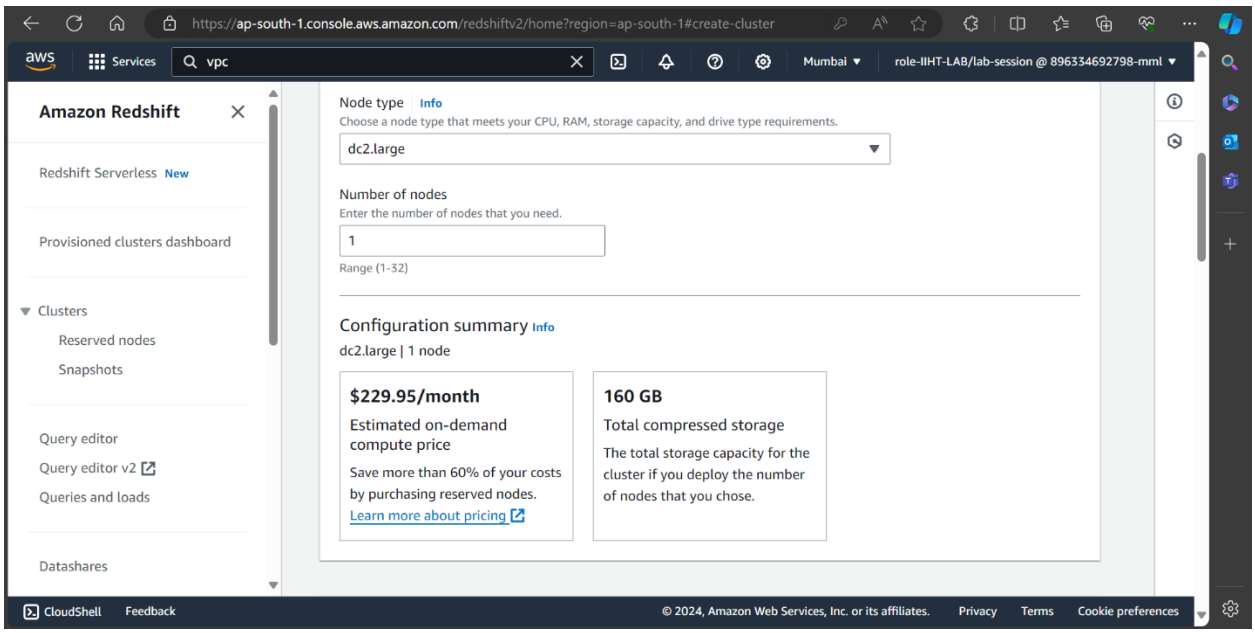
Choose **Clusters**, then choose **Create cluster** and start creating the cluster.



Step 2: In the Cluster configuration section, specify values for Cluster identifier and size of the cluster.

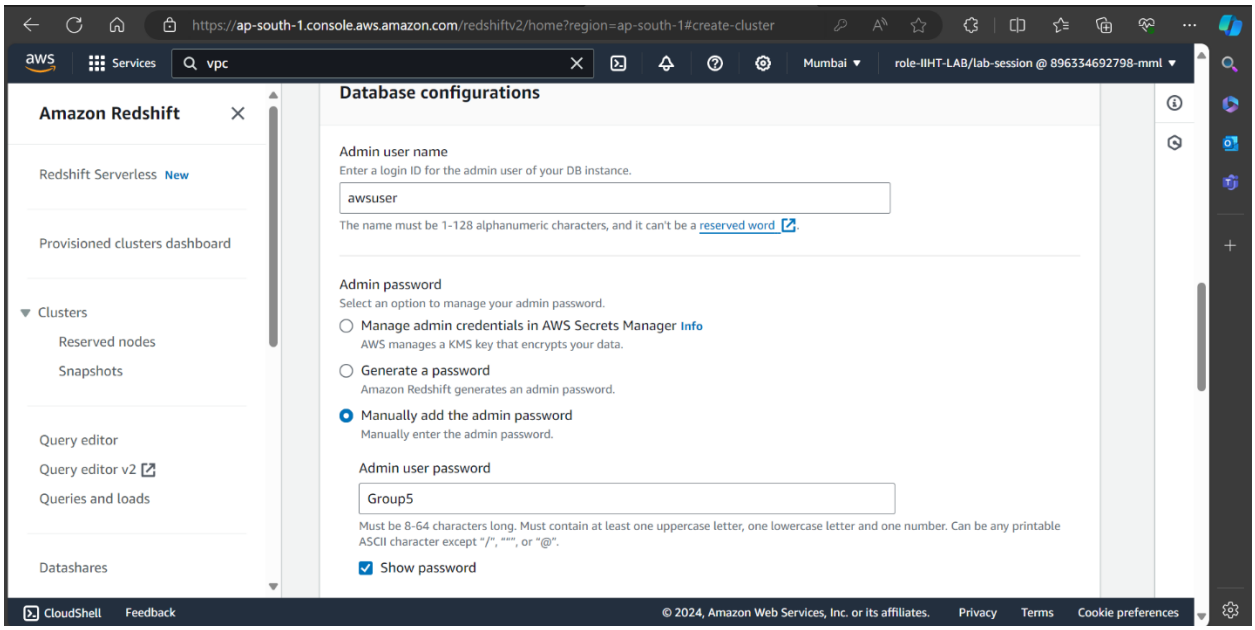


Then choose the **Node type** and number of **Nodes** to size your cluster. Choose dc2.large Node type and 1 for Nodes.

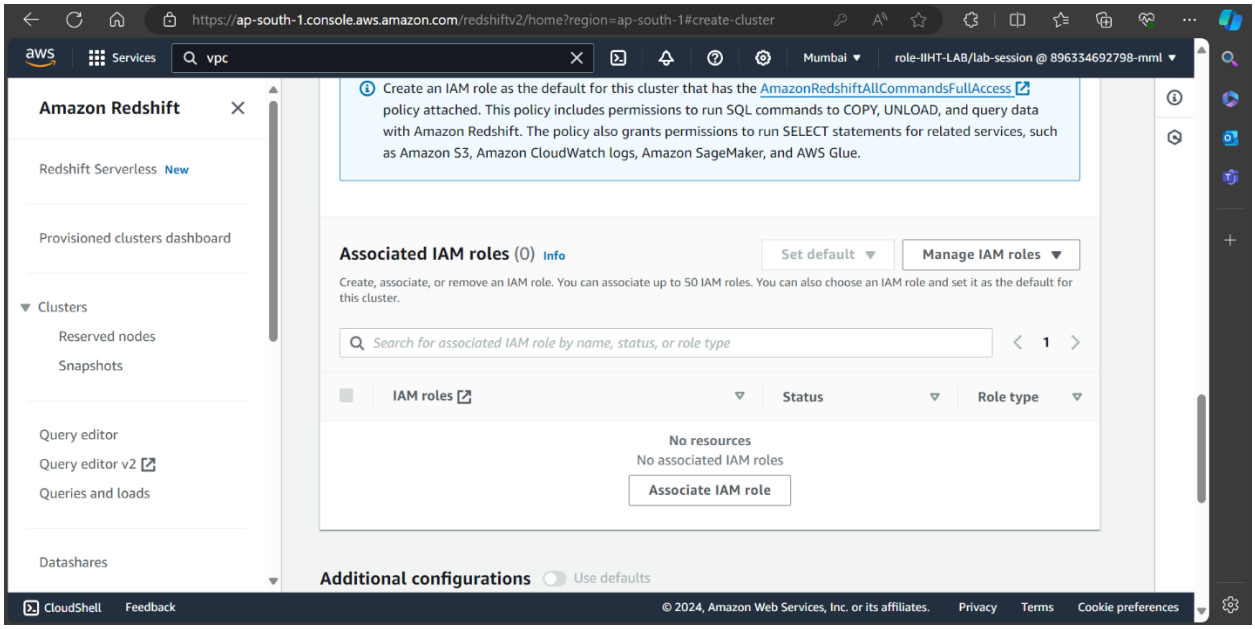


Step 3: In the Database configuration section, specify a value for **Admin username**. Choose Admin password as follows:

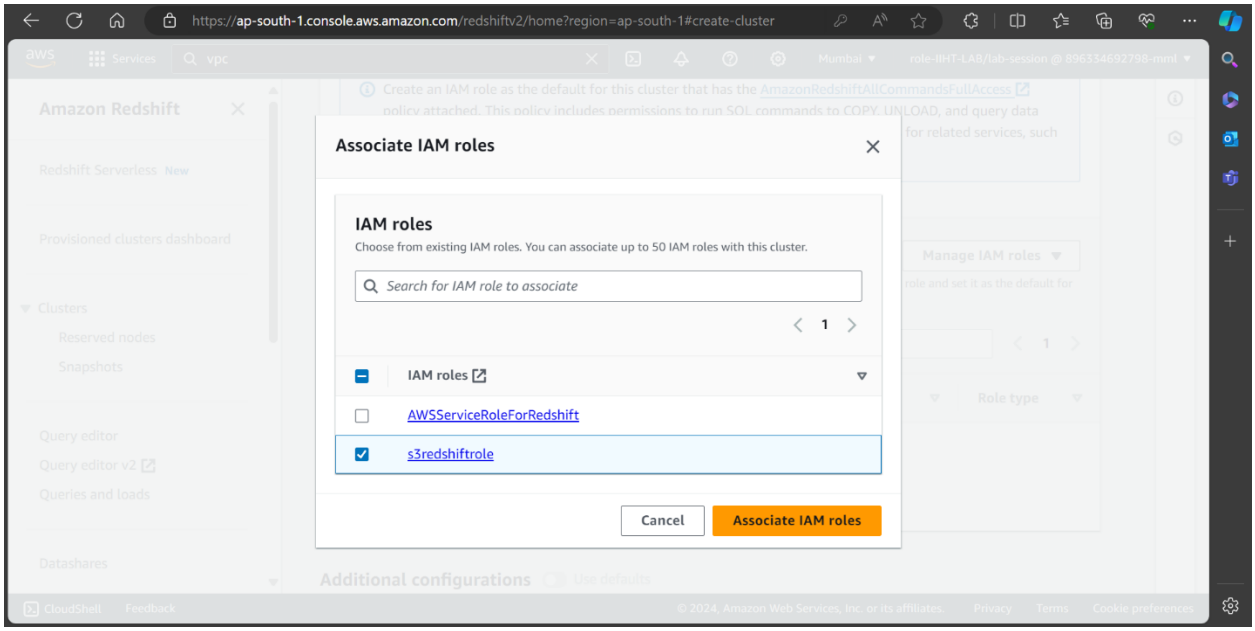
Manually add an admin password (Use your own password).



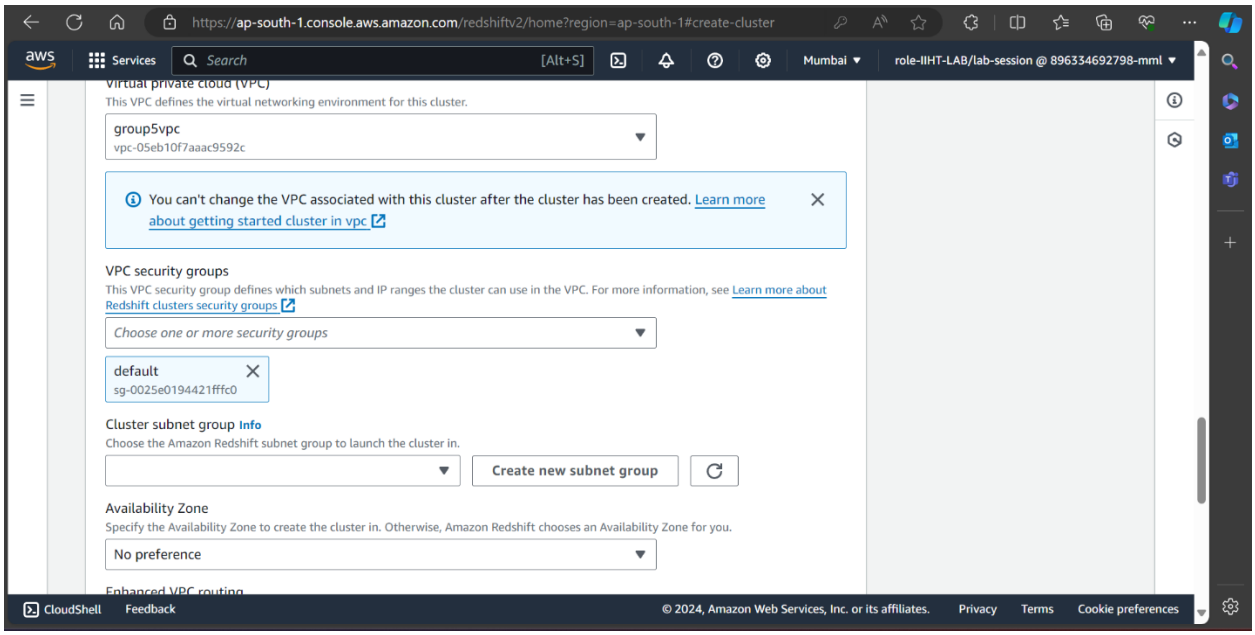
Step 4: Create an IAM role and choose the one you have created for your cluster. Under Cluster permissions, for Manage IAM roles.



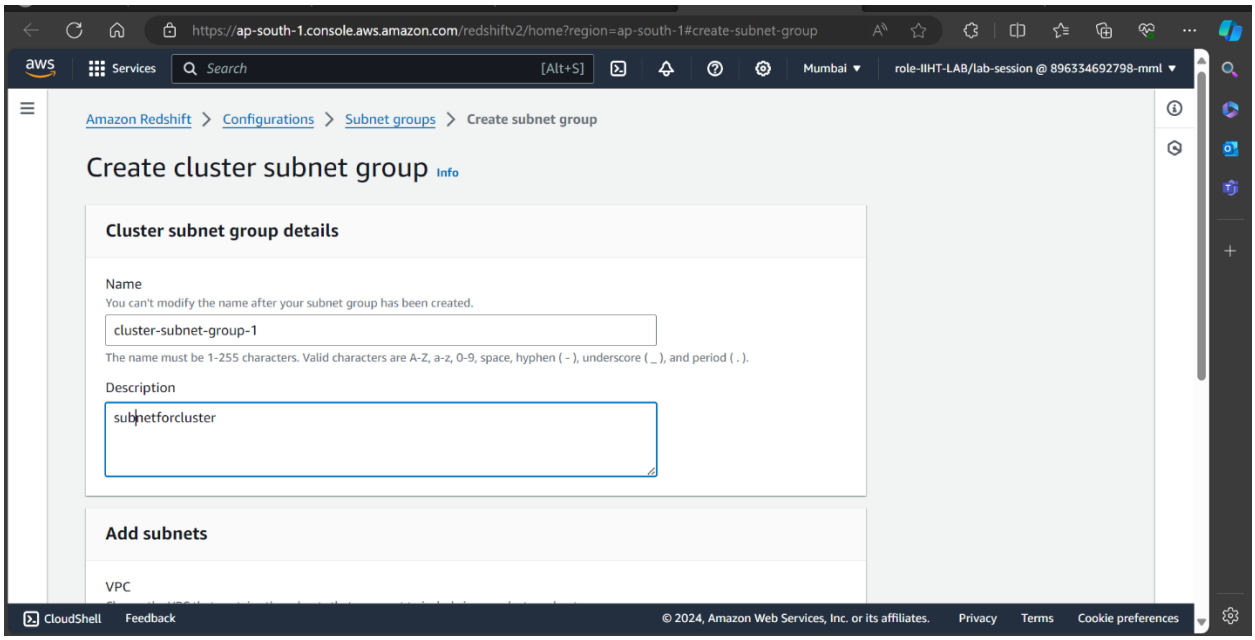
Specify an Amazon S3 bucket for the IAM role to access then Associate IAM role.

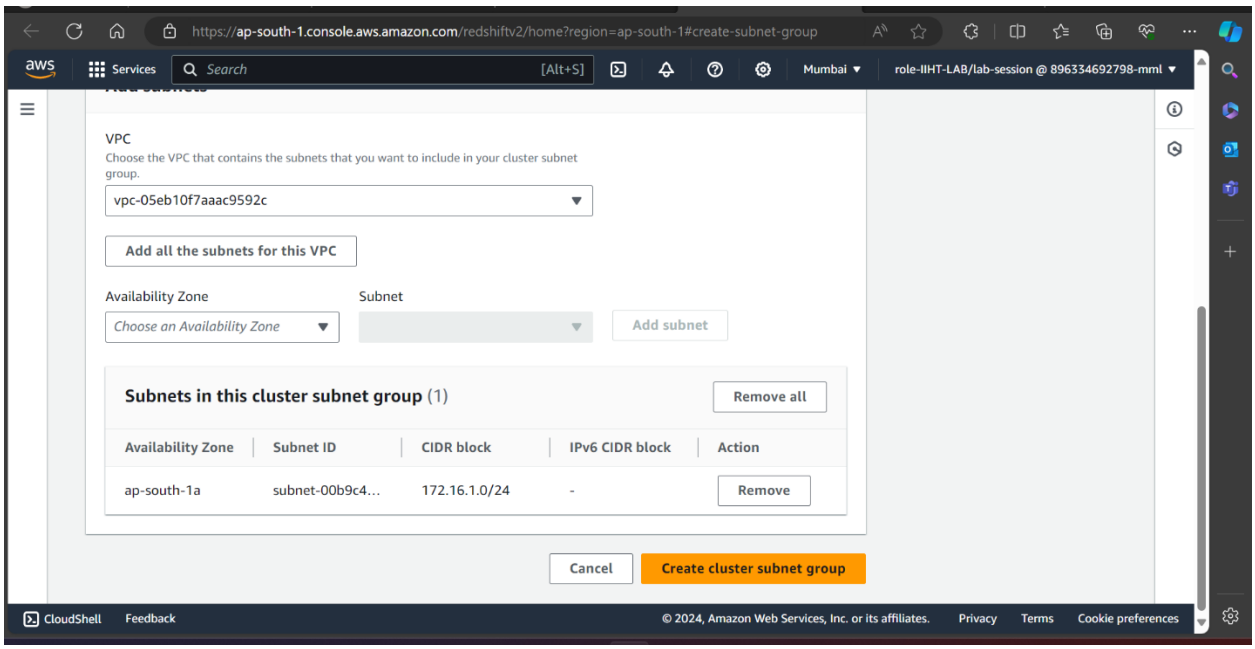


Step 5: Choose the VPC that you have created previously and set the VPC security groups as default.

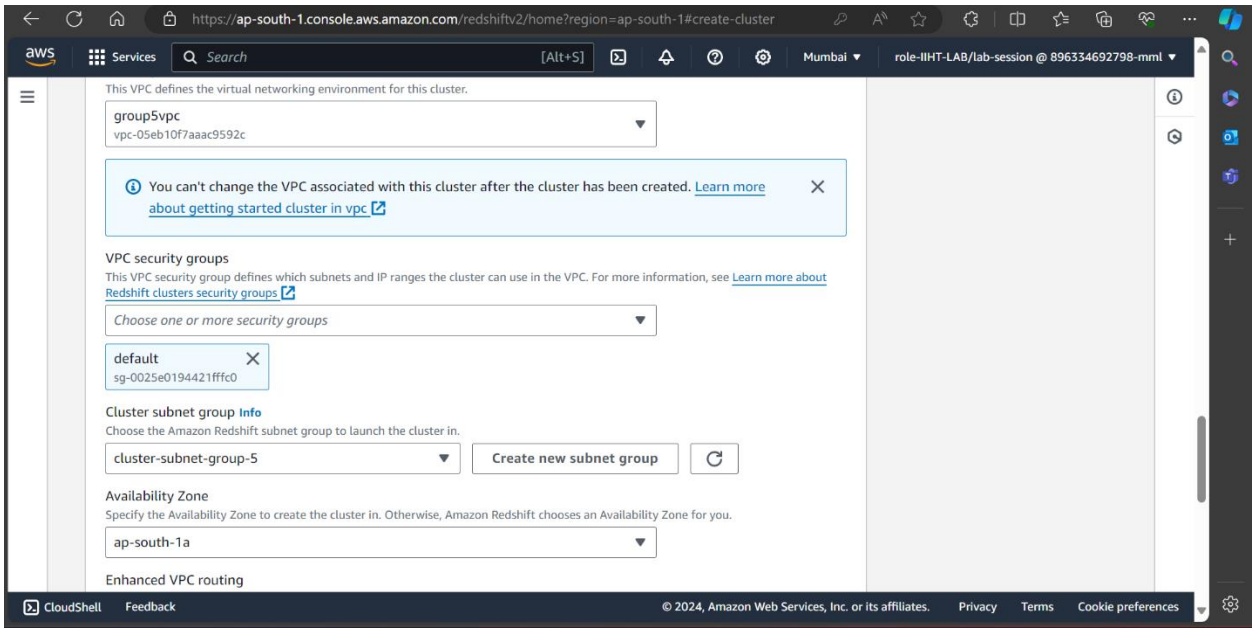


Step 6: Create a cluster subnet group. A cluster subnet group allows you to specify a set of subnets in your VPC.

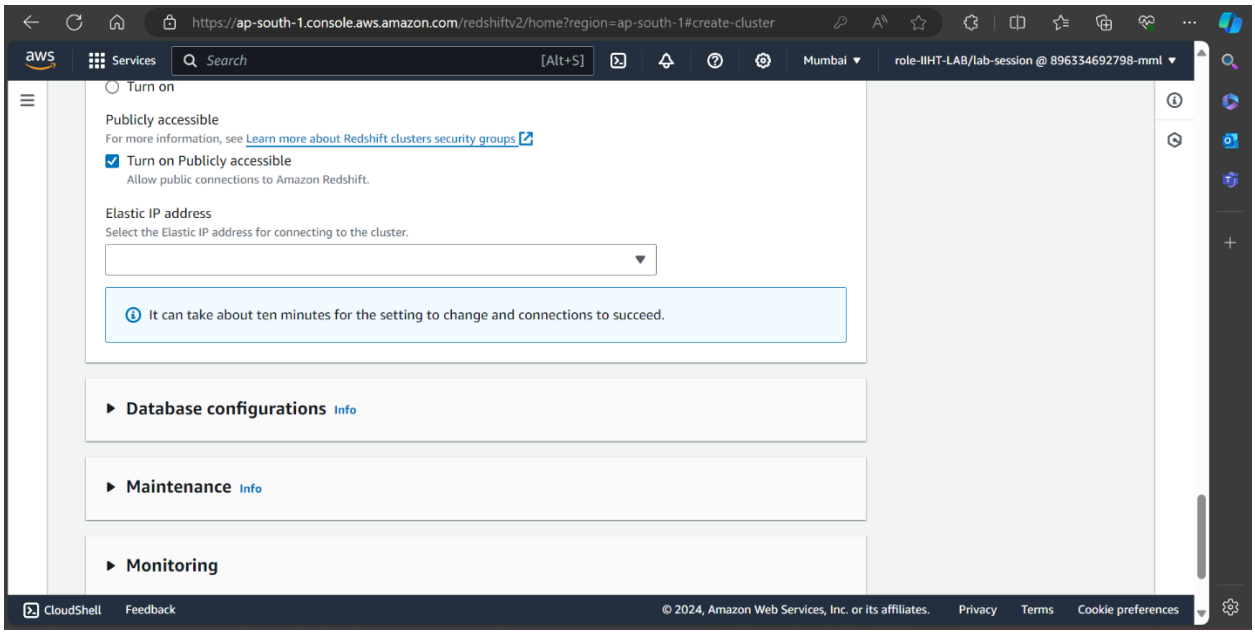




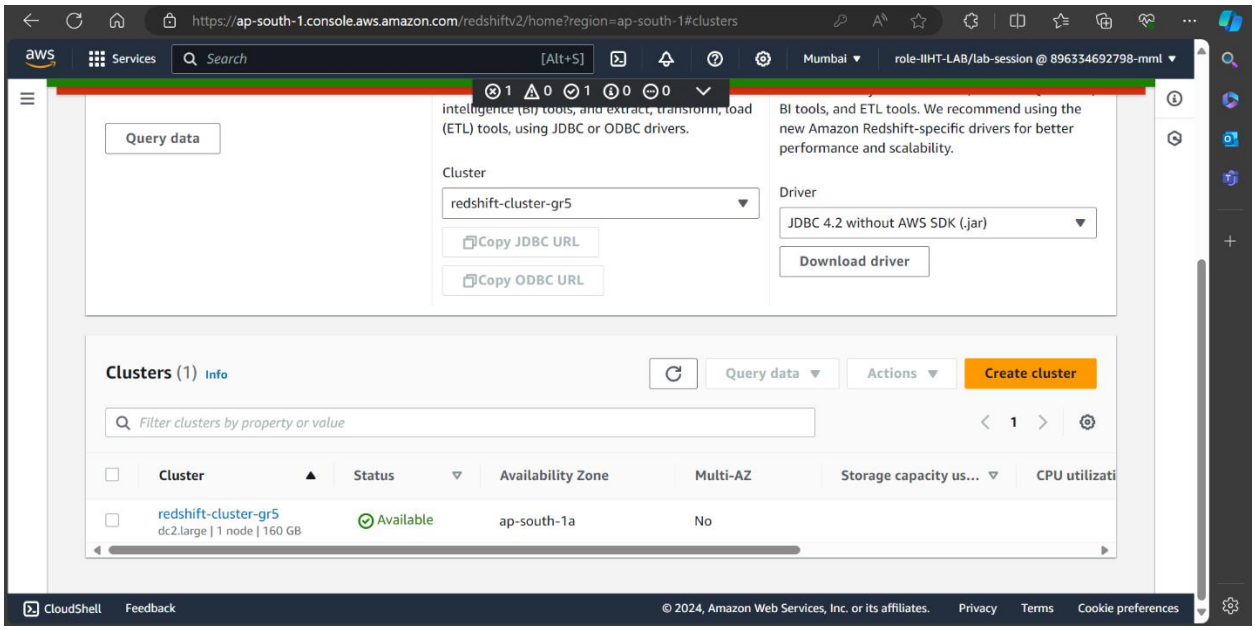
Step 7: Choose the Availability Zone in which location should subnet run.



Turn on **"publicly accessible"**. This allows public connection to redshift.

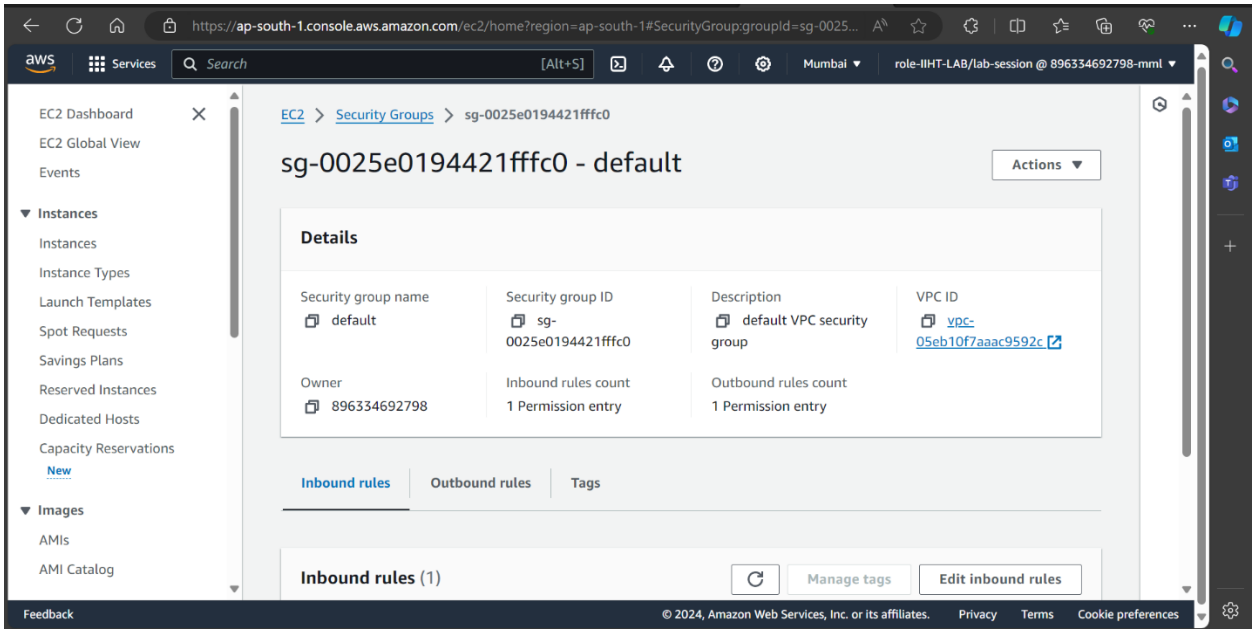


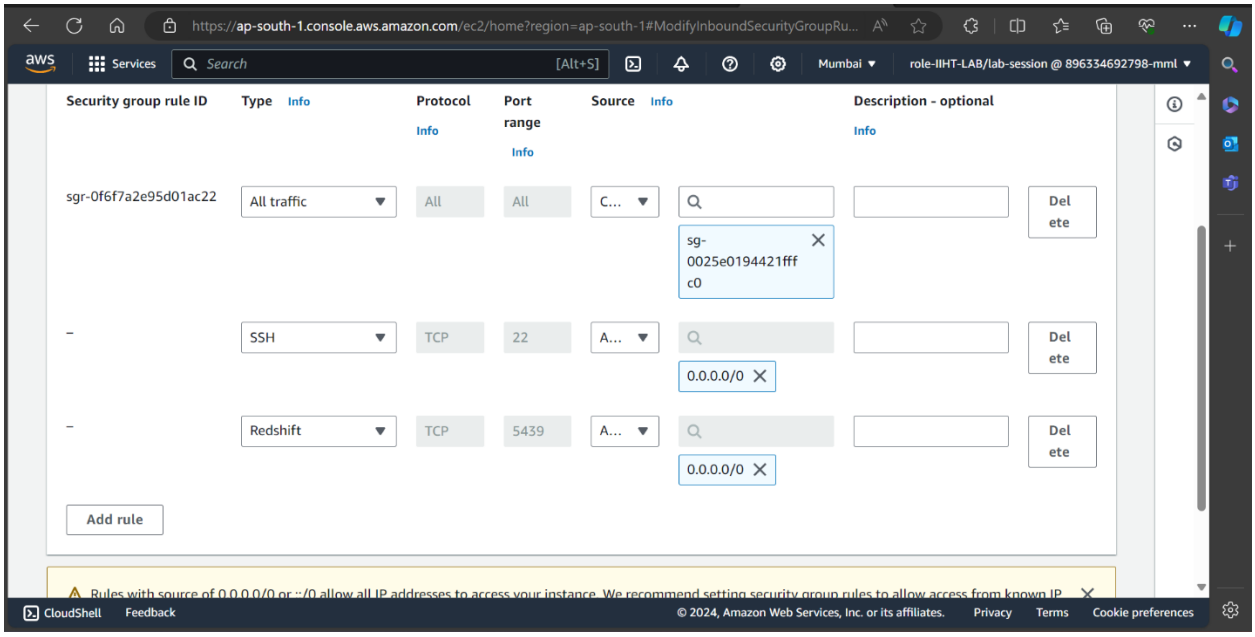
After completing the configuration setup “Create the cluster”



Step 8: Edit inbound rules under security groups and grant inbound access to cluster.

To access from an Amazon EC2 external, add a rule to the security group attached to your cluster that allows inbound traffic.





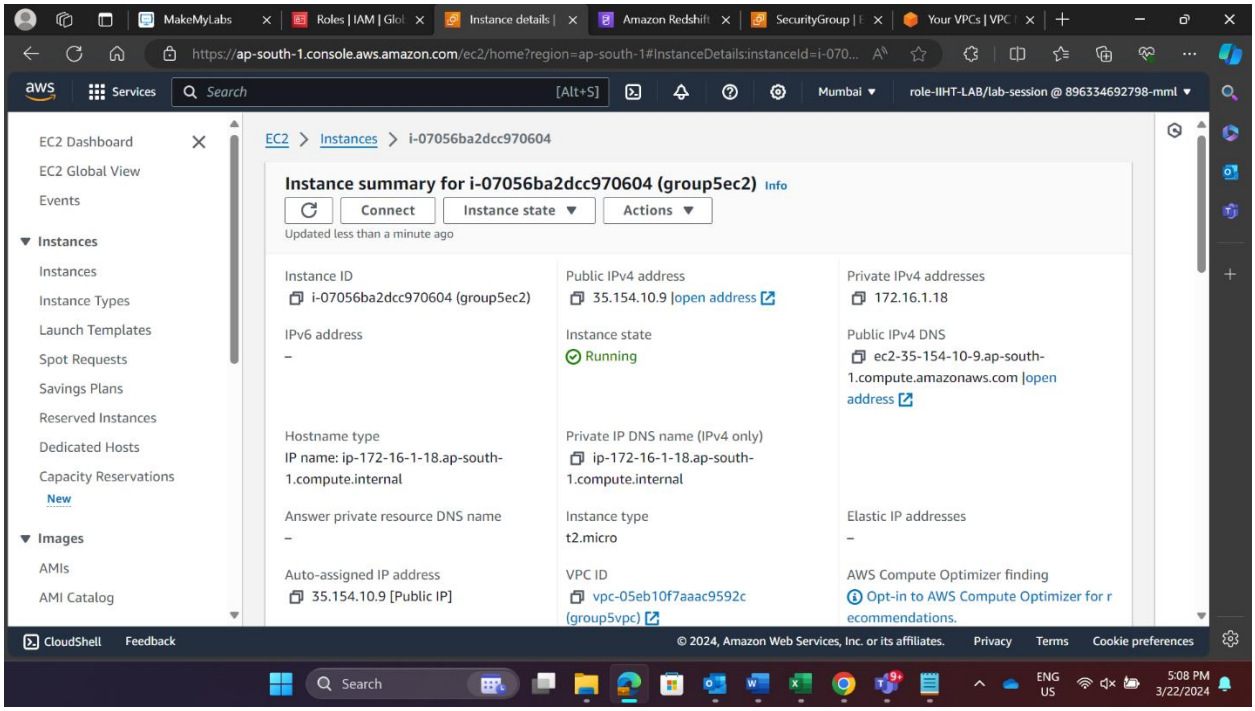
Step 9: After creating your cluster, you can immediately run queries using the Amazon Redshift console.

To query databases through Amazon Redshift cluster,

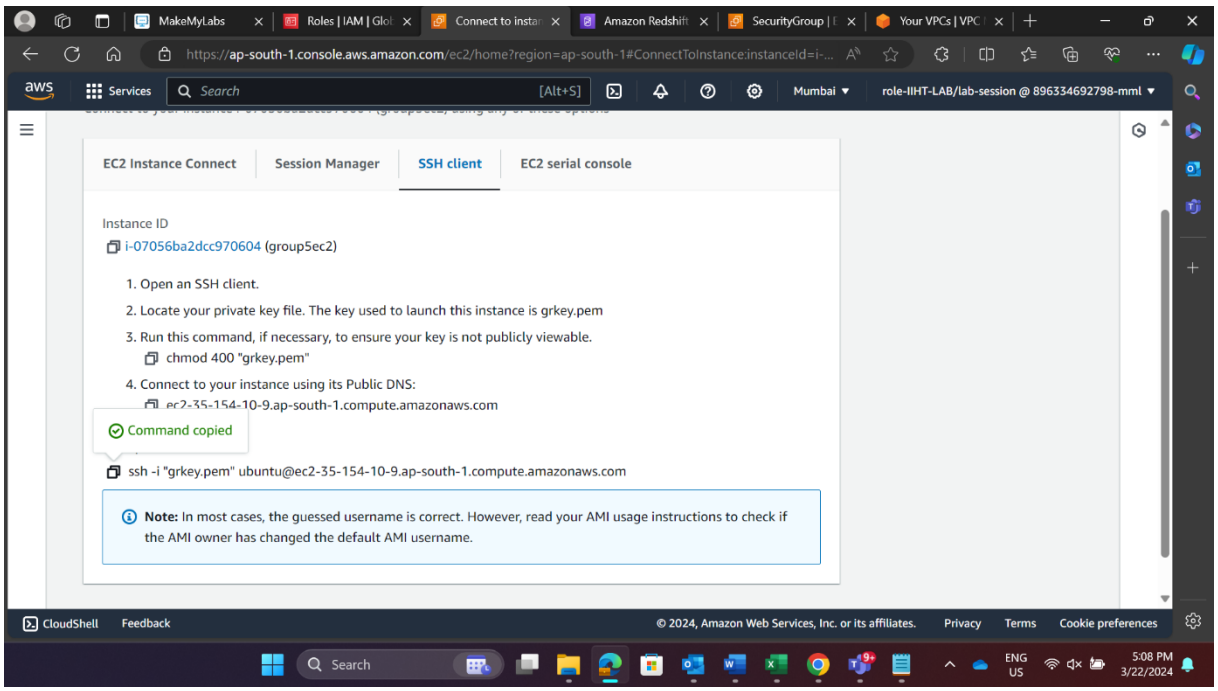
- Connect to your cluster and run queries on the AWS Management Console with one of the query editors.
- Connect to your cluster through an SQL client tool, such as SQL Workbench

Connection to the EC2 instance

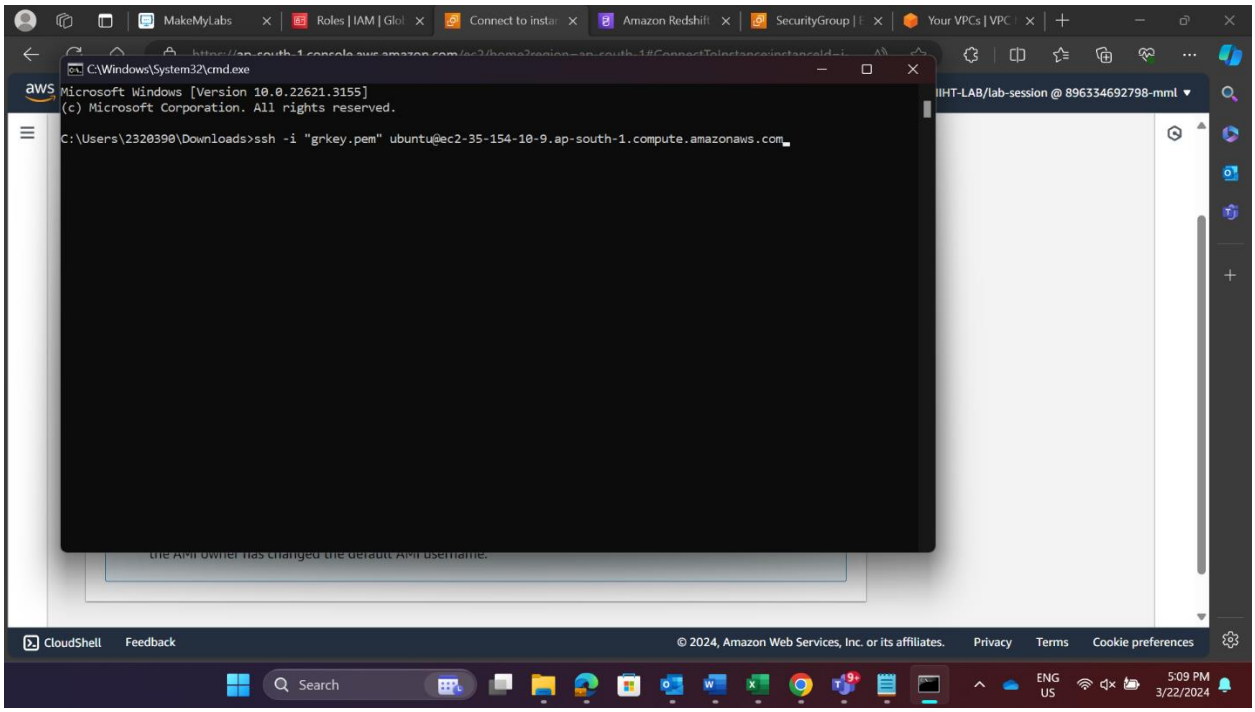
Step 1: Goto your EC2 dashboard and select the instance and click on Connect.

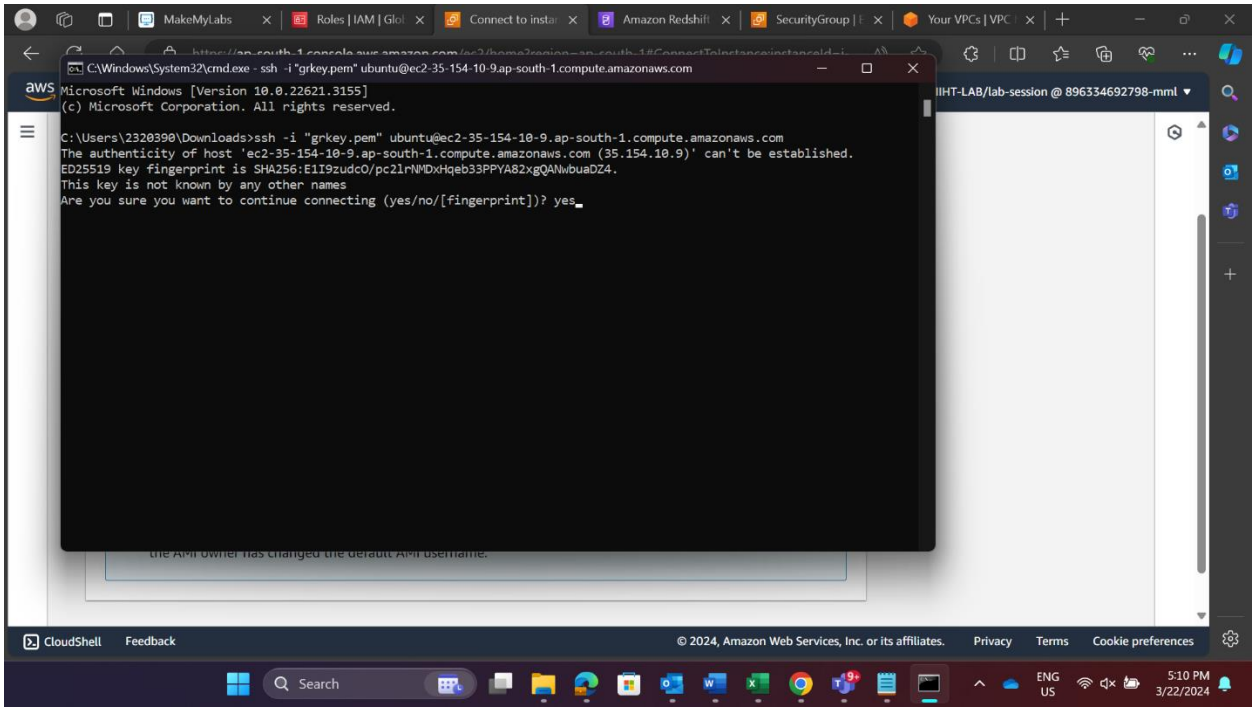


Step 2: Use SSH Client option to connect to your instance and copy the command given as an example.



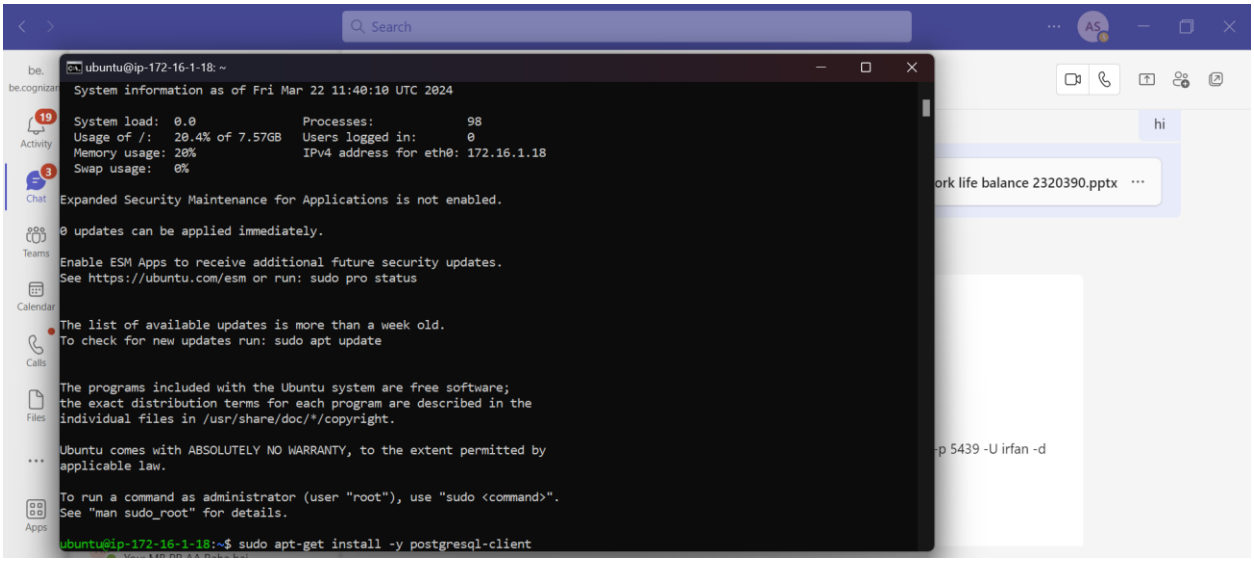
Step 3: Open a terminal or the command prompt on the local machine and paste the command copied in the previous step. Make sure that you are inside the directory where your key pair generated during the instance creation is stored before executing the command.





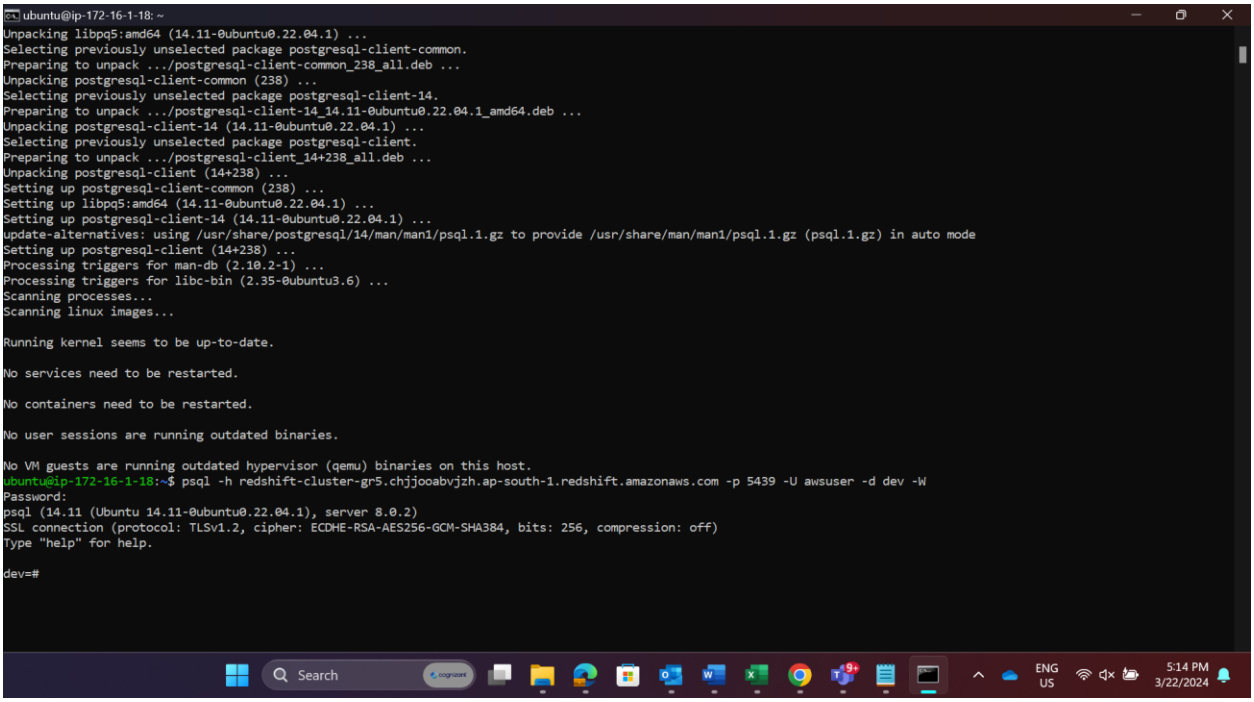
Step 4: Once the above command is successfully executed you will be logged into your ubuntu server. The next step is to install PostgreSQL on your virtual server by running command '**sudo apt-get install -y PostgreSQL-client'**

PostgreSQL is installed successfully on your ubuntu server.

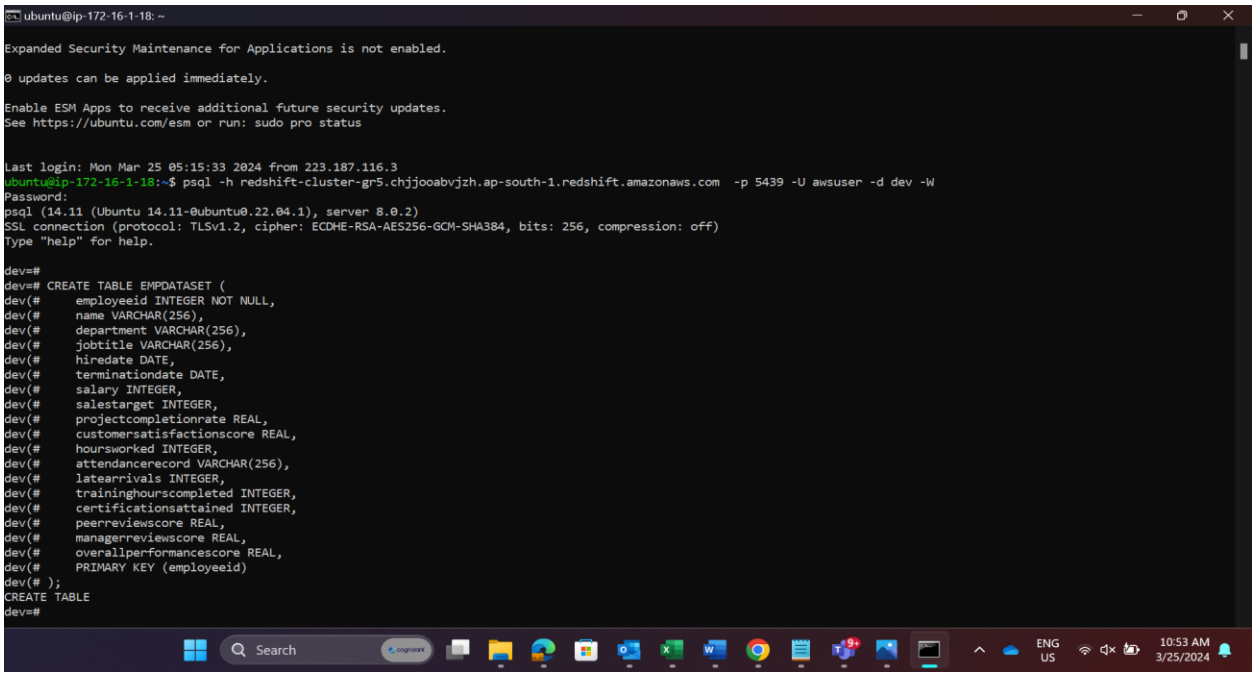


Step 5: After successfully installing the PostgreSQL on your server, the next step is connecting to the redshift cluster. This is done by using the following command.

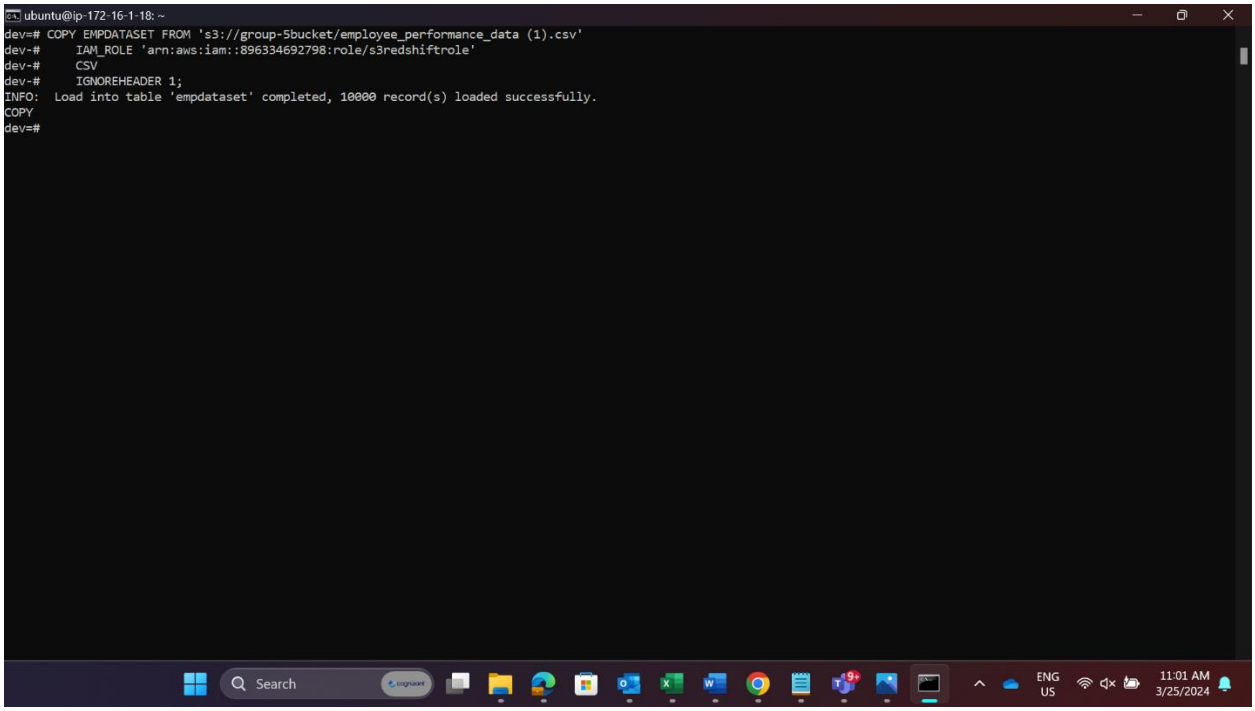
'psql -h redshift-cluster-1.cackughiw5hj.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d dev –W'



Step 6: Once the above command is successfully executed, you can access your database. Now, create the required table (EMPDATASET) using the CREATE statement and define the column names and its attributes.



Step 7: The table has been created successfully. Now the next step is to load the data present in your S3 bucket into the table you have created in the redshift cluster. This is done by using the copy command.

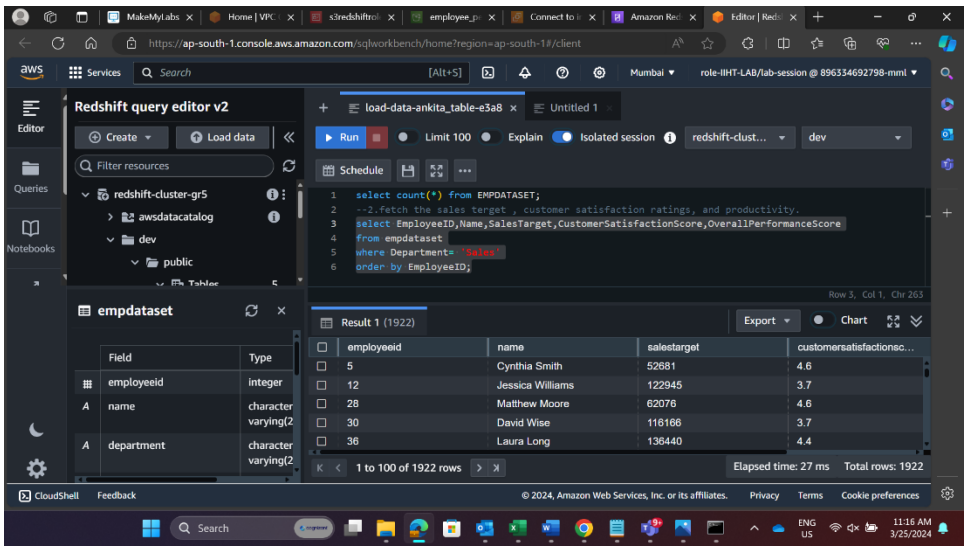


The data has been successfully loaded into the table.

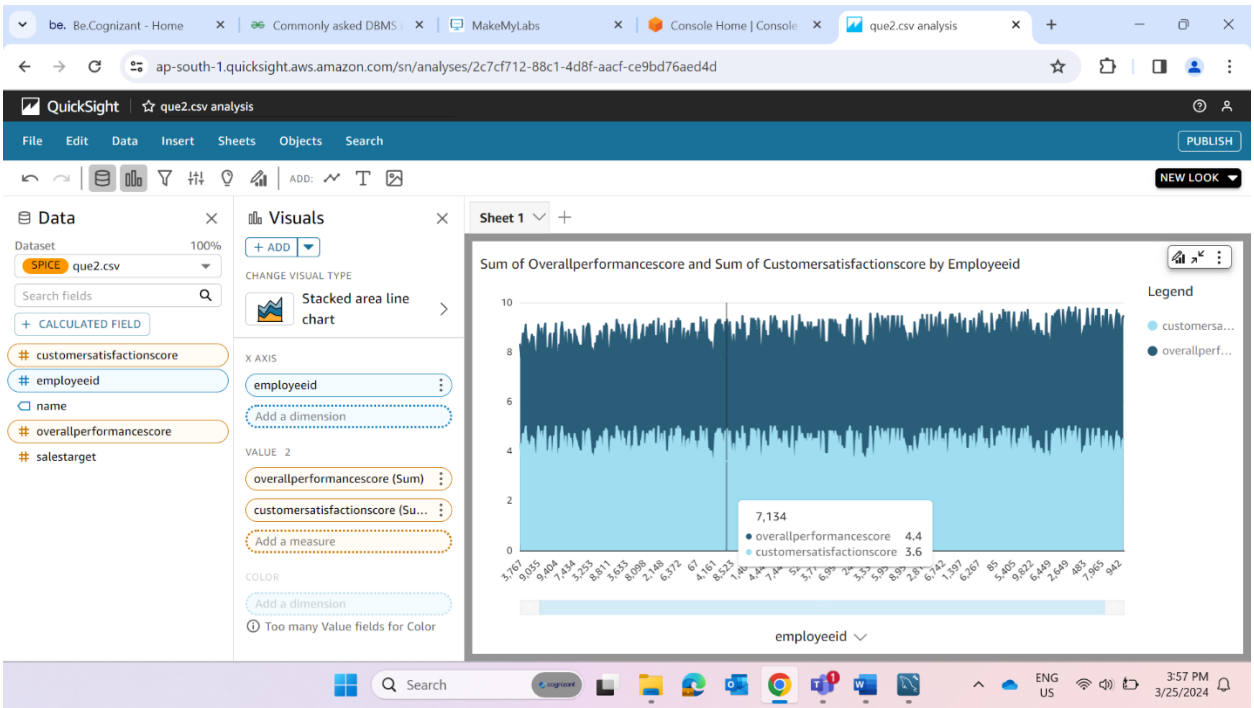
Redshift Queries Generation and Visualization

1. Fetch the sales target, customer satisfaction ratings, and productivity for all the employees of the sales department.

a. Query execution.

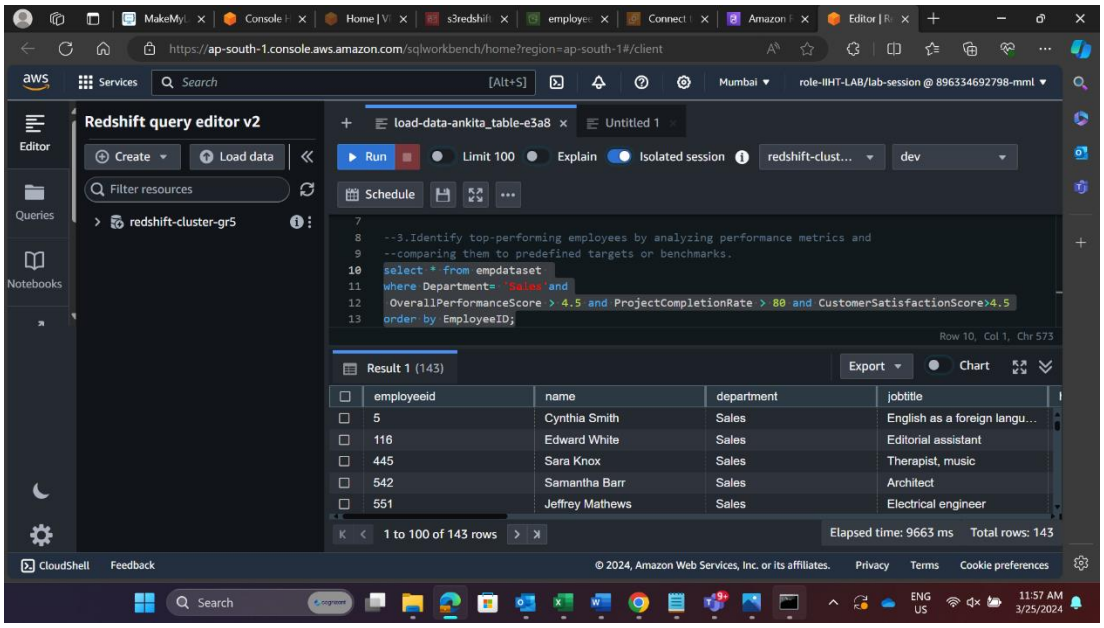


b. Visualization

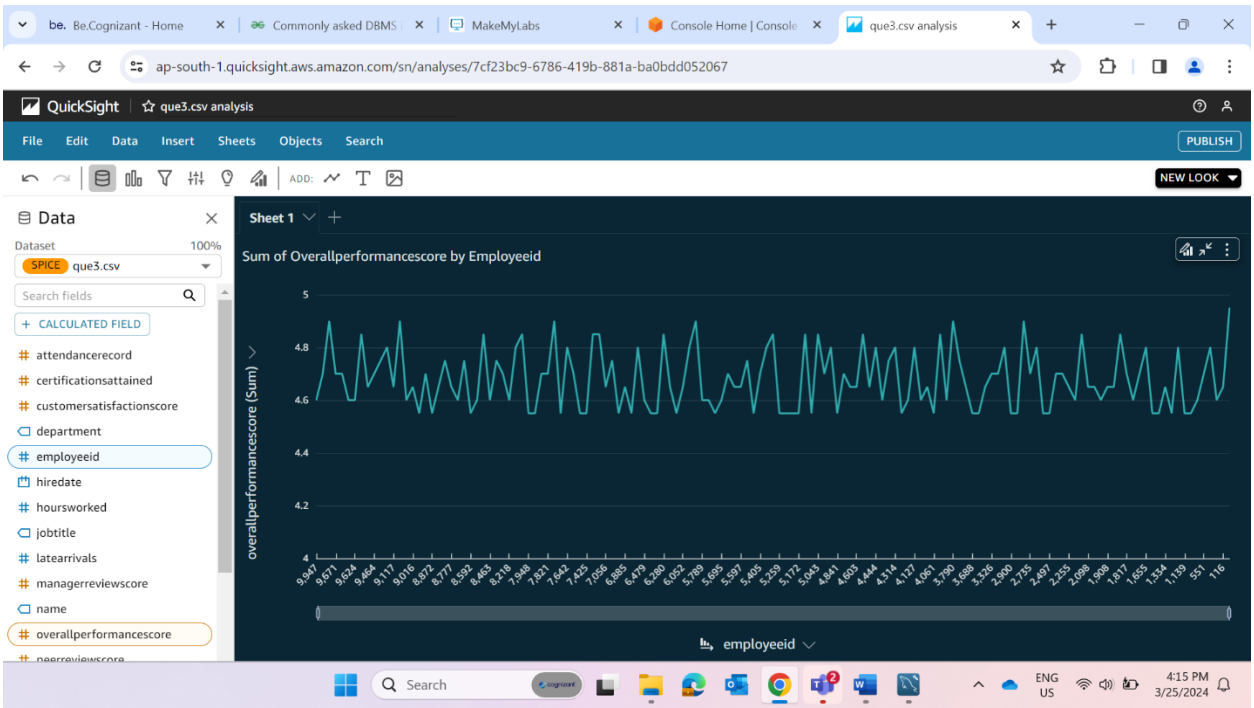


2. Identify top-performing employees by analyzing performance metrics and comparing them to predefined targets or benchmarks.

a. Query Execution

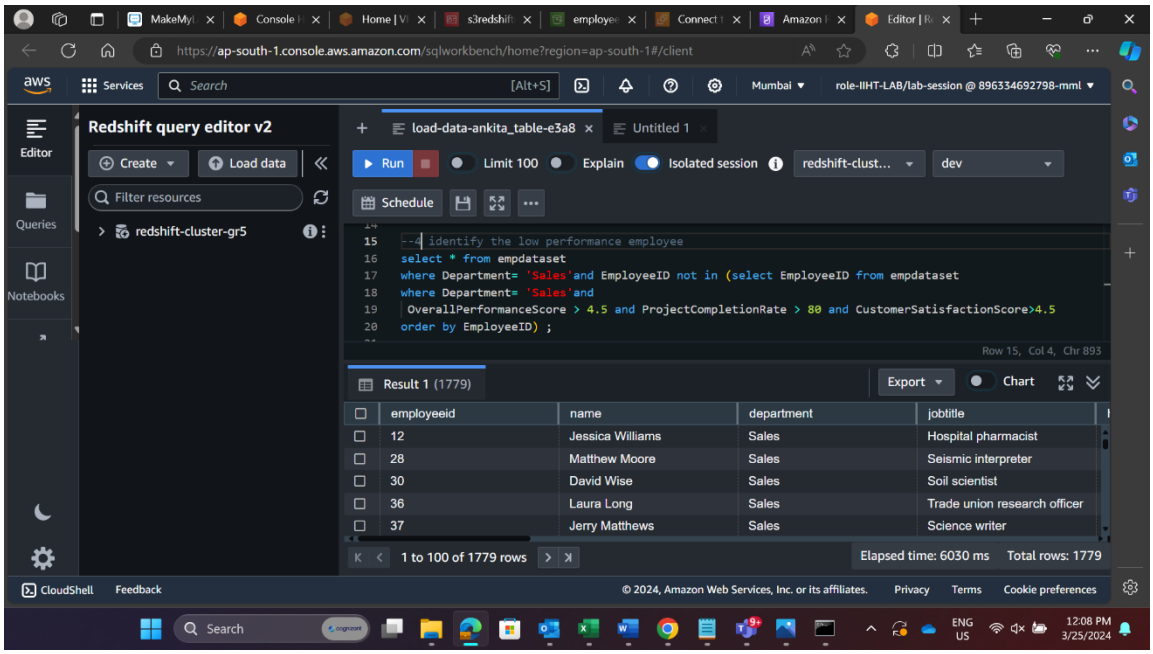


b. Visualization

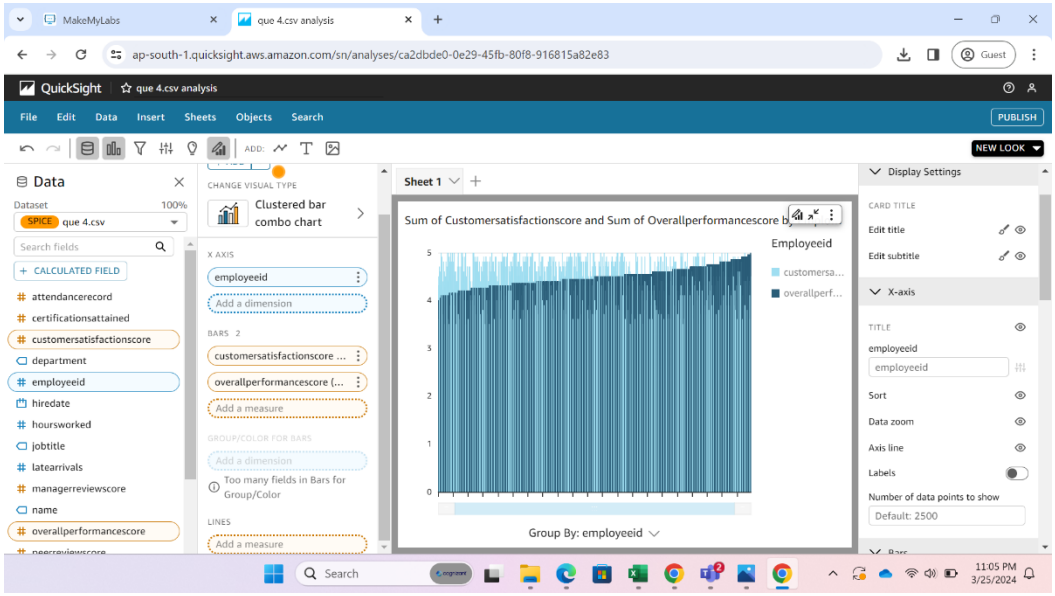


3. Identify the employees having low performance in the sales department.

a. Query Execution

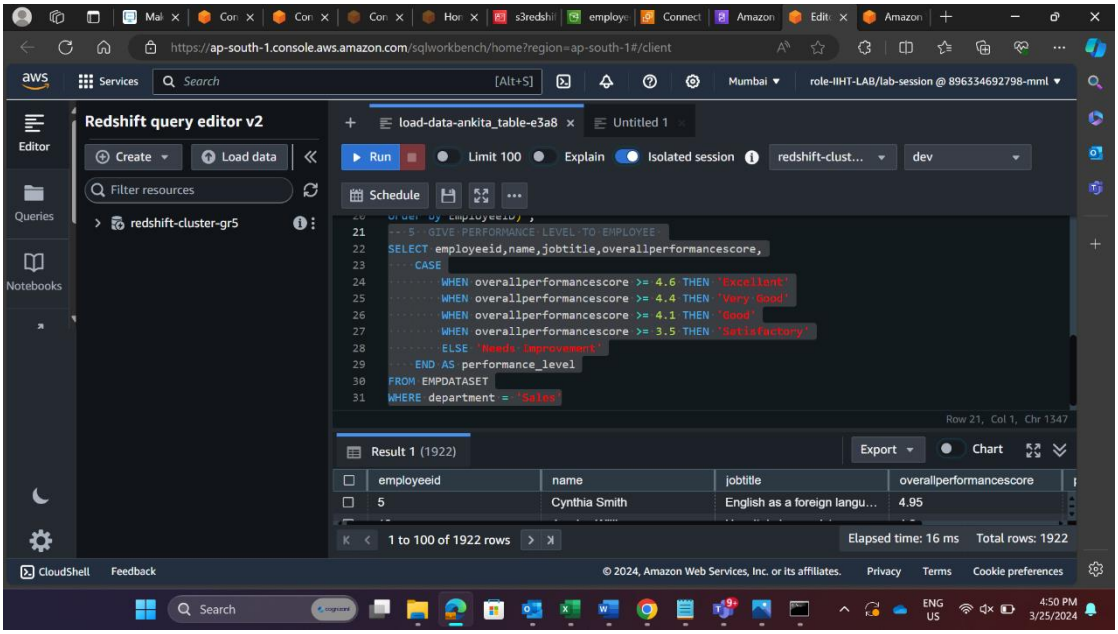


b. Visualization



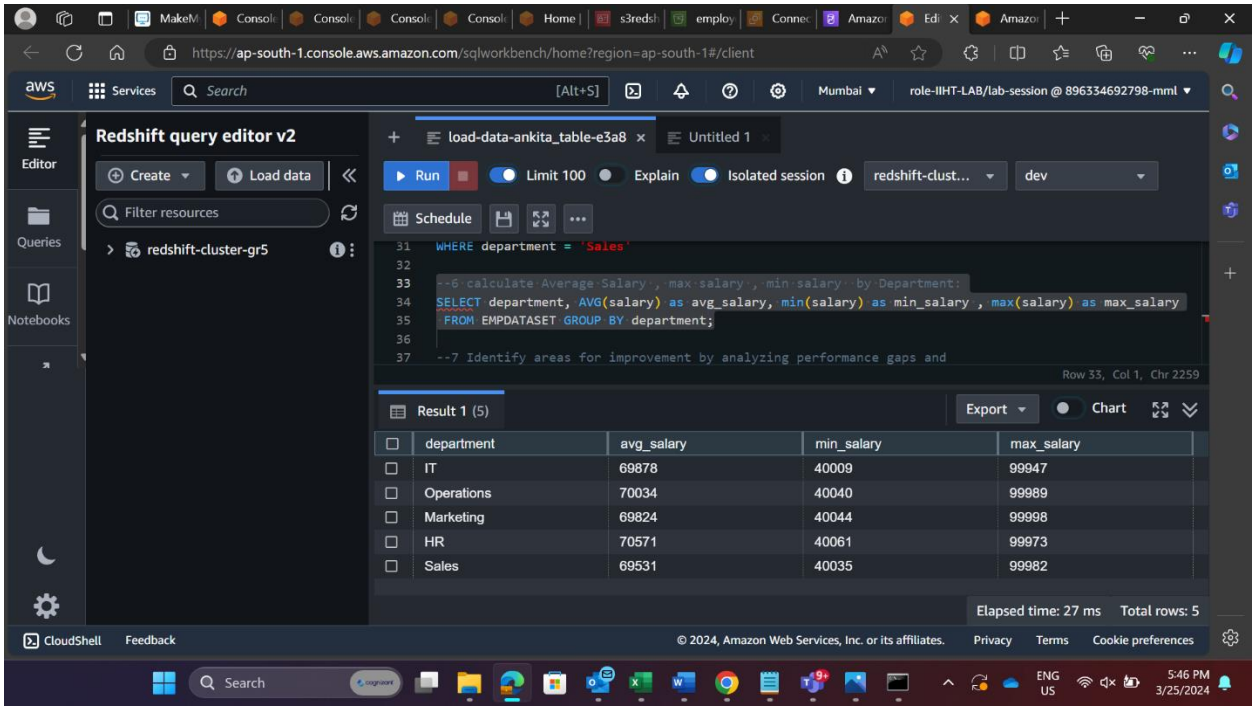
4. Give a performance value to all the employees of the sales department based on some specified criteria.

a. Query Execution

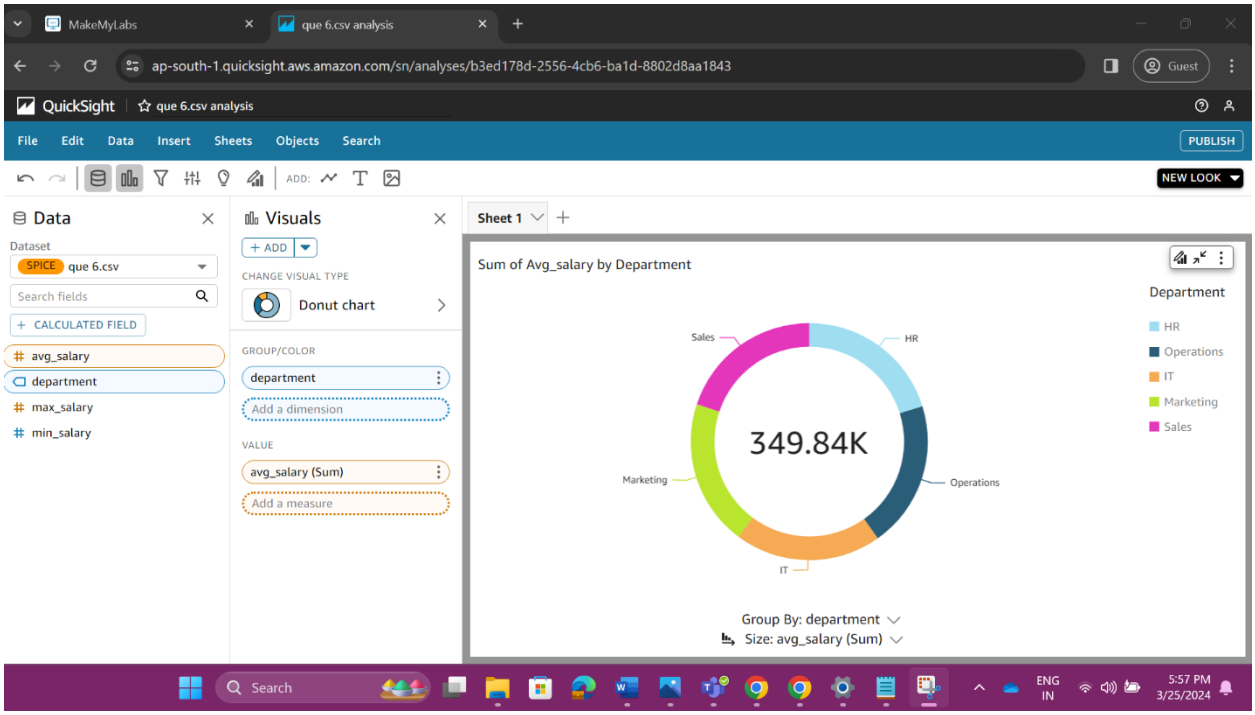


5. Calculate the average, minimum and maximum salary of each department in the employee table.

a. Query Execution

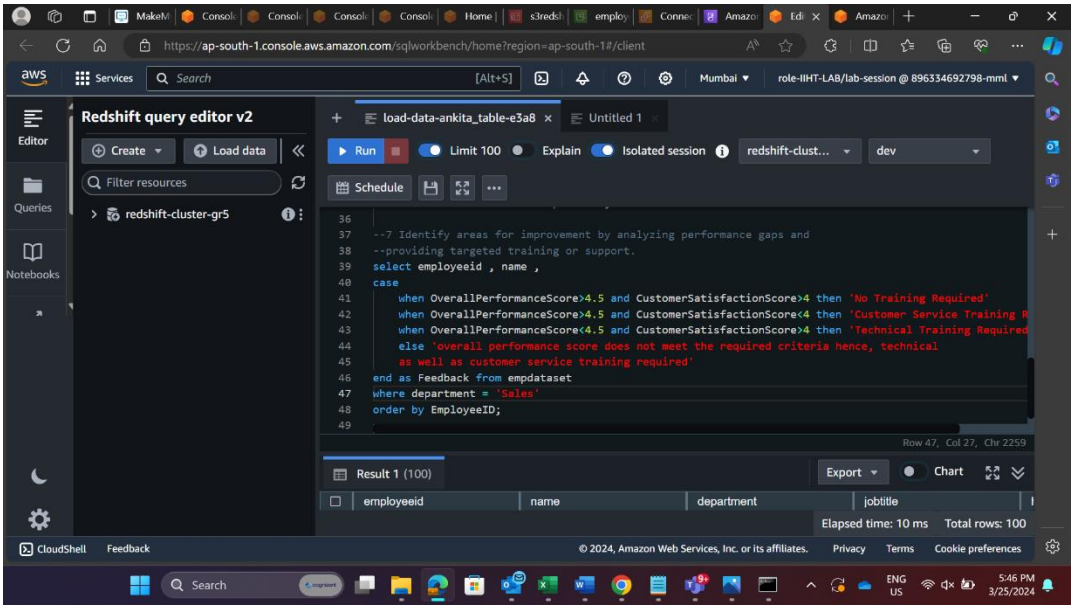


b. Visualization

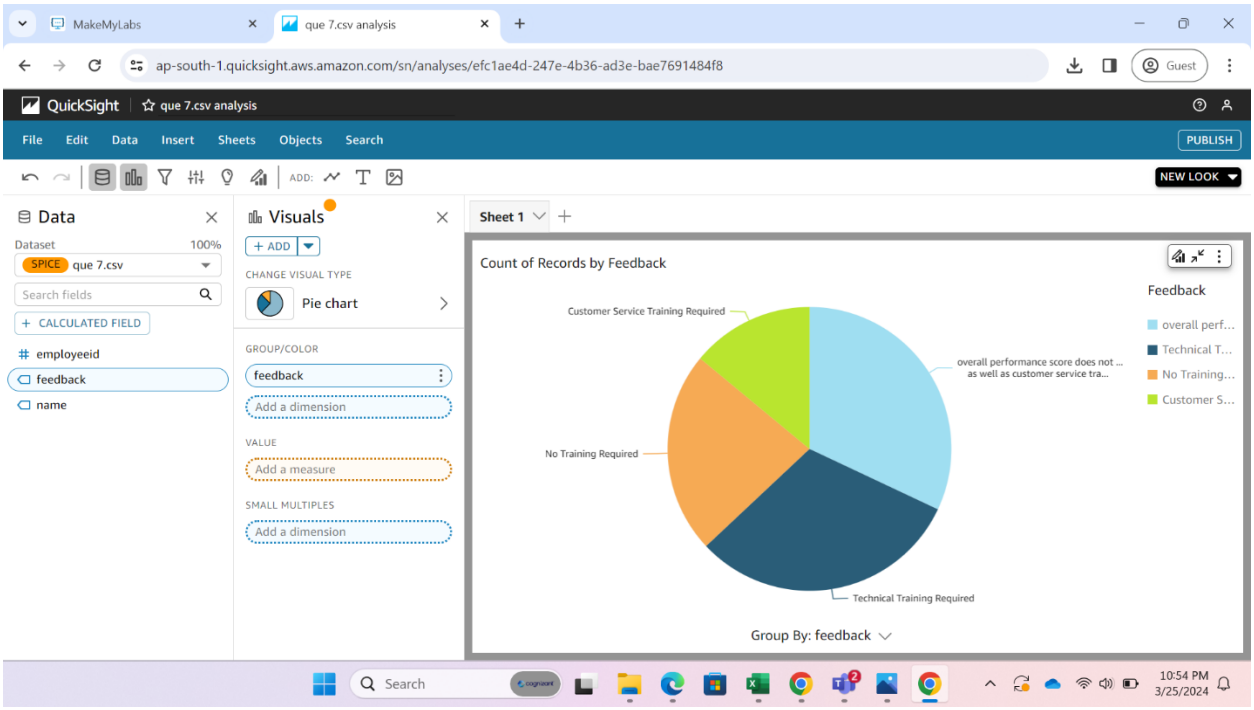


6. Identify areas for improvement by analyzing performance gaps and providing targeted training or support.

a. Query Execution



b. Visualization



CONCLUSION

In conclusion, the project exemplifies the seamless integration and powerful capabilities of AWS services in constructing a robust data infrastructure. By harnessing Amazon S3 as a scalable storage solution, we ensured the efficient and secure handling of our datasets. The coupling of S3 with Amazon Redshift allowed for the establishment of a high-performance data warehousing environment, facilitating quick access and analysis of our structured data. Moreover, the utilization of Amazon EC2 instances enabled us to execute complex SQL operations, ensuring the manipulation and transformation of data according to our analytical needs.

Furthermore, the deployment of AWS Quicksight for visualization added a layer of insight to our data exploration process. With Quicksight's intuitive interface and comprehensive visualization options, we were able to craft informative reports and dashboards that provided stakeholders with clear and actionable insights. This not only enhanced our understanding of the underlying data but also empowered decision-makers to make informed choices based on the analyzed information.

In summary, the successful execution of this project underscores the value proposition of AWS as a leading cloud provider for data analytics initiatives. The combination of S3, Redshift, EC2, and Quicksight offered a cohesive and scalable solution that streamlined the entire data processing pipeline. As organizations continue to grapple with increasing volumes of data, leveraging AWS services provides a reliable and efficient means to extract actionable insights and drive business outcomes. Moving forward, the lessons learned from this project will serve as a foundation for future endeavors in harnessing the full potential of cloud-based data analytics solutions.