



SalesAnalysis

By,

SantanuAdhikary(2320846)

E.DivyaSree(2320625)

A.N.V.Rohit(2320408)

SanmithChetanN(2320907)

ShashankK

Hulmani(2320613)

Vivechan Gowda

G.C(2320755)

Contents

1.Introduction -----	03
a) Amazon Redshift	
b) Amazon S3	
c)Quick Sight Tool	
2.Business Requirements -----	04
3.Overview of the project -----	05
4.Implementation -----	06
5.Visualisation -----	34
6.Conclusion -----	40

Introduction

Regardless of whether you are a small local business or a national corporation, tracking your retail sales is incredibly important. The valuable activity lets you identify potential sales patterns, close more deals in much less time, and make better and informed sales forecasts. However, it is seen that most retail sales managers struggle to keep a tap on their sales performance. This, in turn, impacts their retail productivity.

The amount of data gathered by retail businesses keeps growing at an alarmingly high rate. Sales managers are unable to ensure the adequate utilization and analysis of data, contributing to the issue of data silos. Besides, it results in the most common challenges for retailers: a lack of consistency in the implementation of the sales process, limited line-of-sight for managers, and reps spending too much time not selling.

It is crucial to find a technology solution that can effectively handle all the generated data and use it to unlock meaningful insights into sales performance for the retail industry. This is where **AWS Redshift** comes into the picture. In today's blog, we will understand how using Amazon Redshift can help to significantly **improve sales performance** and tracking.

A). Amazon Redshift:

Amazon Redshift is a fully managed data warehouse service provided by Amazon Web Services (AWS). It's designed for analyzing large datasets and offers scalability, high performance, and SQL compatibility. Redshift uses a distributed architecture with massively parallel processing (MPP), making it efficient for complex queries. Data is stored in a columnar format, which enhances query speed. Whether you're running business intelligence reports or analyzing historical data, Amazon Redshift is a powerful choice for data analytics and reporting.

B). Amazon S3 Storage:

Amazon S3 (Simple Storage Service) is an object storage service provided by Amazon Web Services (AWS). It offers industry-leading scalability, data availability, security, and performance. Here are the key points about Amazon S3:

- Scalability: Amazon S3 allows you to store and retrieve any amount of data from anywhere on the web. It's designed to handle both small and large-scale applications.

- Object Storage: Instead of traditional file storage, Amazon S3 stores data as objects within resources called buckets. These objects can be photos, audio files, videos, or any other type of data.
- Durability and Availability: Amazon S3 provides 99.999999999% durability (meaning your data is highly resistant to loss) and 99.99% availability of objects over a given year.

C). Quick Sight:

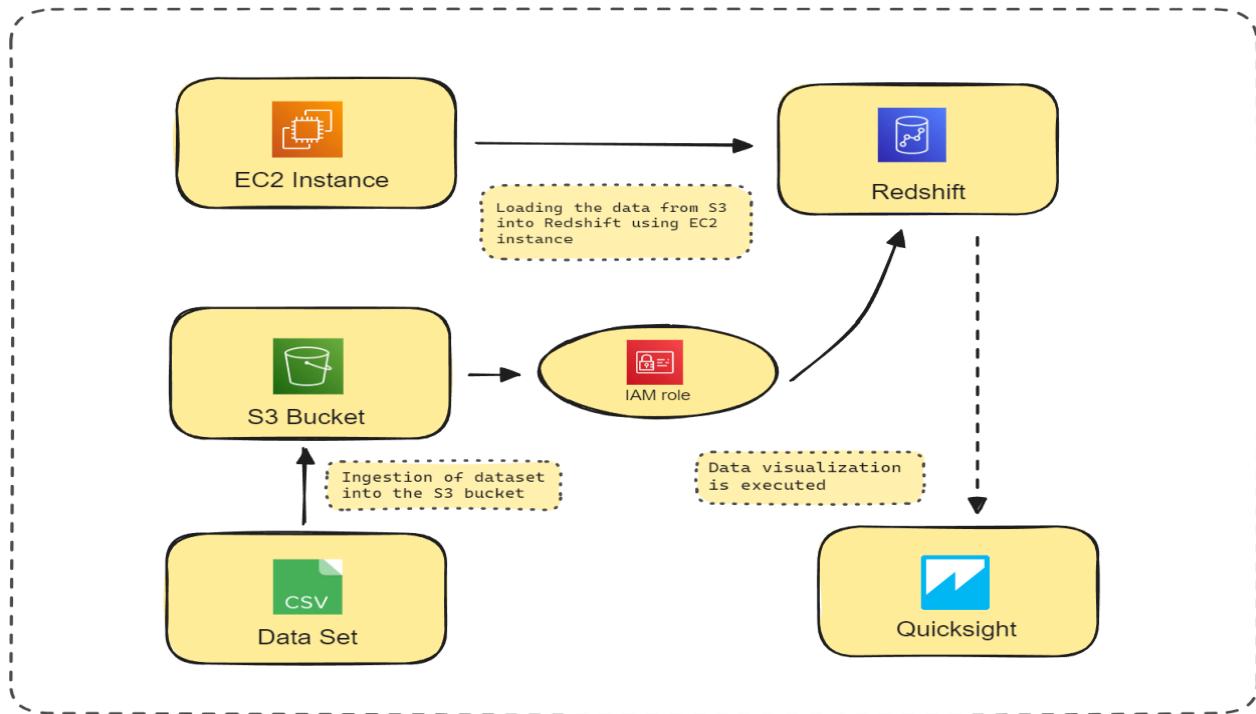
Amazon Quick Sight is a cloud-native, serverless business intelligence (BI) service provided by Amazon Web Services (AWS). Here are the key points about Quick Sight:

- Unified BI at Hyperscale: Quick Sight powers data-driven organizations with unified BI capabilities. It allows users to meet varying analytic needs from the same source of truth through:
- **Modern Interactive Dashboards:** Create interactive visualizations and explore data.
- **Paginated Reports:** Generate critical operational reports and dashboards.
 - **Embedded Analytics:** Embed dashboards or natural language query capabilities in applications.

Business Requirements

1. Import sales data into Amazon Redshift.
2. Create SQL queries to aggregate sales data by product, region, and time period.
3. Identify top-selling products by calculating total sales or units sold.
4. Determine highest revenue-generating regions by analyzing sales revenue per region.
5. Analyze trends over time by plotting sales data over different time periods (e.g., daily, monthly, quarterly).
6. Visualize the sales data using charts and graphs (e.g., bar charts, line graphs) to provide insights into sales performance.

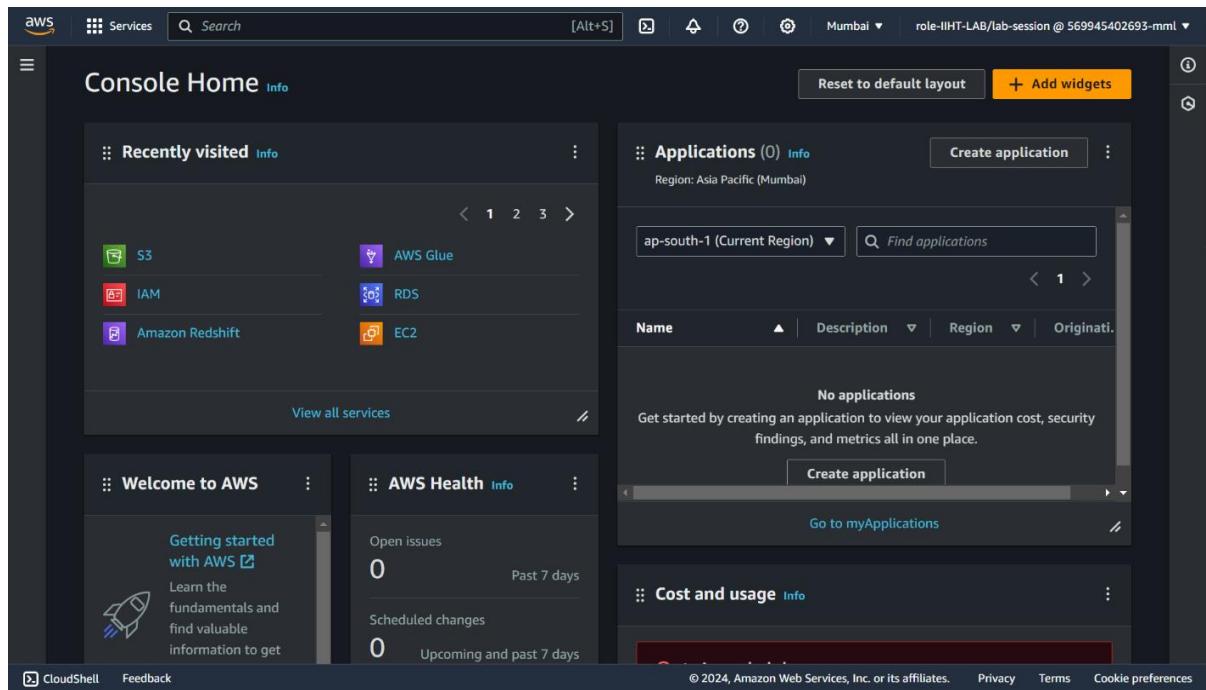
Overview of the Project



1	2	3	4	5	6
Fetching data from kaggle. And Loading the data to S3 bucket.	Creating IAM role for S3 Access from Redshift	Deploying a Redshift Cluster	Deploy EC2 instance ,connect to Redshift, Load the data from S3 to Redshift .	performing data analysis using Redshift Query Editor	Loading the data to Quick sight and identifying trends and patterns through visualization.

Implementation of Project

Step 1: Open AWS console

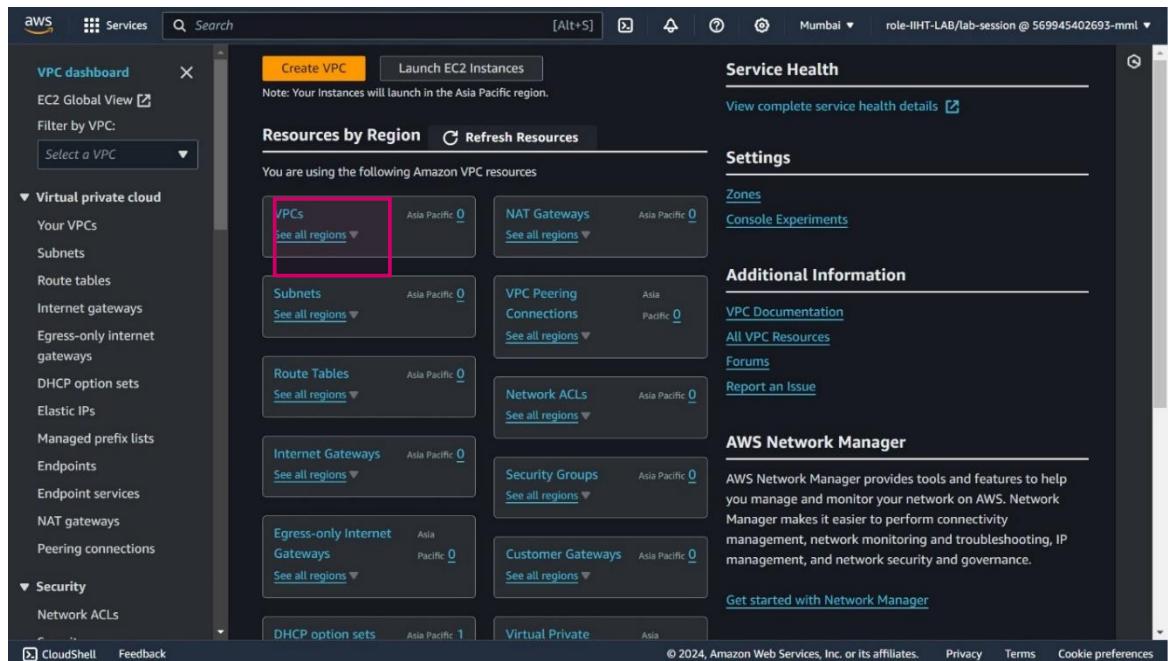


The screenshot shows the AWS Console Home page. The top navigation bar includes the AWS logo, Services, a search bar, and account information (Mumbai, role-IIHT-LAB/lab-session). Below the navigation is the 'Console Home' header with a 'Reset to default layout' and 'Add widgets' button. The main content area is divided into several sections:

- Recently visited:** Shows links to S3, AWS Glue, IAM, RDS, Amazon Redshift, and EC2. A 'View all services' link is also present.
- Welcome to AWS:** Includes a 'Getting started with AWS' section with a rocket icon and a 'Learn the fundamentals and find valuable information to get' link.
- AWS Health:** Shows 'Open issues: 0' and 'Scheduled changes: 0'.
- Applications:** Shows 'No applications' with a note to 'Get started by creating an application to view your application cost, security findings, and metrics all in one place.' It includes a 'Create application' button and a 'Go to myApplications' link.
- Cost and usage:** Shows a large red bar indicating high costs.

At the bottom, there are links for CloudShell, Feedback, and a copyright notice: © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences.

Step 2: Create VPC

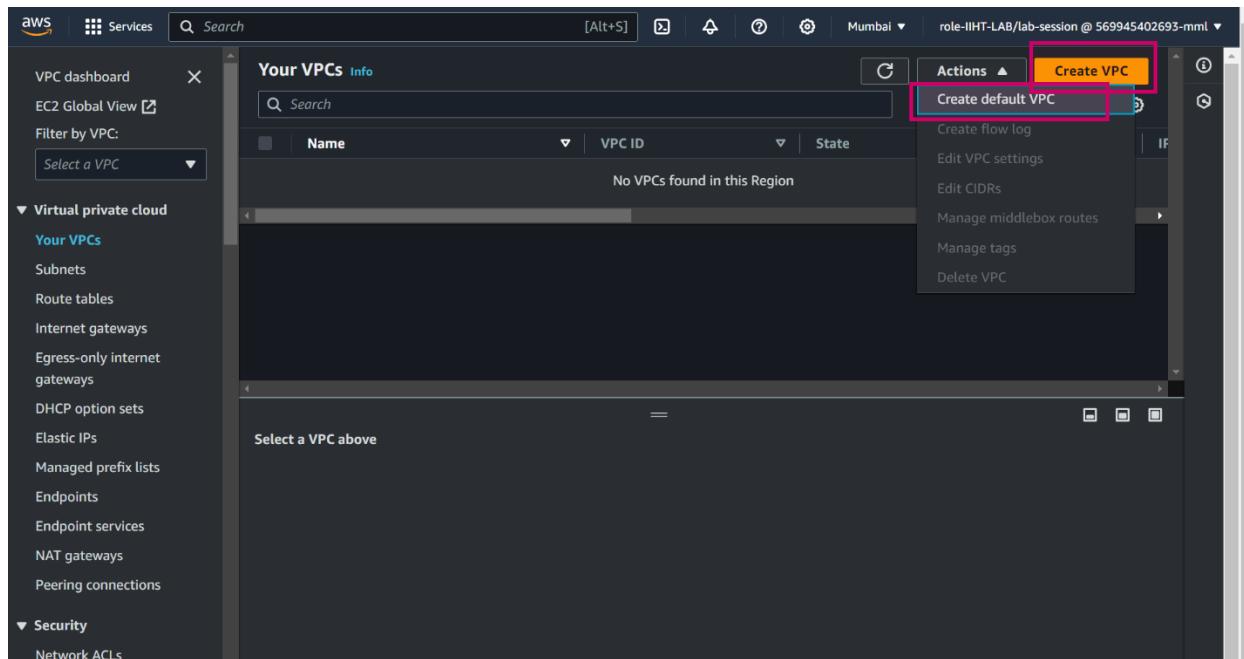


The screenshot shows the VPC dashboard in the AWS Console. The top navigation bar includes the AWS logo, Services, a search bar, and account information (Mumbai, role-IIHT-LAB/lab-session). Below the navigation is the 'VPC dashboard' header with a 'Create VPC' and 'Launch EC2 Instances' button. The main content area is divided into several sections:

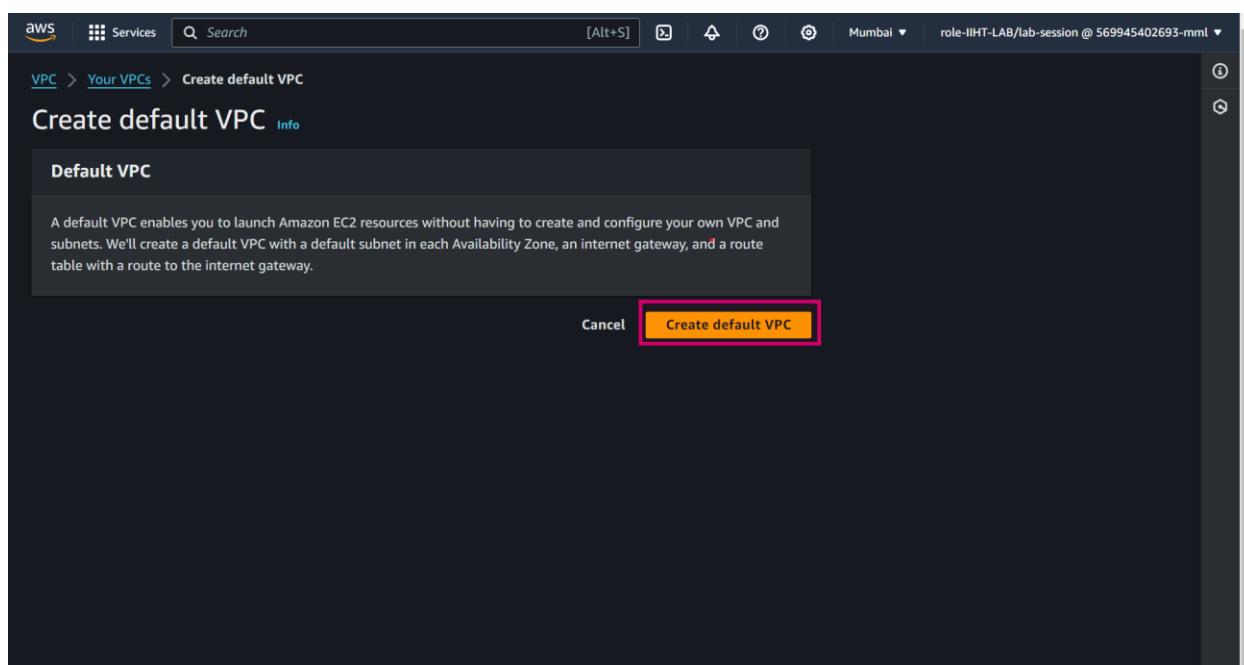
- VPC dashboard:** Includes a 'EC2 Global View' and a 'Filter by VPC' dropdown set to 'Select a VPC'.
- Virtual private cloud:** Lists Your VPCs, Subnets, Route tables, Internet gateways, Egress-only Internet gateways, DHCP option sets, Elastic IPs, Managed prefix lists, Endpoints, Endpoint services, NAT gateways, and Peering connections.
- Resources by Region:** Shows various Amazon VPC resources:
 - VPCs:** Asia Pacific 0 (highlighted with a pink box)
 - Subnets:** Asia Pacific 0
 - Route Tables:** Asia Pacific 0
 - Internet Gateways:** Asia Pacific 0
 - Egress-only Internet Gateways:** Asia Pacific 0
 - DHCP option sets:** Asia Pacific 1
 - NAT Gateways:** Asia Pacific 0
 - VPC Peering Connections:** Asia Pacific 0
 - Network ACLs:** Asia Pacific 0
 - Security Groups:** Asia Pacific 0
 - Customer Gateways:** Asia Pacific 0
- Service Health:** Includes a 'View complete service health details' link.
- Settings:** Includes 'Zones' and 'Console Experiments'.
- Additional Information:** Includes 'VPC Documentation', 'All VPC Resources', 'Forums', and 'Report an Issue'.
- AWS Network Manager:** Describes the tool for managing and monitoring network on AWS. It includes a 'Get started with Network Manager' link.

At the bottom, there are links for CloudShell, Feedback, and a copyright notice: © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences.

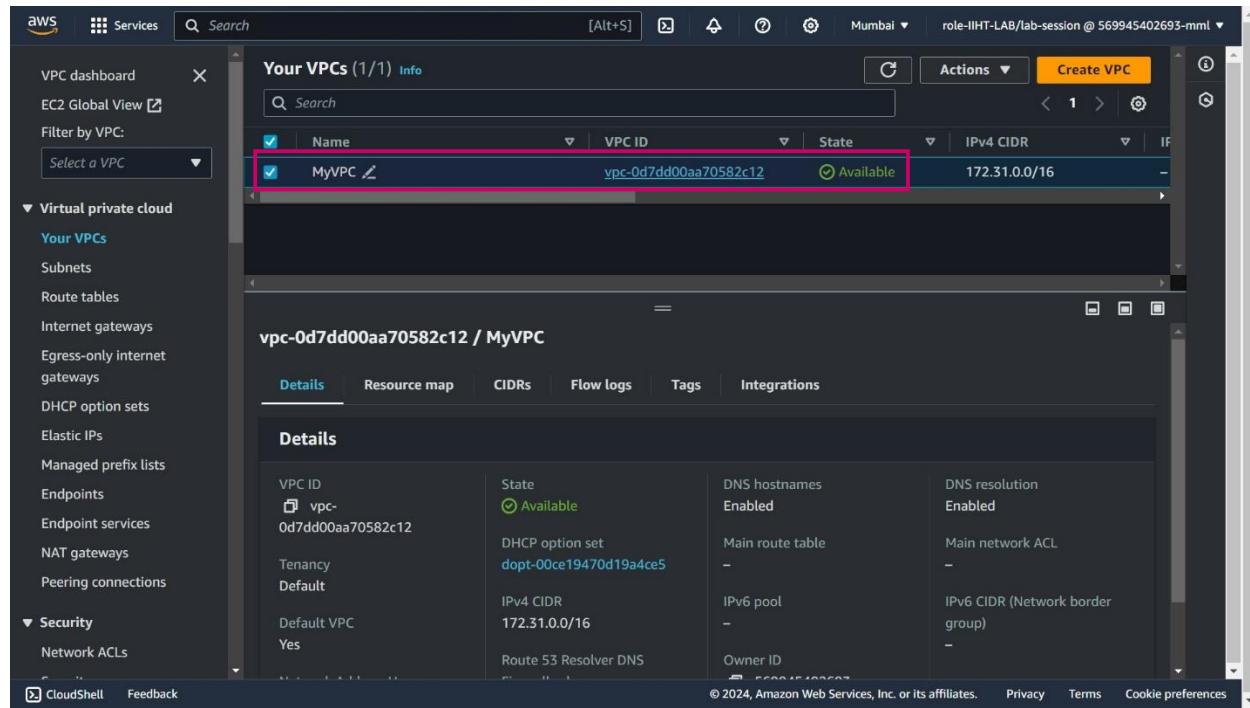
Open VPC and create default VPC and add subnets



The screenshot shows the AWS VPC dashboard. On the left, there's a sidebar with options like 'Your VPCs', 'Subnets', 'Route tables', etc. The main area is titled 'Your VPCs' and shows a message 'No VPCs found in this Region'. At the top right, there's an 'Actions' menu with several options: 'Create VPC' (highlighted with a pink box), 'Create default VPC', 'Create flow log', 'Edit VPC settings', 'Edit CIDRs', 'Manage middlebox routes', 'Manage tags', and 'Delete VPC'.



The screenshot shows the 'Create default VPC' wizard. The title bar says 'VPC > Your VPCs > Create default VPC'. The main section is titled 'Default VPC' and contains the following text: 'A default VPC enables you to launch Amazon EC2 resources without having to create and configure your own VPC and subnets. We'll create a default VPC with a default subnet in each Availability Zone, an internet gateway, and a route table with a route to the internet gateway.' At the bottom, there are 'Cancel' and 'Create default VPC' buttons, with 'Create default VPC' also highlighted with a pink box.



The screenshot shows the AWS VPC dashboard. On the left sidebar, under 'Virtual private cloud', 'Your VPCs' is selected. A table titled 'Your VPCs (1/1)' lists one entry: 'MyVPC' (vpc-0d7dd00aa70582c12). The row for 'MyVPC' is highlighted with a pink border. The 'Actions' button is visible at the top right of the table. Below the table, the details for 'vpc-0d7dd00aa70582c12 / MyVPC' are displayed in a card, with the 'Details' tab selected. The card contains the following information:

VPC ID	State	DNS hostnames	DNS resolution
vpc-0d7dd00aa70582c12	Available	Enabled	Enabled
Tenancy	DHCP option set	Main route table	Main network ACL
Default	dopt-00ce19470d19a4ce5	-	-
Default VPC	IPv4 CIDR	IPv6 pool	IPv6 CIDR (Network border group)
Yes	172.31.0.0/16	-	-
	Route 53 Resolver DNS	Owner ID	
	route53resolver.12345678901234567890	12345678901234567890	

Create subnet

VPC ID
Create subnets in this VPC
vpc-003a627b3c7ad7116 (MyVPC)

Associated VPC CIDRs
IPv4 CIDRs
10.0.0.0/16

Subnet settings
Specify the CIDR blocks and Availability Zone for the subnet.

Subnet 1 of 3

Subnet name
Create a tag with a key of 'Name' and a value that you specify.
Public1A
The name can be up to 256 characters long.

Availability Zone
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
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.
10.0.0.0/16

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

Subnet settings
Specify the CIDR blocks and Availability Zone for the subnet.

Subnet 1 of 3

Subnet name
Create a tag with a key of 'Name' and a value that you specify.
Public1A
The name can be up to 256 characters long.

Availability Zone
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
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.
10.0.0.0/16

IPv4 subnet CIDR block
10.0.1.0/24 256 IPs

Tags - optional
Key Value - optional
Name Public1A Remove Add new tag You can add 49 more tags. Remove

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

Subnet 2 of 3

Subnet name
Create a long worth key or 'Name' and a value that you specify.
Public1B

The name can be up to 256 characters long.

Availability Zone [Info](#)
Asia Pacific (Mumbai) / ap-south-1b

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.
10.0.0.0/16

IPv4 subnet CIDR block
10.0.2.0/24 256 IPs

▼ Tags - optional

Key	Value - optional
<input type="text" value="Name"/>	<input type="text" value="Public1B"/> X Remove

Add new tag
You can add 49 more tags.
[Remove](#)

Subnet 3 of 3

Subnet name
Public1C

Availability Zone [Info](#)
Choose the zone which you want will be used, or let Amazon choose one for you.
Asia Pacific (Mumbai) / ap-south-1c

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.
10.0.0.0/16

IPv4 subnet CIDR block
10.0.3.0/24 256 IPs

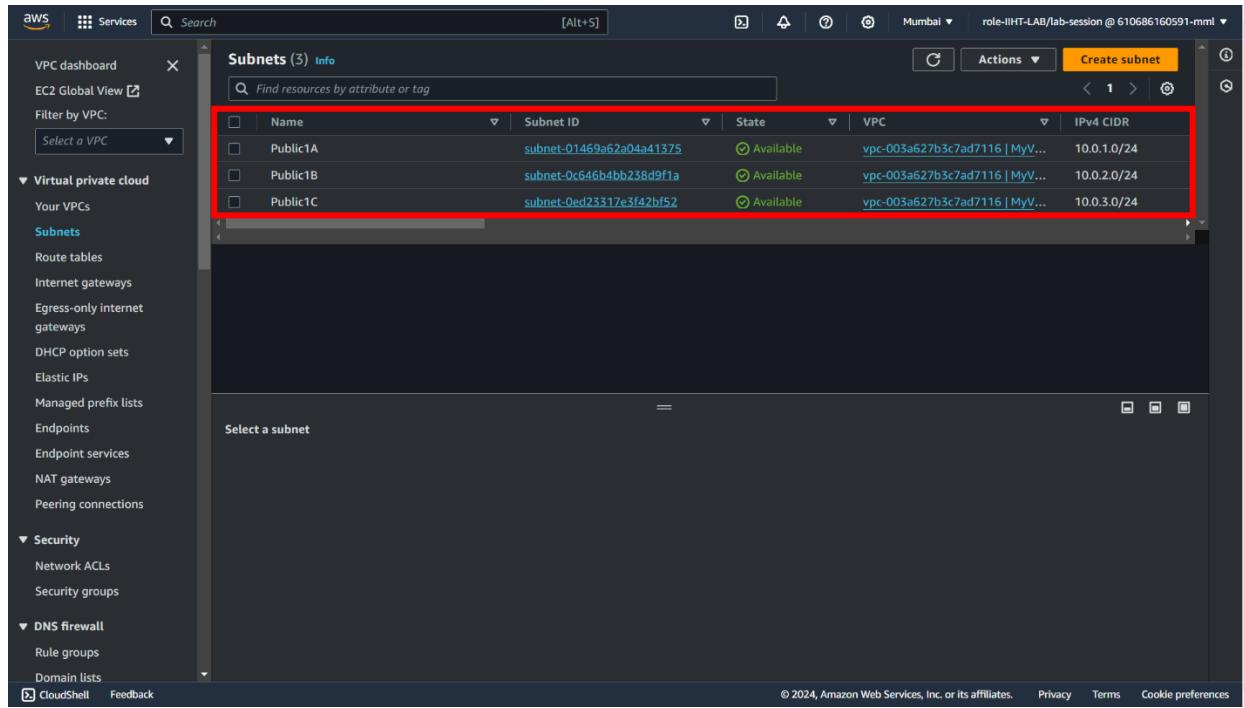
▼ Tags - optional

Key	Value - optional
<input type="text" value="Name"/>	<input type="text" value="Public1C"/> X Remove

Add new tag
You can add 49 more tags.
[Remove](#)

[Add new subnet](#)

[Cancel](#) [Create subnet](#)

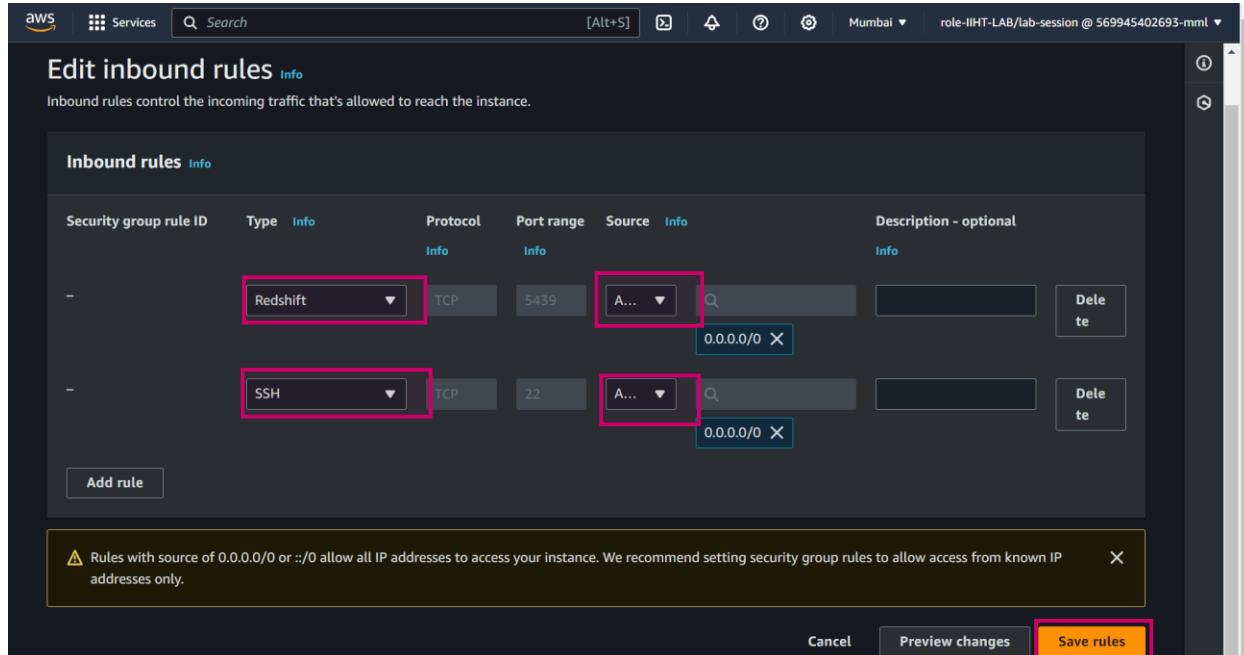


Subnets (3) Info

Name	Subnet ID	State	VPC	IPv4 CIDR
Public1A	subnet-01469a62a04a41375	Available	vpc-003a627b5c7ad7116 MyV...	10.0.1.0/24
Public1B	subnet-0c646b4bb238d9f1a	Available	vpc-003a627b5c7ad7116 MyV...	10.0.2.0/24
Public1C	subnet-0ed23317e3f42bf52	Available	vpc-003a627b5c7ad7116 MyV...	10.0.3.0/24

Step3: Creating Security Groups

There is a default security group we need to change the inbound rules to give access to redshift from S3



Edit inbound rules

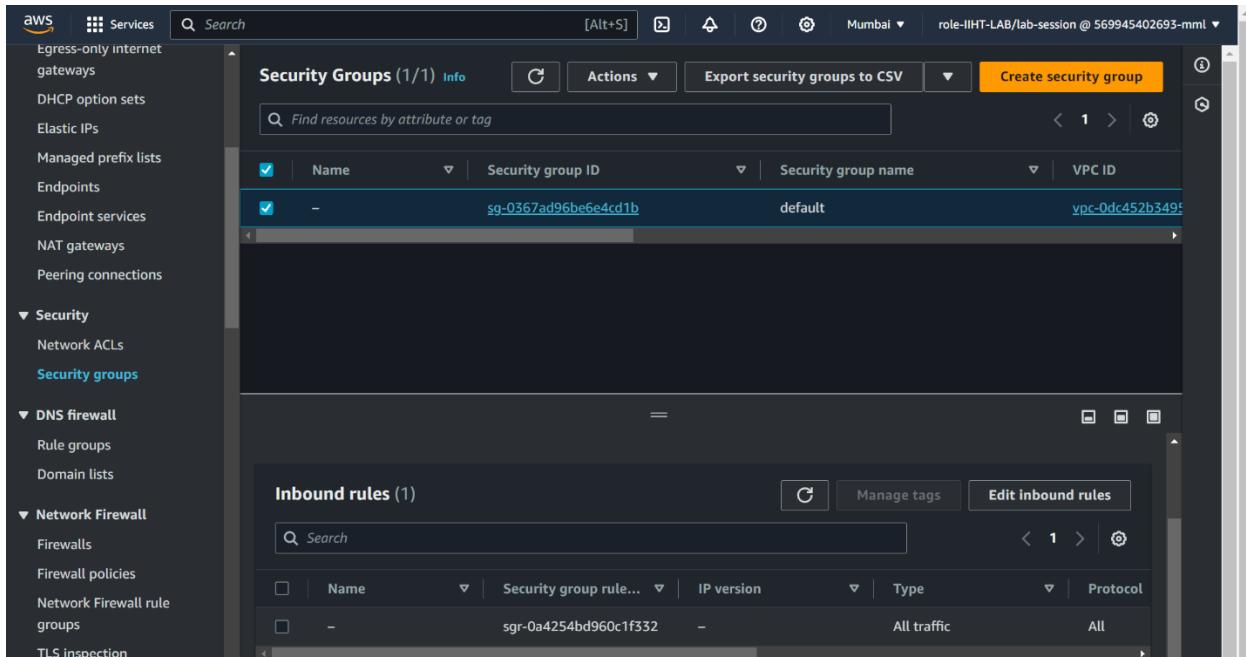
Inbound rules control the incoming traffic that's allowed to reach the instance.

Inbound rules

Security group rule ID	Type	Protocol	Port range	Source	Description - optional
-	Redshift	TCP	5439	A...	0.0.0.0/0
-	SSH	TCP	22	A...	0.0.0.0/0

Warning: ⚠ Rules with source of 0.0.0.0/0 or ::/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

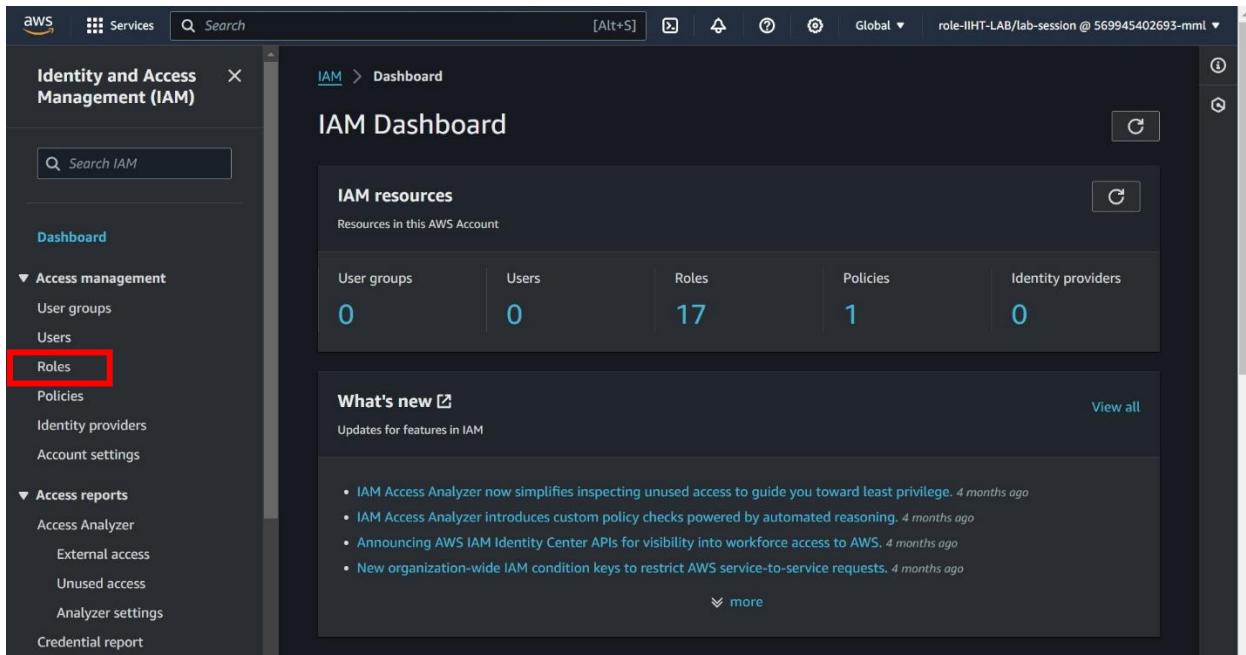
Buttons: Cancel, Preview changes, Save rules



The screenshot shows the AWS CloudFormation console. A single stack named "Redshift-Stack" is listed, containing one resource: "Redshift Cluster". The status of the stack is "CREATE_COMPLETE". The cluster has a name of "redshift-cluster-1" and is associated with a VPC ID of "vpc-0dc452b3495". The creation time was "2023-09-12T10:25:00Z".

Step 4 : Open IAM Dashboard to create role for Redshift to access the S3

It is used to securely access the AWS services and resources. It helps to enhance the security by ensuring that only authorized individuals or system can interact with our AWS.

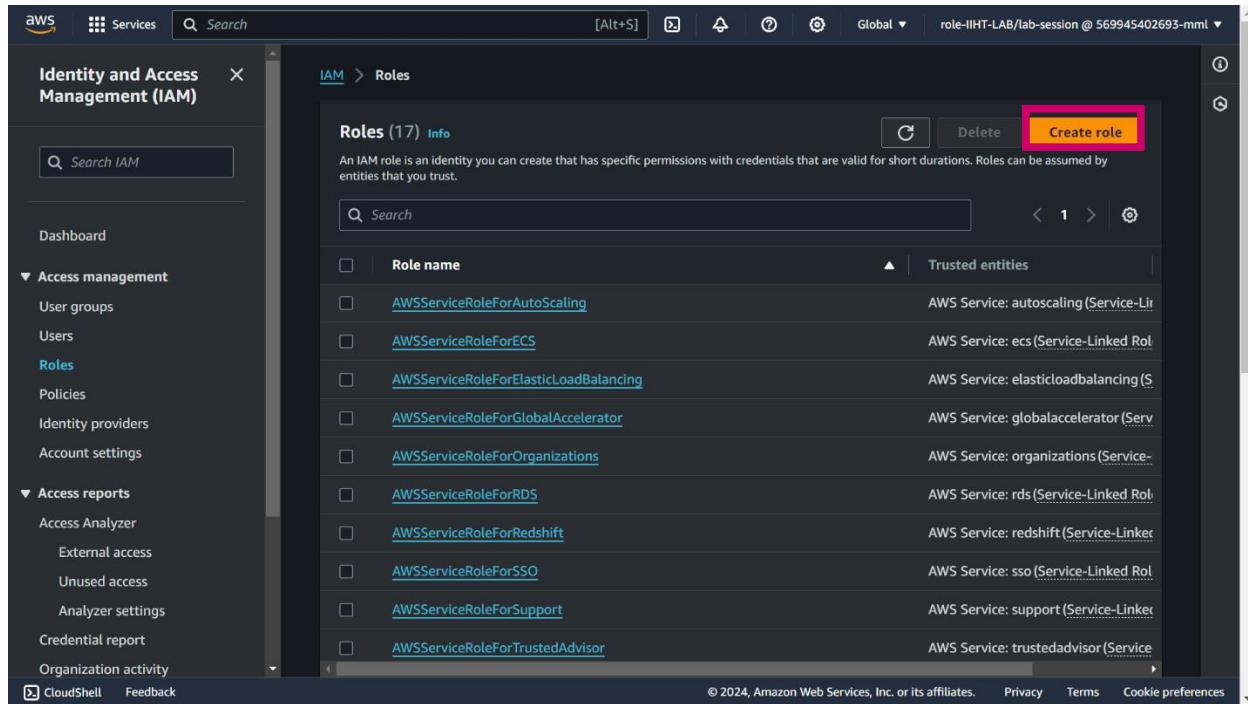


The screenshot shows the AWS IAM Dashboard. The left sidebar is expanded to show the "Access management" section, with the "Roles" option highlighted by a red box. The main dashboard displays the following statistics:

User groups	Users	Roles	Policies	Identity providers
0	0	17	1	0

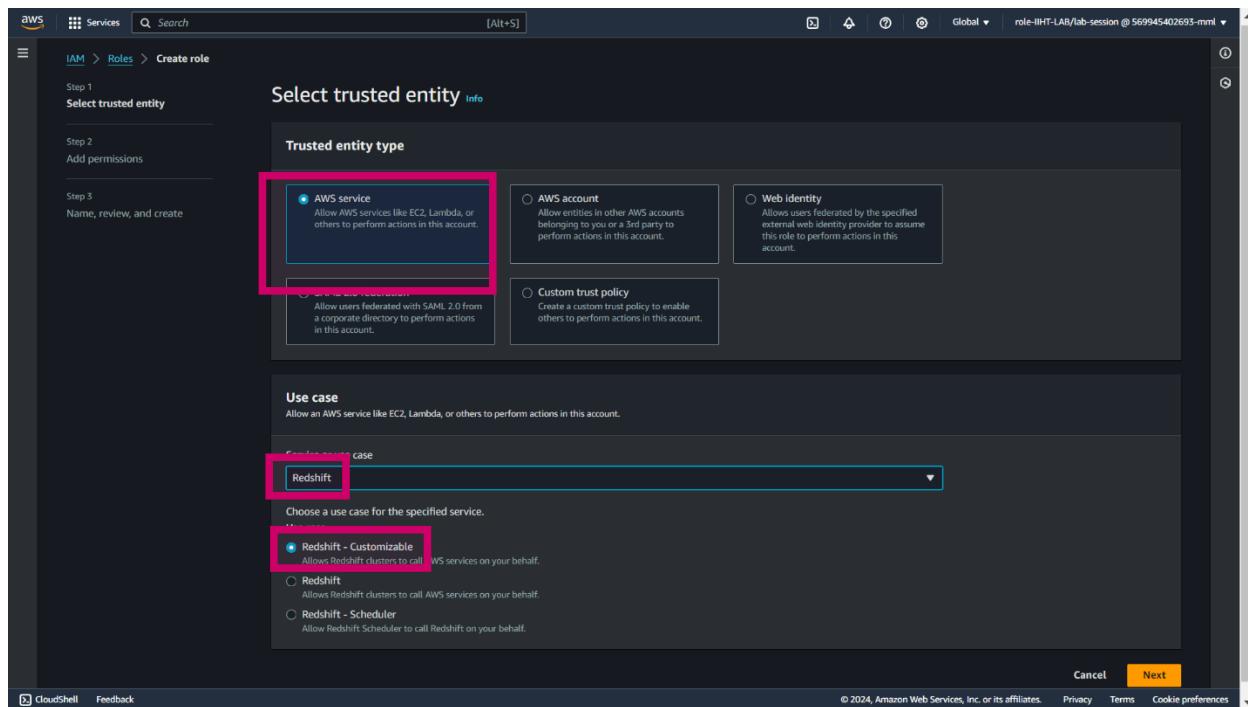
Below the statistics, there is a "What's new" section listing recent updates:

- IAM Access Analyzer now simplifies inspecting unused access to guide you toward least privilege. 4 months ago
- IAM Access Analyzer introduces custom policy checks powered by automated reasoning. 4 months ago
- Announcing AWS IAM Identity Center APIs for visibility into workforce access to AWS. 4 months ago
- New organization-wide IAM condition keys to restrict AWS service-to-service requests. 4 months ago



The screenshot shows the AWS IAM Roles page. The left sidebar is collapsed. The main area displays a table of 17 IAM roles, each with a checkbox, a role name, and a description of the trusted entity. The 'Create role' button at the top right is highlighted with a red box.

	Role name	Trusted entities
<input type="checkbox"/>	AWSServiceRoleForAutoScaling	AWS Service: autoscaling (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForECS	AWS Service: ecs (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForElasticLoadBalancing	AWS Service: elasticloadbalancing (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForGlobalAccelerator	AWS Service: globalaccelerator (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForOrganizations	AWS Service: organizations (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForRDS	AWS Service: rds (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)
<input type="checkbox"/>	AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service-Linked Role)



The screenshot shows the 'Create role' wizard, Step 1: Select trusted entity. It has three tabs: Step 1 (Select trusted entity), Step 2 (Add permissions), and Step 3 (Name, review, and create). The 'Select trusted entity' tab is active. The 'Trusted entity type' section shows four options: 'AWS service' (selected and highlighted with a red box), 'AWS account', 'Web Identity', and 'Custom trust policy'. Below this, the 'Use case' section shows a dropdown menu with 'Redshift' selected (highlighted with a red box) and a list of use cases: 'Redshift - Customizable' (selected and highlighted with a red box), 'Redshift', and 'Redshift - Scheduler'.

Add permissions

Permissions policies (1/913) [Info](#)

Choose one or more policies to attach to your new role.

Policy name	Type	Description
<input type="checkbox"/> AmazonDMSRedshiftS3Role	AWS managed	Provides access to manage S3 settings for...
<input checked="" type="checkbox"/> AmazonS3FullAccess	AWS managed	Provides full access to all buckets via the ...
<input type="checkbox"/> AmazonS3ObjectLambdaExecutionRolePolicy	AWS managed	Provides AWS Lambda functions permis...
<input type="checkbox"/> AmazonS3OutpostsFullAccess	AWS managed	Provides full access to Amazon S3 on Out...
<input type="checkbox"/> AmazonS3ReadOnlyAccess	AWS managed	Provides read only access to Amazon S3 ...
<input type="checkbox"/> AWSBackupServiceRolePolicyForS3Backup	AWS managed	Policy containing permissions necessary f...
<input type="checkbox"/> AWSBackupServiceRolePolicyForS3Restore	AWS managed	Policy containing permissions necessary f...
<input type="checkbox"/> QuickSightAccessForS3StorageManagementAnalytic...	AWS managed	Policy used by QuickSight team to access ...

Set permissions boundary - *optional*

[Cancel](#) [Previous](#) **Next**

Name, review, and create

Role details

Role name **Success**

Add a short explanation for this role **Allows Redshift clusters to call AWS services on your behalf.**

Step 1: Select trusted entities

Trust policy

```

1: {
2:   "Version": "2012-10-17",
3:   "Statement": [
4:     {
5:       "Effect": "Allow",
6:       "Action": "sts:AssumeRole",
7:       "Principal": "*"
8:     }
9:   ],
10:  "Service": "redshift.amazonaws.com"
11: }
12: }
13: }
14: }
15: }
16: }

```

Step 2: Add permissions

Permissions policy summary

Policy name	Type	Attached as
AmazonS3FullAccess	AWS managed	Permissions policy

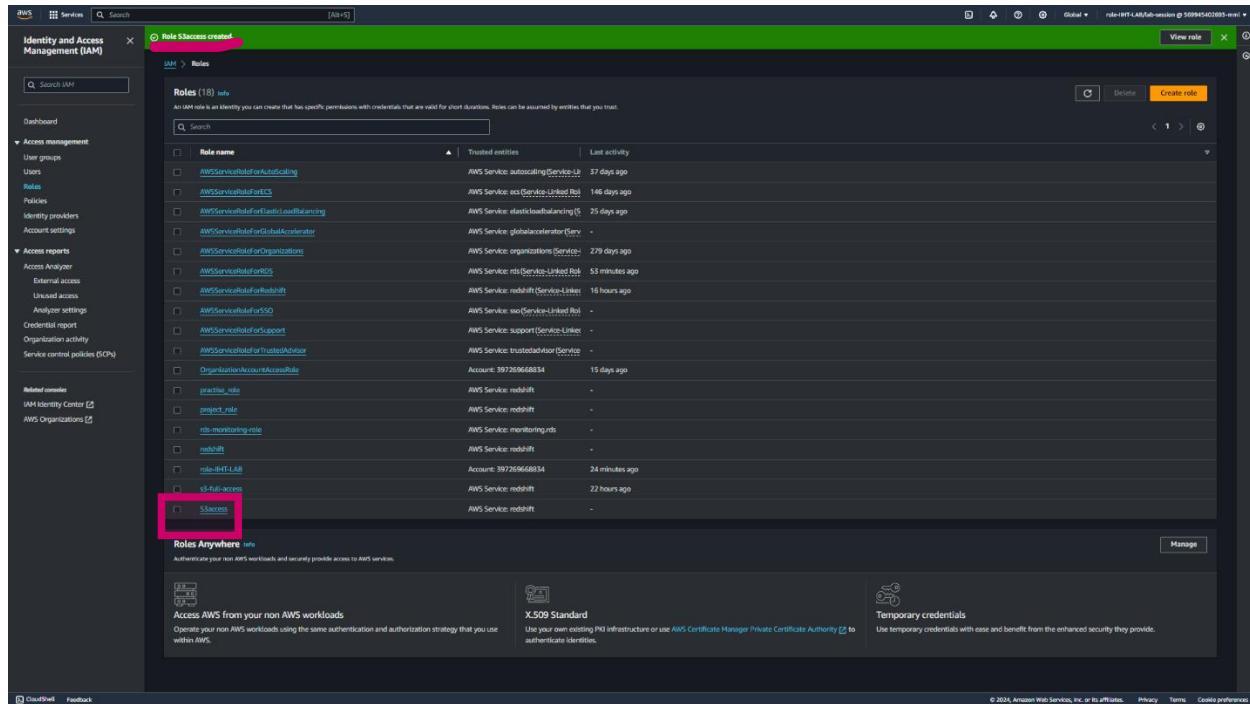
Step 3: Add tags

Add tags - *optional*

No tags associated with the resource.

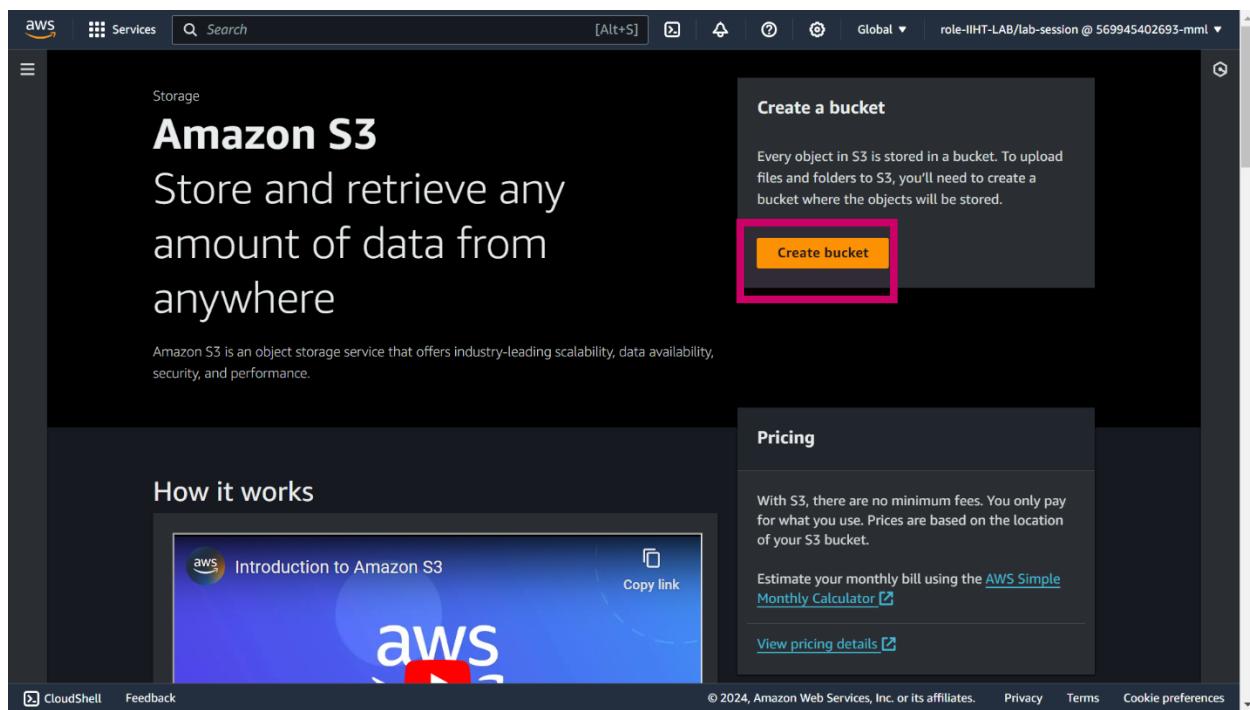
[Add new tag](#)

[Create role](#)



The screenshot shows the AWS Identity and Access Management (IAM) service. In the top left, there's a search bar and a 'View rule' button. On the left sidebar, under 'Access management', 'Roles' is selected. The main area displays a table of roles with columns for 'Role name', 'Trusted entities', and 'Last activity'. A new role, 'S3Access', is listed at the bottom of the table, highlighted with a red box. The 'Create role' button is visible in the top right corner.

Step 5: Amazon S3 Bucket creation



The screenshot shows the AWS S3 service. At the top, there's a search bar and a 'Create bucket' button. Below it, there's a large 'Amazon S3' heading with the subtext 'Store and retrieve any amount of data from anywhere'. A paragraph explains that S3 is an object storage service. On the right, there's a 'Create a bucket' section with a sub-section titled 'Pricing'.

Create a bucket

Every object in S3 is stored in a bucket. To upload files and folders to S3, you'll need to create a bucket where the objects will be stored.

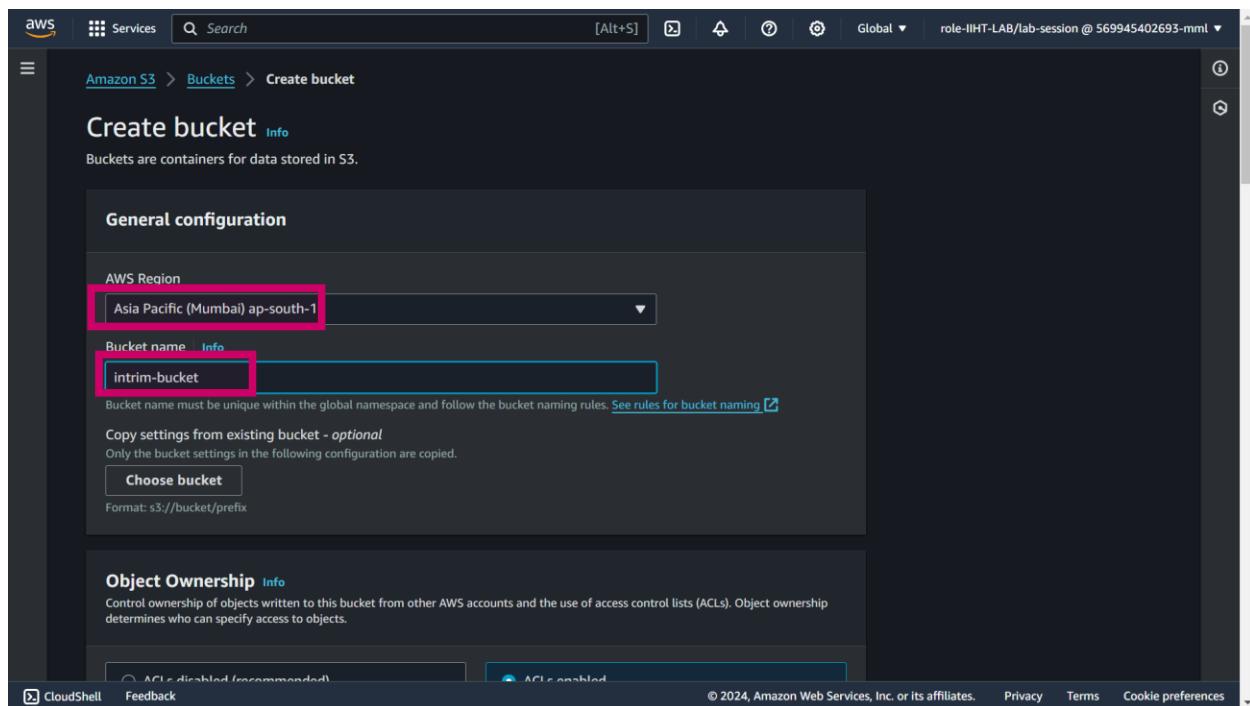
Create bucket

Pricing

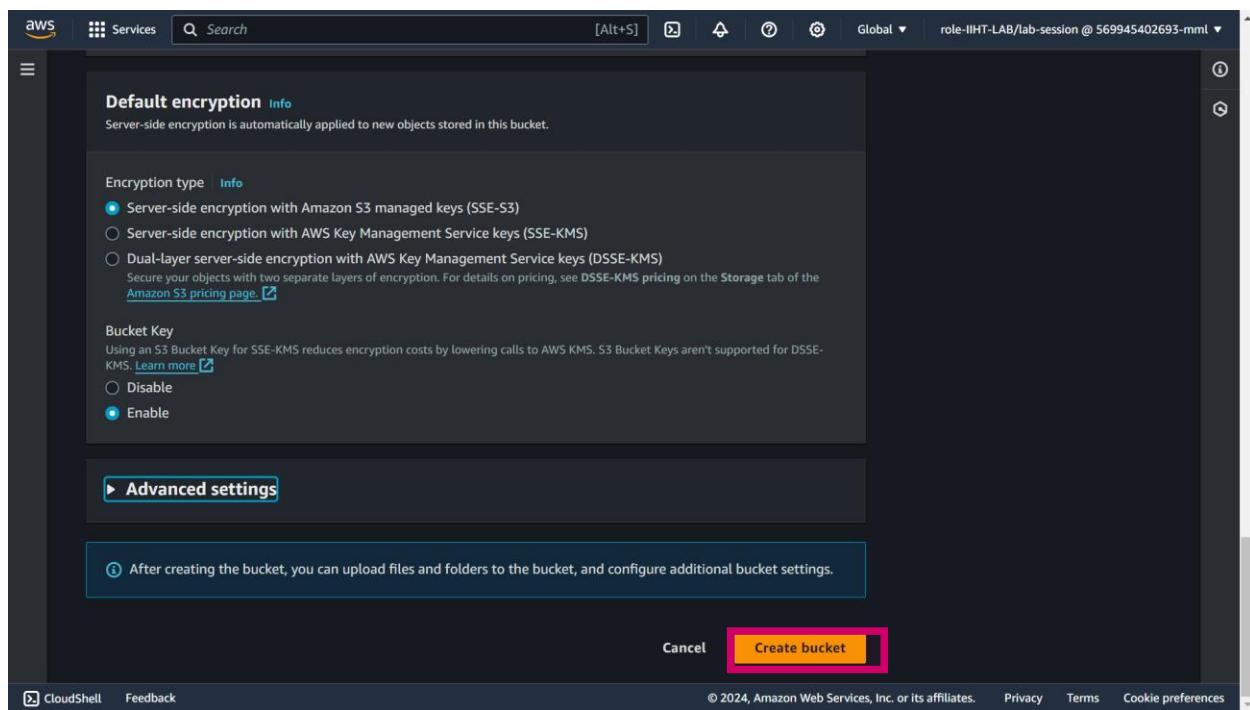
With S3, there are no minimum fees. You only pay for what you use. Prices are based on the location of your S3 bucket.

Estimate your monthly bill using the [AWS Simple Monthly Calculator](#)

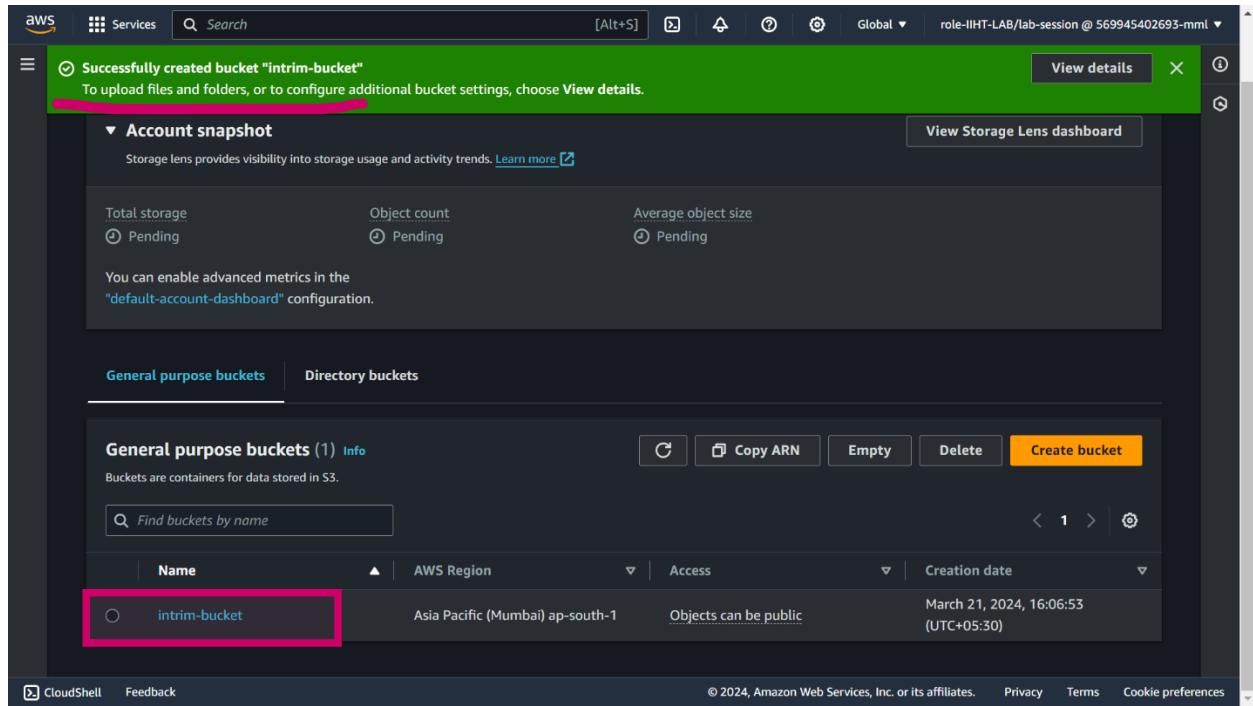
[View pricing details](#)



The screenshot shows the 'Create bucket' wizard in the AWS S3 console. The 'General configuration' step is active. A pink box highlights the 'AWS Region' dropdown set to 'Asia Pacific (Mumbai) ap-south-1'. A blue box highlights the 'Bucket name' input field containing 'intrim-bucket'. Below the input field is a note: 'Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming.' A 'Copy settings from existing bucket - optional' section is present, with a 'Choose bucket' button and a note: 'Only the bucket settings in the following configuration are copied.' A 'Format: s3://bucket/prefix' placeholder is shown. The 'Object Ownership' step is partially visible below.



The screenshot shows the 'Default encryption' step of the 'Create bucket' wizard. A pink box highlights the 'Encryption type' section, which includes three options: 'Server-side encryption with Amazon S3 managed keys (SSE-S3)' (selected), 'Server-side encryption with AWS Key Management Service keys (SSE-KMS)', and 'Dual-layer server-side encryption with AWS Key Management Service keys (DSS-E-KMS)'. A note below states: 'Secure your objects with two separate layers of encryption. For details on pricing, see DSSE-KMS pricing on the Storage tab of the Amazon S3 pricing page.' A blue box highlights the 'Bucket Key' section, which contains a note: 'Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS.' It includes 'Learn more' and 'Disable' (unchecked) and 'Enable' (checked) buttons. A blue box also highlights the 'Advanced settings' button. A note at the bottom states: 'After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.' A yellow box highlights the 'Create bucket' button at the bottom right.

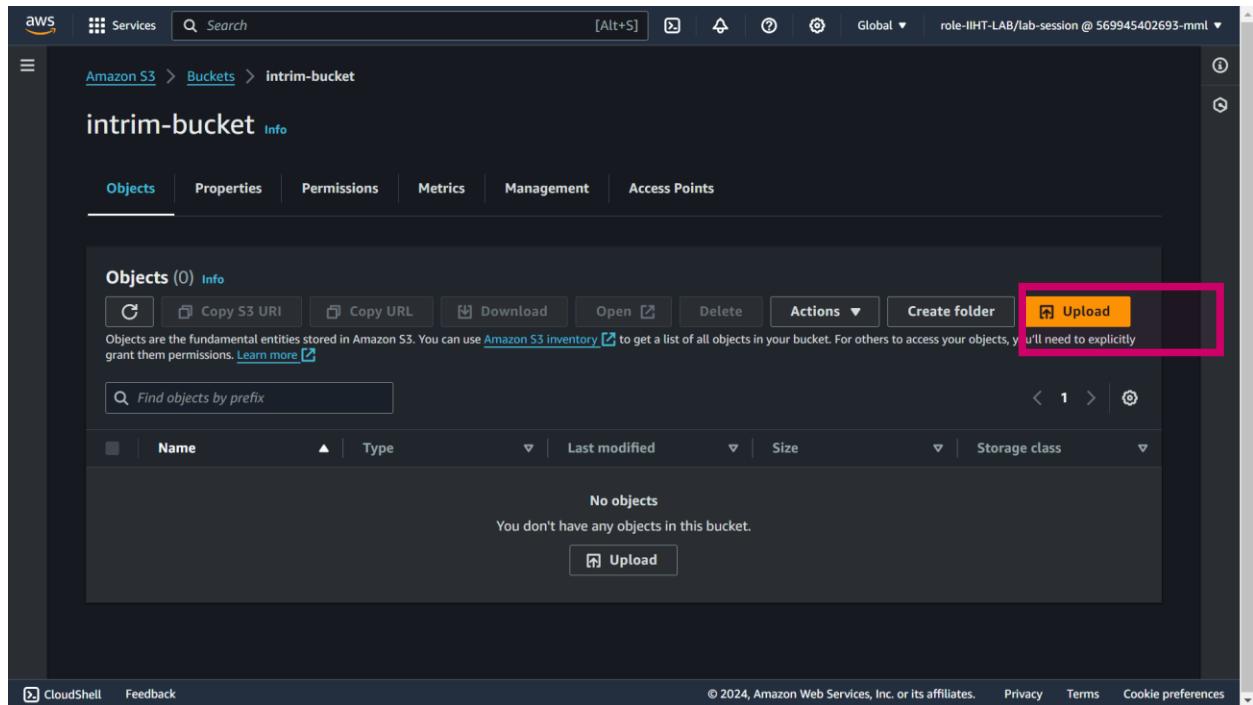


The screenshot shows the AWS S3 service dashboard. At the top, a green success message states: "Successfully created bucket 'intrim-bucket'. To upload files and folders, or to configure additional bucket settings, choose View details." Below this, the "Account snapshot" section displays metrics: Total storage (Pending), Object count (Pending), and Average object size (Pending). A note says, "You can enable advanced metrics in the 'default-account-dashboard' configuration." Below the snapshot, there are tabs for "General purpose buckets" (selected) and "Directory buckets". Under "General purpose buckets", there is a table with one row:

Name	AWS Region	Access	Creation date
intrim-bucket	Asia Pacific (Mumbai) ap-south-1	Objects can be public	March 21, 2024, 16:06:53 (UTC+05:30)

At the bottom of the page, there are links for CloudShell, Feedback, and a footer with copyright information: "© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".

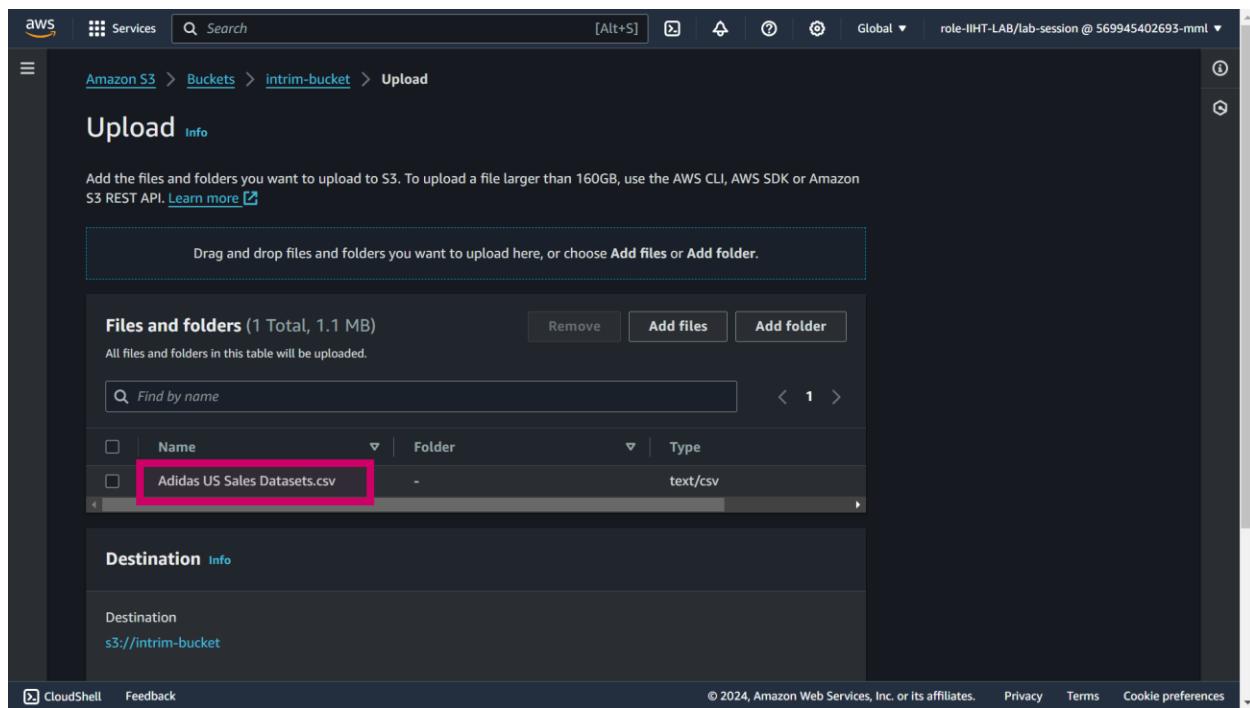
Step6: Uploading .csv file into S3 bucket



The screenshot shows the "intrim-bucket" page within the AWS S3 service. The navigation path is "Amazon S3 > Buckets > intrim-bucket". The main area is titled "intrim-bucket" and contains tabs for Objects, Properties, Permissions, Metrics, Management, and Access Points. The "Objects" tab is selected, showing a table with one row:

Name	Type	Last modified	Size	Storage class
No objects				

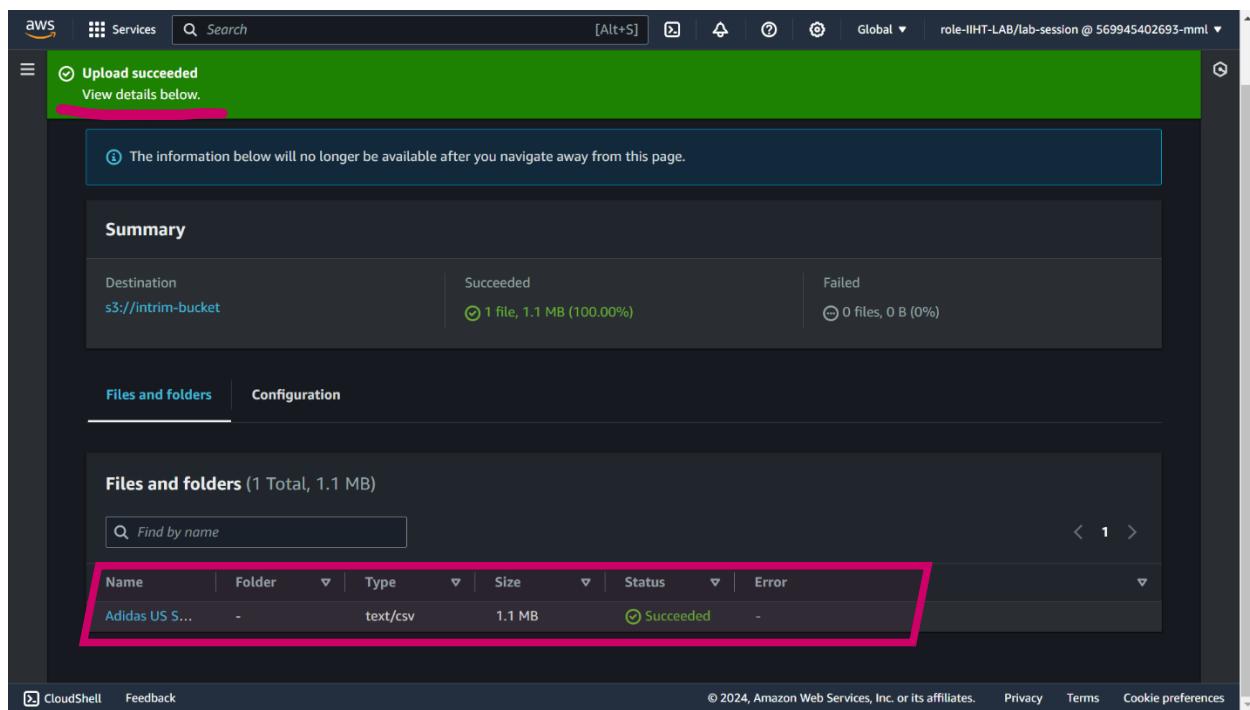
Below the table, a message says, "You don't have any objects in this bucket." There is a prominent yellow "Upload" button with a cloud icon. Above the table, there are several actions: Copy S3 URI, Copy URL, Download, Open, Delete, Actions, Create folder, and another Upload button. A note at the top of the objects table reads: "Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions." At the bottom of the page, there are links for CloudShell, Feedback, and a footer with copyright information: "© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".



The screenshot shows the AWS S3 'Upload' interface. At the top, the navigation bar includes 'Services' (with 'Amazon S3' selected), 'Search', and various global settings. The main area shows the path 'Amazon S3 > Buckets > intrim-bucket > Upload'. Below this, the 'Upload' section has a sub-header 'Info'. A note says: 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)'.

A large dashed box area is labeled 'Drag and drop files and folders you want to upload here, or choose Add files or Add folder.' Below this, a table titled 'Files and folders (1 Total, 1.1 MB)' lists one item: 'Adidas US Sales Datasets.csv' (text/csv). There are 'Remove', 'Add files', and 'Add folder' buttons above the table. The table has columns for 'Name', 'Folder', and 'Type'.

The 'Destination' section shows 'Destination' set to 's3://intrim-bucket'. At the bottom, there are links for 'CloudShell', 'Feedback', and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates.' and 'Cookie preferences'.



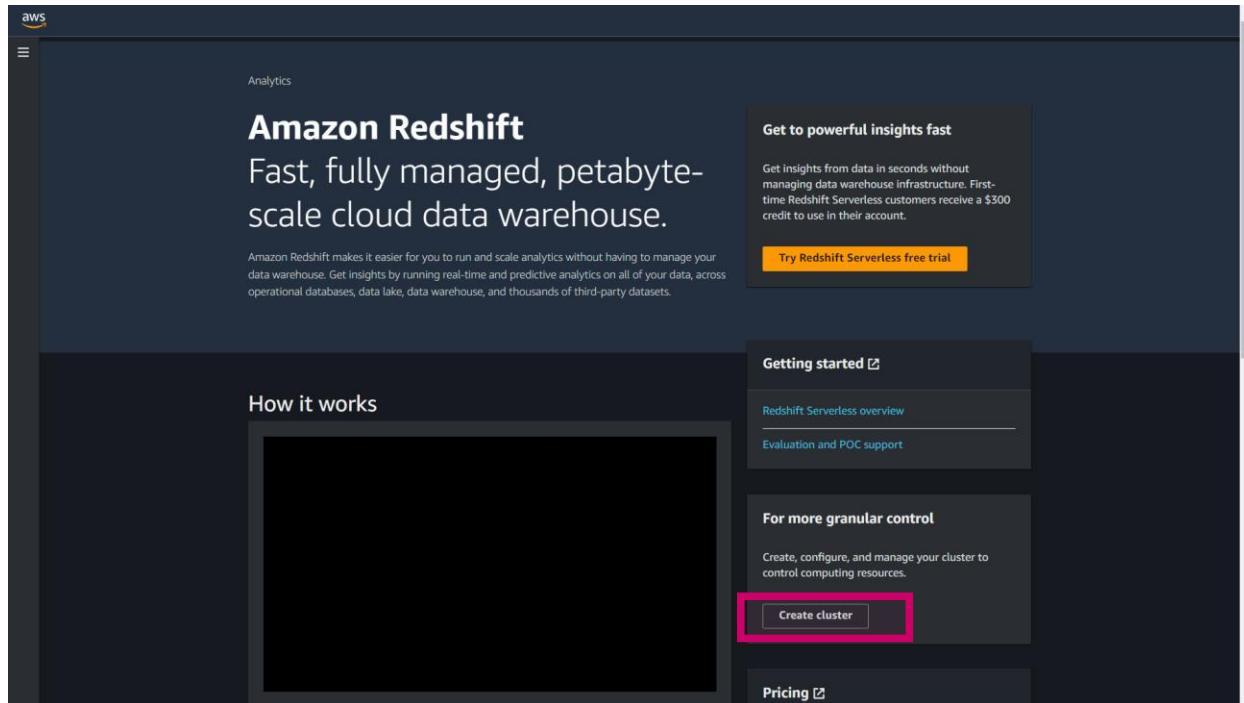
The screenshot shows the AWS S3 'Upload successful' summary page. The top navigation bar and global settings are identical to the previous screenshot. A green banner at the top says 'Upload succeeded' with a link to 'View details below.' Below this, a message states: 'The information below will no longer be available after you navigate away from this page.'

The 'Summary' section shows the destination 's3://intrim-bucket' and upload results: 'Succeeded' (1 file, 1.1 MB (100.00%)) and 'Failed' (0 files, 0 B (0%)).

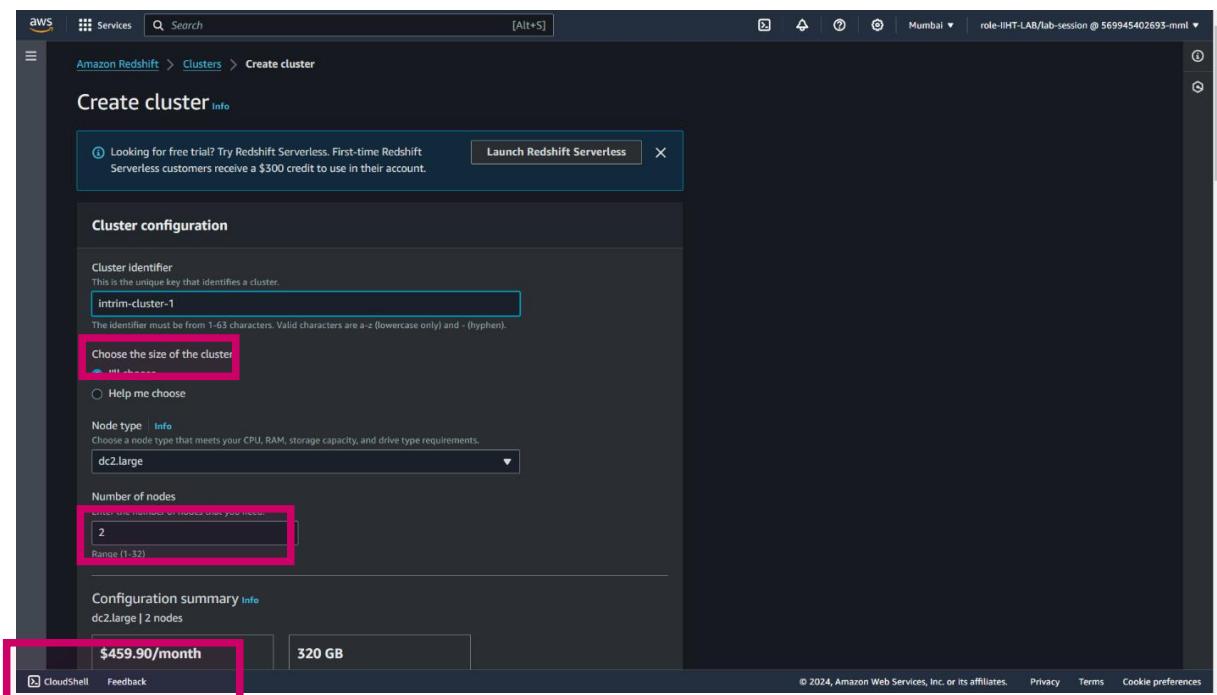
The 'Files and folders' section displays the uploaded file 'Adidas US S...' (text/csv, 1.1 MB) with a status of 'Succeeded'. This table has columns for 'Name', 'Folder', 'Type', 'Size', 'Status', and 'Error'.

At the bottom, there are links for 'CloudShell', 'Feedback', and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates.' and 'Cookie preferences'.

Step 7: Creating Cluster



The screenshot shows the Amazon Redshift landing page. At the top, there's a banner with the text "Get to powerful insights fast" and a button "Try Redshift Serverless free trial". Below the banner, there's a section titled "How it works" with a large image placeholder. To the right, there are three main sections: "Getting started", "For more granular control", and "Pricing". The "Getting started" section includes links to "Redshift Serverless overview" and "Evaluation and POC support". The "For more granular control" section has a "Create cluster" button. The "Pricing" section is partially visible.



The screenshot shows the "Create cluster" page in the AWS Management Console. The URL in the address bar is "Amazon Redshift > Clusters > Create cluster". The page has a header "Create cluster" with a "Info" link. A callout box says "Looking for free trial? Try Redshift Serverless. First-time Redshift Serverless customers receive a \$300 credit to use in their account." and a "Launch Redshift Serverless" button. The main form is titled "Cluster configuration". It includes fields for "Cluster identifier" (containing "intrrim-cluster-1"), "Choose the size of the cluster" (with a dropdown menu showing "dc2.large" highlighted), "Help me choose" (radio button), "Node type" (dropdown menu showing "dc2.large"), "Number of nodes" (input field containing "2" highlighted with a red box), and "Configuration summary" (showing "dc2.large | 2 nodes"). At the bottom, there's a summary table with two rows: one for "CloudShell" (cost "\$459.90/month") and one for "Feedback". The entire "CloudShell" row is highlighted with a thick red border.

Step 8: Creating Cluster Subnet Group

aws Services Search [Alt+S] Mumbai role-IIHT-LAB/lab-session @ 569945402693-mm1

Cluster subnet group details

Name You can't modify the name after your subnet group has been created.

Description

Add subnets

VPC Choose the VPC that contains the subnets that you want to include in your cluster subnet

Availability Zone	Subnet	Action
Choose an Availability Zone	Subnet-07e203...	Remove

Subnets in this cluster subnet group (1)

Availability Zone	Subnet ID	CIDR block	IPv6 CIDR block	Action
ap-south-1a	subnet-07e203...	172.31.32.0/20	-	<input type="button" value="Remove"/>

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

aws Services Search [Alt+S] Mumbai role-IIHT-LAB/lab-session @ 569945402693-mm1

Specify the Availability zone to create the cluster in. Otherwise, Amazon Redshift chooses an availability zone for you.

Enhanced VPC routing Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the Internet. [Learn more about getting started cluster in vpc](#)

Turn off Turn on

Publicly accessible

Turn on Publicly accessible Allow public connections to Amazon Redshift.

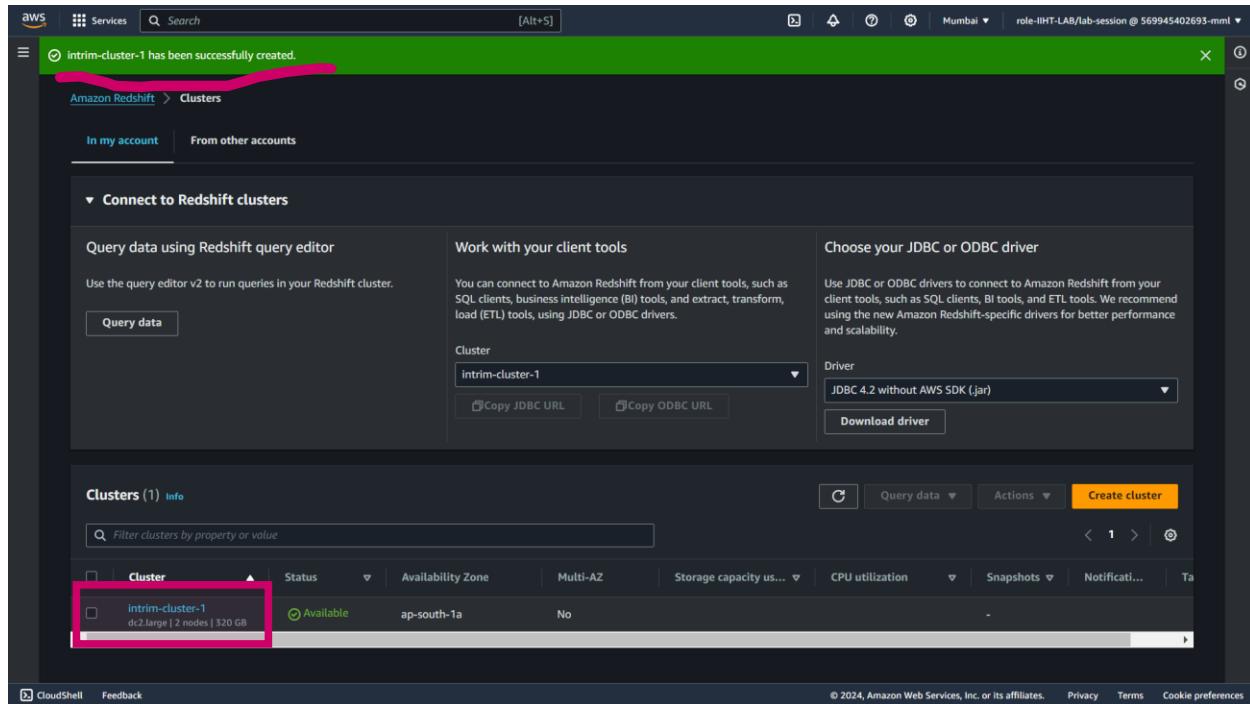
Elastic IP address Select the Elastic IP address for connecting to the cluster.

ⓘ It can take about ten minutes for the setting to change and connections to succeed.

Create cluster

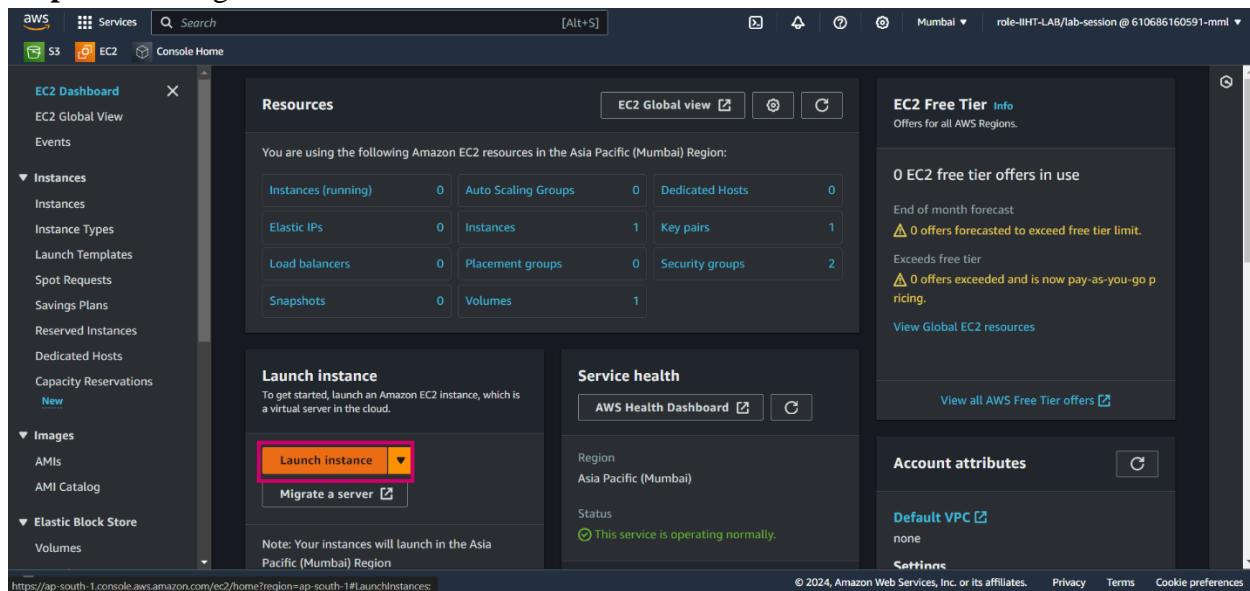
Database configurations Maintenance Monitoring Backup

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



The screenshot shows the AWS Amazon Redshift console. At the top, a green success banner displays the message: "intrim-cluster-1 has been successfully created." Below the banner, the navigation bar includes "Amazon Redshift > Clusters". Under the "Clusters" section, there is a "Connect to Redshift clusters" panel with three tabs: "Query data using Redshift query editor", "Work with your client tools", and "Choose your JDBC or ODBC driver". The "Cluster" dropdown is set to "intrim-cluster-1", and the "Driver" dropdown is set to "JDBC 4.2 without AWS SDK (.jar)". A "Download driver" button is also present. Below this, the main "Clusters (1) Info" table lists one cluster: "intrim-cluster-1" (dc2.large | 2 nodes | 320 GB), which is "Available" in "ap-south-1a". The table includes columns for Cluster, Status, Availability Zone, Multi-AZ, Storage capacity used, CPU utilization, Snapshots, Notifications, and Tags. At the bottom of the page, there are links for "CloudShell", "Feedback", and copyright information: "© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".

Step 9 : Creating an EC2 instance



The screenshot shows the AWS EC2 Dashboard. The left sidebar includes sections for "EC2 Dashboard", "EC2 Global View", "Events", "Instances" (with sub-options like Instances, Instance Types, Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Capacity Reservations), "Images" (with sub-options like AMIs, AMI Catalog), and "Elastic Block Store" (with sub-options like Volumes). The main content area is titled "Resources" and displays statistics for Amazon EC2 resources in the Asia Pacific (Mumbai) Region. It includes tables for Instances (running), Auto Scaling Groups, Dedicated Hosts, Elastic IPs, Instances, Key pairs, Load balancers, Placement groups, Security groups, Snapshots, and Volumes. Below this, there are two main sections: "Launch instance" (with a large "Launch instance" button highlighted with a pink box) and "Service health" (showing the region as Asia Pacific (Mumbai) and status as "This service is operating normally"). To the right, there are sections for "EC2 Free Tier" (0 offers in use, end of month forecast, 0 offers forecasted to exceed free tier limit), "View Global EC2 resources", and "Account attributes" (Default VPC set to none). The bottom of the page shows the URL "https://ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#LaunchInstances:" and copyright information: "© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".

Launch an instance Info

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags Info

Name Add additional tags

Application and OS Images (Amazon Machine Image) Info

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below.

Search our full catalog including 1000s of application and OS images

Quick Start

Amazon Linux macOS Ubuntu Windows Red Hat SUSE Linux Browse more AMIs Including AMIs from

CloudShell Feedback

Summary

Number of instances Info
1

Software Image (AMI)
Canonical, Ubuntu, 22.04 LTS, ...read more
ami-007020fd9c84e18c7

Virtual server type (instance type)
t2.micro

Firewall (security group)
redshiftsg

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 usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

Cancel Launch instance Review commands

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

Instance type Info | Get advice

Instance type
t2.micro Family: t2 1 vCPU 1 GiB Memory Current generation: true Free tier eligible
On-Demand Linux base pricing: 0.0174 USD per Hour
On-Demand Windows base pricing: 0.017 USD per Hour
On-Demand RHEL base pricing: 0.0724 USD per Hour
On-Demand SUSE base pricing: 0.0174 USD per Hour

All generations Compare instance types

Additional costs apply for AMIs with pre-installed software

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 Create new key pair

Network settings Info

Network Info vpc-0c27605fb8c2d0696 | MyVPC
Subnet Info

Summary

Number of instances Info
1

Software Image (AMI)
Canonical, Ubuntu, 22.04 LTS, ...read more
ami-007020fd9c84e18c7

Virtual server type (instance type)
t2.micro

Firewall (security group)
redshiftsg

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 usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

Cancel Launch instance Review commands

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

aws Services Search [Alt+S]

EC2 Console Home

linuxKey Create new key pair

Network settings Info

Network Info
vpc-0c27605fb8c2d0696 | MyVPC

Subnet Info
subnet-0189a90f24a209c18 | Public1A

Auto-assign public IP Info
Enable

Additional charges apply when outside of free tier allowance

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

Common security groups Info
Select security groups

redshiftsg sg-0582302ab92d09061 X
VPC: vpc-0c27605fb8c2d0696

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

Configure storage Info

Advanced

CloudShell Feedback

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

Summary

Number of instances Info
1

Software Image (AMI)
Canonical, Ubuntu, 22.04 LTS, ...read more
ami-007020fd9cb84e18c7

Virtual server type (instance type)
t2.micro

Firewall (security group)
redshiftsg

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 usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

Cancel Launch instance Review commands

Summary

Number of instances Info
1

Software Image (AMI)
Canonical, Ubuntu, 22.04 LTS, ...read more
ami-007020fd9cb84e18c7

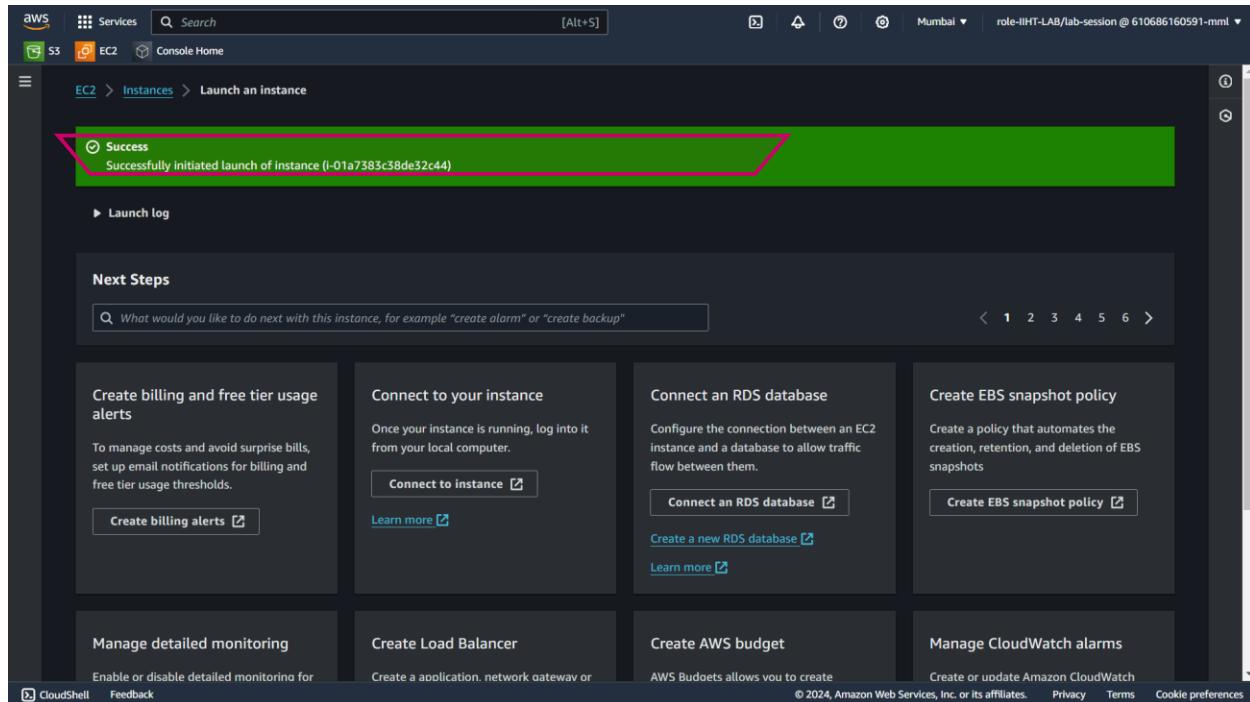
Virtual server type (instance type)
t2.micro

Firewall (security group)
redshiftsg

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 usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

Cancel Launch instance Review commands



Success
Successfully initiated launch of instance (i-01a7383c38de32c44)

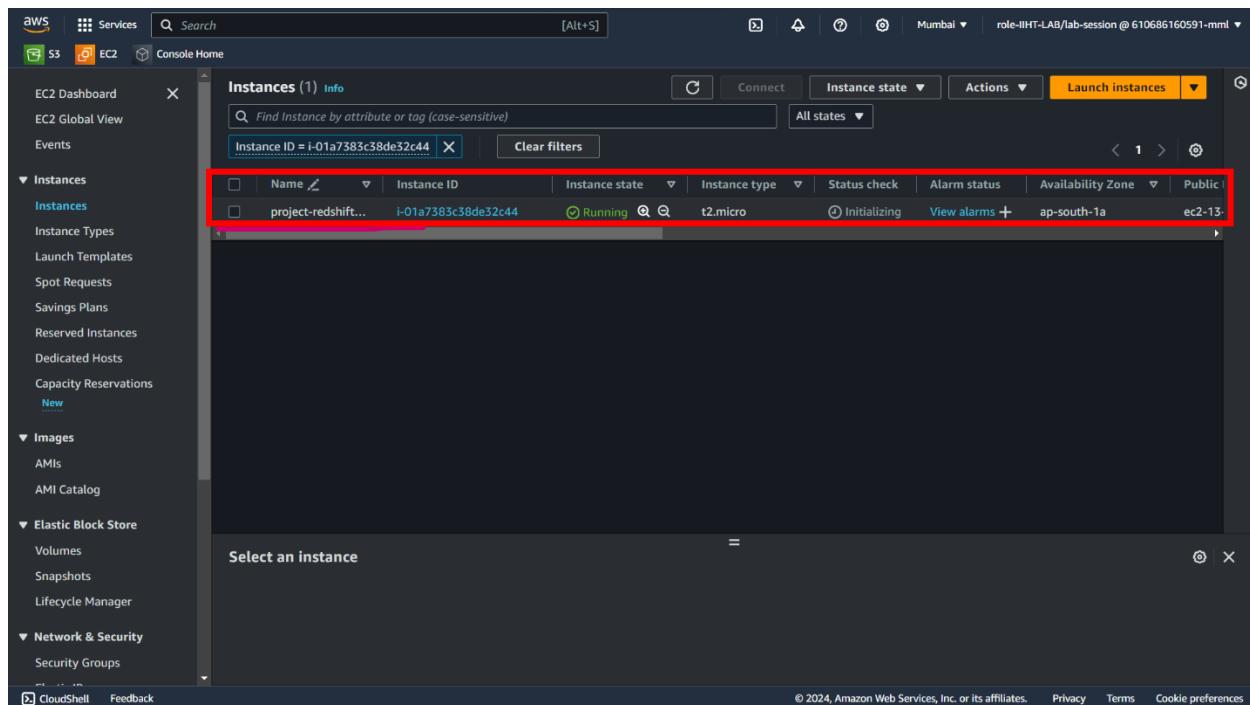
[Launch log](#)

Next Steps

What would you like to do next with this instance, for example "create alarm" or "create backup"

Create billing and free tier usage alerts To manage costs and avoid surprise bills, set up email notifications for billing and free tier usage thresholds. Create billing alerts	Connect to your instance Once your instance is running, log into it from your local computer. Connect to instance Learn more	Connect an RDS database Configure the connection between an EC2 instance and a database to allow traffic flow between them. Connect an RDS database Create a new RDS database Learn more	Create EBS snapshot policy Create a policy that automates the creation, retention, and deletion of EBS snapshots Create EBS snapshot policy
Manage detailed monitoring Enable or disable detailed monitoring for	Create Load Balancer Create a application, network oateway or	Create AWS budget AWS Budgets allows you to create	Manage CloudWatch alarms Create or update Amazon CloudWatch

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



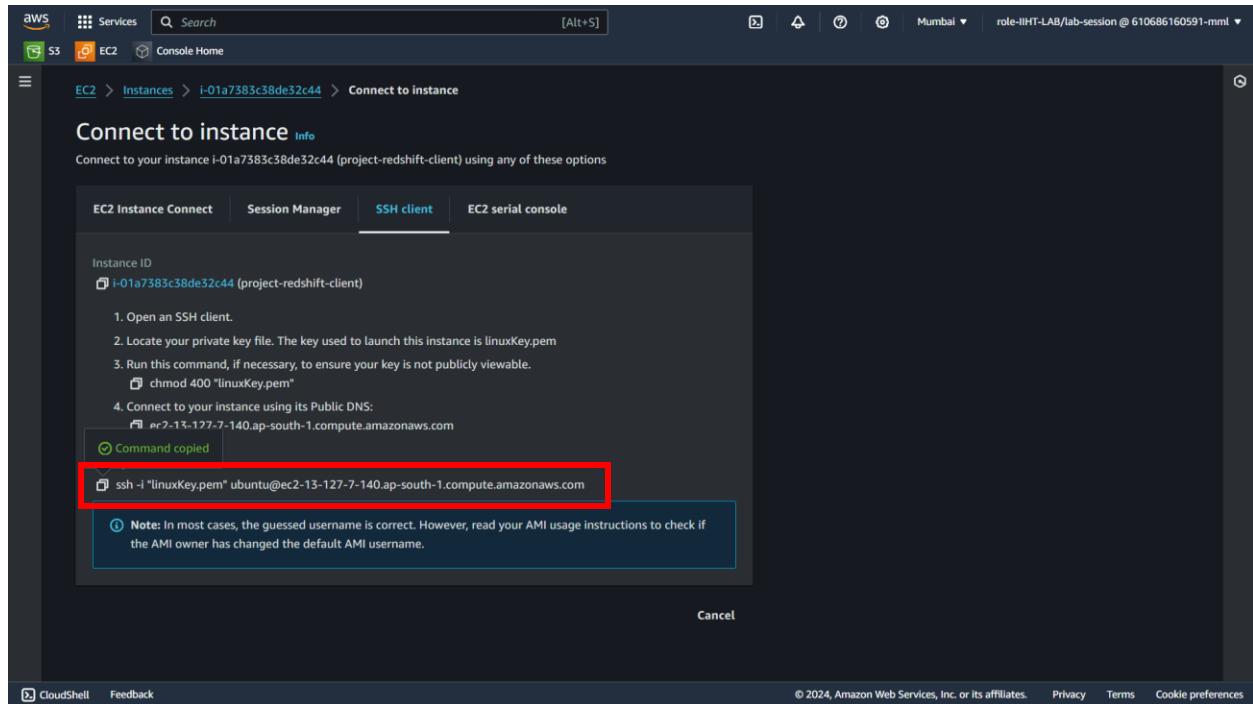
Instances (1) Info

Find Instance by attribute or tag (case-sensitive)
Instance ID = i-01a7383c38de32c44

<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
<input type="checkbox"/>	project-redshift...	i-01a7383c38de32c44	Running	t2.micro	Initializing	View alarms	ap-south-1a	ec2-13-

Select an instance

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



Step 10 : Open terminal and run the commands to create a table in Ec2 instance.

These are the commands to run the table in Ec2 instance.

- “sudo apt-get update
- sudo apt-get install -y postgresql-client && psql --version
- psql -h redshift-cluster-1.cjlluv3z3ar1.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d dev -W”

Syntax : pql -h <hostname> -p <port> -U <user> -d <database> -W

```

PS C:\Users\2320846> cd downloads
PS C:\Users\2320846\downloads> ssh -i "linuxKey.pem" ubuntu@ec2-13-127-7-140.ap-south-1.compute.amazonaws.com
The authenticity of host 'ec2-13-127-7-140.ap-south-1.compute.amazonaws.com (13.127.7.140)' can't be established.
ED25519 key fingerprint is SHA256:iAINGLoqfOLVshcfpgNaja3pJ4SzbJHmNFFV2FVPo0.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-13-127-7-140.ap-south-1.compute.amazonaws.com' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.5.0-1014-aws x86_64)

 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support: https://ubuntu.com/pro

 System information as of Thu Mar 21 14:42:13 UTC 2024

 System load: 0.0      Processes:         98
 Usage of /: 28.4% of 7.57GB   Users logged in:     0
 Memory usage: 19%          IPv4 address for eth0: 10.0.1.38
 Swap usage:  0%

Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-10-0-1-38:~$
```

- Connecting to the Redshift cluster from ec2 through postgreSQL client

```

ubuntu@ip-10-0-1-38:~$ psql -h redshift-cluster-1.c4hapfhrjfor.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d dev -W
Password:
psql (14.11 (Ubuntu 14.11-0ubuntu0.22.04.1), server 8.0.2)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 256, compression: off)
Type "help" for help.
```

- create database for our use

```

dev=# create database project ;
CREATE DATABASE
dev=#

```

- Again we connect to Redshift and login to the created database
- Create a table

```

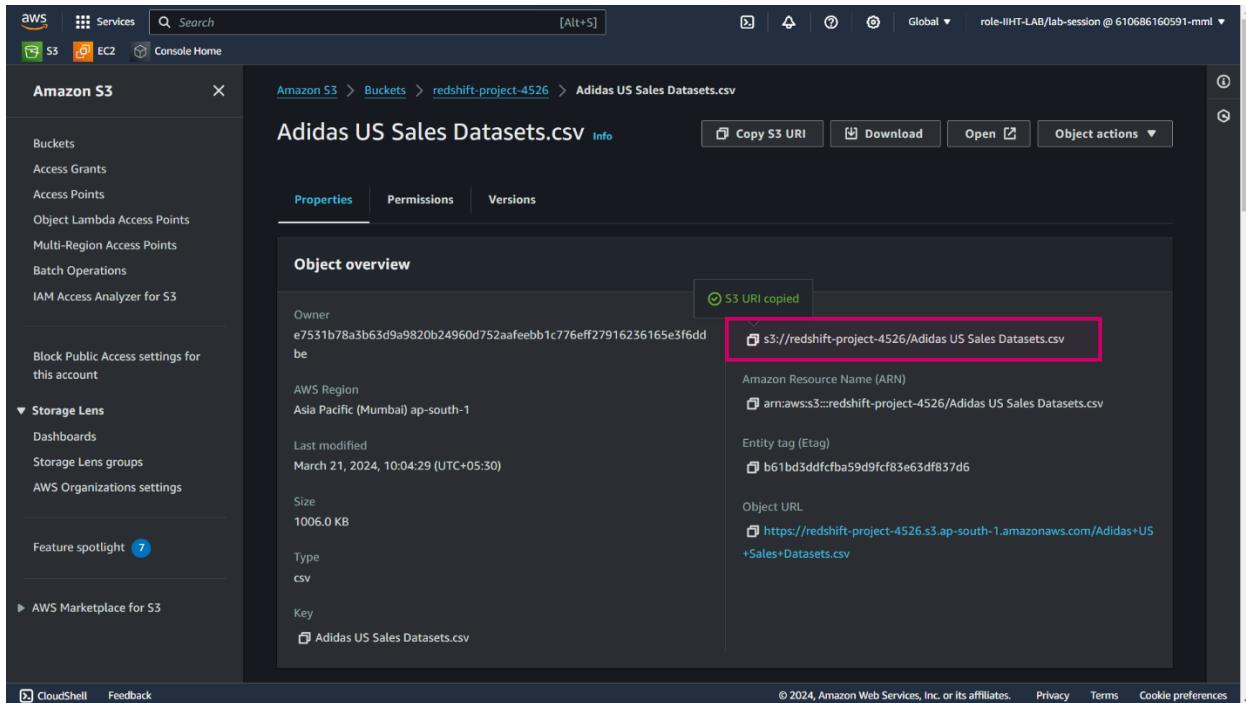
dev@ip-10-0-1-38:~$ psql -h redshift-cluster-1.c4hapfhjrjfor.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d project -W
Password:
psql (14.11 (Ubuntu 14.11-0ubuntu0.22.04.1), server 8.0.2)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 256, compression: off)
Type "help" for help.

project=# \d
Did not find any relations.
project=#
CREATE TABLE public.sales_data (
    retailer character varying(255) ENCODE lzo,
    retailer_id integer ENCODE az64,
    Invoice_date character varying(255) ENCODE lzo,
    region
        character varying(255) ENCODE lzo,
    state character varying(255) ENCODE lzo,
    city character varying(255) ENCODE lzo,
    product character varying(255) ENCODE lzo,
    price_per_unit numeric(10, 2) ENCODE az64,
    units_sold integer ENCODE az64,
    total_sales numeric(10, 2) ENCODE az64,
    operating_profit numeric(10, 2) ENCODE az64,
    operating_margin numeric(5, 2) ENCODE az64,
    sales_method character varying(255) ENCODE lzo
) DISTSTYLE AUTO;
CREATE TABLE
project=#

```

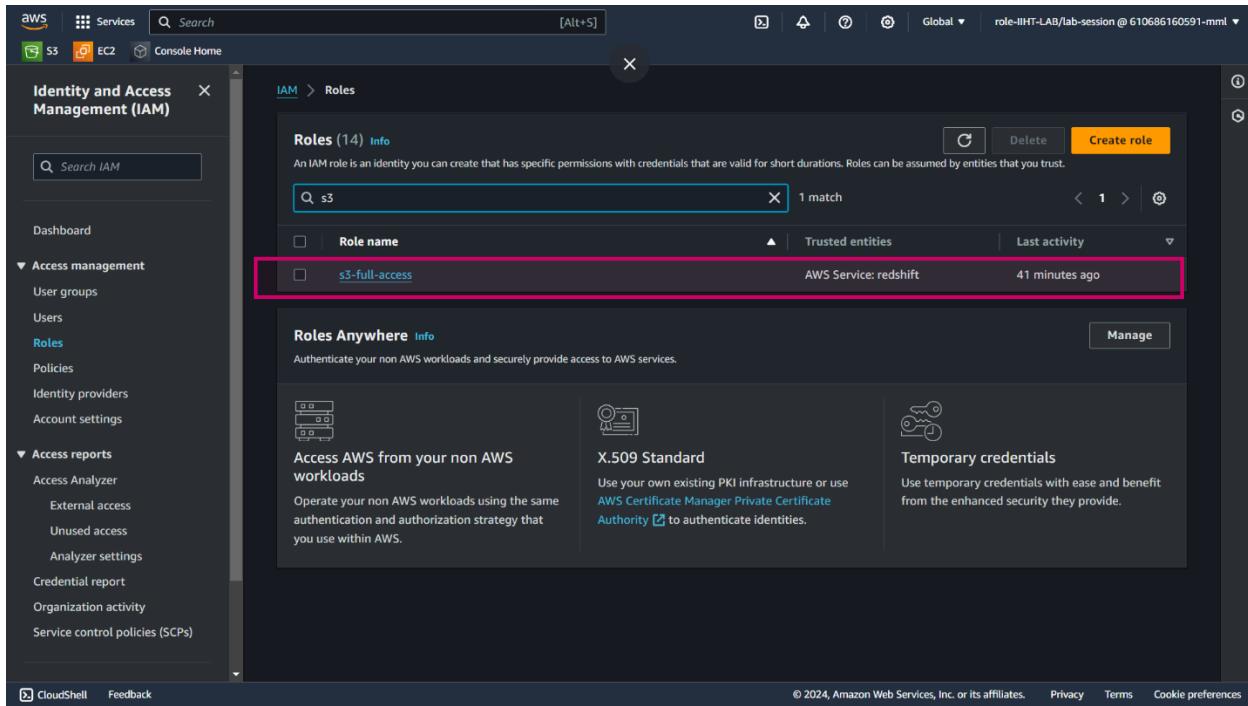
Step 11: Load the data from S3 to Redshift

To load the data we need S3 object URI

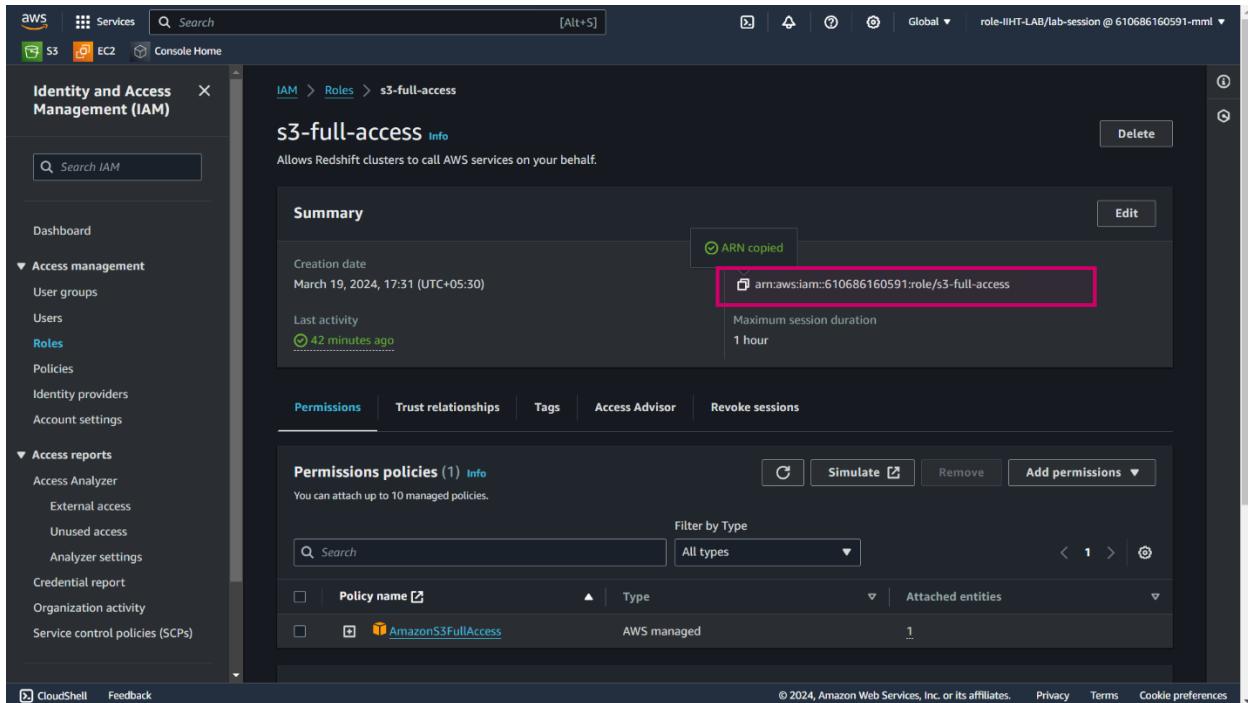


The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with navigation links like Buckets, Access Grants, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, and others. The main area displays a file named 'Adidas US Sales Datasets.csv' located in the bucket 'redshift-project-4526'. The object details page includes sections for Properties, Permissions, and Versions. Under the Properties section, there's an 'Object overview' table with columns for Owner, AWS Region, Last modified, Size, Type, and Key. A prominent 'Copy S3 URI' button is visible above the table, and its resulting value, 's3://redshift-project-4526/Adidas US Sales Datasets.csv', is also highlighted with a pink box.

We need the IAM S3 Role ARN that we created earlier



The screenshot shows the AWS IAM Roles page. A search bar at the top left contains 's3'. Below it, a table lists roles, with 's3-full-access' highlighted by a pink border. The table columns include 'Role name', 'Trusted entities', and 'Last activity'. The 's3-full-access' row shows 'AWS Service: redshift' and '41 minutes ago'. At the bottom right of the table, there is a 'Manage' button.



The screenshot shows the detailed view of the 's3-full-access' role. The ARN of the role, 'arn:aws:iam::610686160591:role/s3-full-access', is highlighted with a pink border. The 'Permissions' tab is selected, showing a single policy named 'AmazonS3FullAccess' attached to the role.

Then use the following command to load the data

```

project=# 
COPY project.public.sales_data FROM 's3://redshift-project-4526/Adidas US Sales Datasets.csv' IAM_ROLE 'arn:aws:iam::610686160591:role/s3-full-access' FORMAT AS CSV DELIMITER ',' QUOTE '\"' IGNOREHEADER 1 REGION AS 'ap-south-1' ;

INFO: Load into table 'sales_data' completed, 9648 record(s) loaded successfully.
COPY
project=# select * from sales_data limit 10 ;

```

COPY <tablename> FROM <S3 bucket URI> IAM_ROLE <IAM role A> FORMAT AS CSV
 DELIMITER<file delimiter> REGION AS <cluster region>

retailer	retailer_id	invoice_date	region	state	city	product	price_per_unit	units_sold	total_sales	operating_profit	operating_margin	sales_method
Foot Locker	1185732	01-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1200	600000.00	300000.00	50.00	In-store
Foot Locker	1185732	02-01-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	1000	500000.00	150000.00	30.00	In-store
Foot Locker	1185732	03-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	1000	400000.00	140000.00	35.00	In-store
Foot Locker	1185732	04-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	850	382500.00	133875.00	35.00	In-store
Foot Locker	1185732	05-01-2020	Northeast	New York	New York	Men's Apparel	60.00	900	540000.00	108000.00	30.00	In-store
Foot Locker	1185732	06-01-2020	Northeast	New York	New York	Women's Apparel	50.00	1000	500000.00	125000.00	25.00	In-store
Foot Locker	1185732	07-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1250	625000.00	312500.00	50.00	In-store
Foot Locker	1185732	08-01-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	900	450000.00	135000.00	30.00	Outlet
Foot Locker	1185732	21-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	950	380000.00	133000.00	35.00	Outlet
Foot Locker	1185732	22-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	825	371250.00	129930.00	35.00	Outlet
Foot Locker	1185732	23-01-2020	Northeast	New York	New York	Men's Apparel	60.00	900	540000.00	162000.00	30.00	Outlet
Foot Locker	1185732	24-01-2020	Northeast	New York	New York	Women's Apparel	50.00	1000	500000.00	125000.00	25.00	Outlet
Foot Locker	1185732	25-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1220	610000.00	305000.00	50.00	Outlet
Foot Locker	1185732	26-01-2020	Northeast	New York	New York	Women's Street Footwear	50.00	950	450000.00	135000.00	35.00	Outlet
Foot Locker	1185732	27-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	950	380000.00	133000.00	35.00	Outlet
Foot Locker	1185732	28-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	800	360000.00	126000.00	35.00	Outlet
Foot Locker	1185732	29-01-2020	Northeast	New York	New York	Men's Apparel	60.00	850	510000.00	153000.00	30.00	Outlet
Foot Locker	1185732	30-01-2020	Northeast	New York	New York	Women's Apparel	50.00	950	475000.00	118750.00	25.00	Outlet
Foot Locker	1185732	31-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1200	600000.00	300000.00	50.00	Outlet
Foot Locker	1185732	01-02-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	900	450000.00	135000.00	30.00	Outlet
(28 rows)												

Database schema :

	Field	Type	NL	CMP
A	retailer	character varying(50)	NULL	lzo
#	retailerid	integer	NULL	az64
A	invoice_date	character varying(50)	NULL	lzo
A	region	character varying(40)	NULL	lzo
A	state	character varying(40)	NULL	lzo
A	city	character varying(40)	NULL	lzo
A	product	character varying(100)	NULL	lzo
#	priceperunit	numeric(10,2)	NULL	az64
#	units_sold	integer	NULL	az64
#	total_sales	numeric(12,2)	NULL	az64
#	operatingprofit	numeric(12,2)	NULL	az64
#	operatingmargin	numeric(5,2)	NULL	az64
A	salesmethod	character varying(50)	NULL	lzo

Step 12: Data Cleaning

Looking at the data we found some data mismatch , so we have to clean the data first before performing any analysis on the data .

- First we have to change the invoice_date colum from character variable to date format

```
-- changing the invoice_date from character varying(50) to date format

-- step 1 : Adding new column
ALTER TABLE sales_data
ADD COLUMN order_date DATE;

-- Step 2: Update the new column with the converted values
UPDATE sales_data
SET order_date = TO_DATE(invoice_date, 'DD-MM-YYYY');

-- Step 3: Drop the old column
ALTER TABLE sales_data
DROP COLUMN INVOICE_DATE;

SELECT * FROM SALES_DATA;
```

Now the schema looking like this

	Field	Type	NL	CMP
A	retailer	character varying(50)	NULL	Izo
#	retailerid	integer	NULL	az64
A	region	character varying(40)	NULL	Izo
A	state	character varying(40)	NULL	Izo
A	city	character varying(40)	NULL	Izo
A	product	character varying(100)	NULL	Izo
#	priceperunit	numeric(10,2)	NULL	az64
#	units_sold	integer	NULL	az64
#	total_sales	numeric(12,2)	NULL	az64
#	operatingprofit	numeric(12,2)	NULL	az64
#	operatingmargin	numeric(5,2)	NULL	az64
A	salesmethod	character varying(50)	NULL	Izo
#	order_date	date	NULL	az64

- We observed that in the dataset the some column's data was mismatching , these columns are

- Total_Sales column
- Operating profit column

```
select priceperunit , units_sold , total_sales , operatingprofit , operatingmargin from sales_data ;
```

	priceperunit	units_sold	total_sales	operatingprofit	operatingmargin
□	50	1200	600000	300000	50
□	50	1000	500000	150000	30
□	40	1000	400000	140000	35
□	45	850	382500	133875	35
□	60	900	540000	162000	30
□	50	1000	500000	125000	25
□	50	1250	625000	312500	50

So we perform some more data cleaning for

Total_Sales :

```
25 -- Data cleaning for total sales column
26
27 -- ADDING A NEW COL
28 ALTER TABLE sales_data
29 ADD COLUMN TOTAL_COST INT;
30
31
32 -- MANUALLY MULTIPLY UNITS SOLD * PRICE PER UNIT
33 UPDATE sales_data
34 SET TOTAL_COST = PRICEPERUNIT * UNITS SOLD;
35
36
37 --DROP OLD COLUMN
38 ALTER TABLE SALES_DATA
39 DROP COLUMN TOTAL_SALES;
40
41
42 -- RENAME COLUMN TO THE ORIGINAL NAME
43 ALTER TABLE SALES_DATA
44 RENAME TOTAL_COST TO TOTAL_SALES;
45
```

	priceperunit	units_sold	total_sales	operatingprofit	operatingmargin
□	50	1200	60000	300000	50
□	50	1000	50000	150000	30
□	40	1000	40000	140000	35
□	45	850	38250	133875	35
□	60	900	54000	162000	30
□	50	1000	50000	125000	25

Operating profit :

```

52  -- Data cleaning for operating profit column
53
54
55  -- add NEW PROFIT col
56  ALTER TABLE sales_data
57  ADD COLUMN new_profit INT;
58
59  -- MANUALLY CALCULATING THE PROFIT
60  UPDATE sales_data
61  SET new_profit = ( operatingmargin * total_sales ) / 100 ;
62
63  --DROP OLD COLUMN OPERATING_MARGIN
64  ALTER TABLE SALES_DATA
65  DROP COLUMN operatingprofit;
66
67  -- RENAME THE NEW COL TO THE ORIGINAL COLUMN NAME
68  ALTER TABLE SALES_DATA
69  RENAME new_profit TO operating_profit;
70

```

	priceperunit	units_sold	total_sales	operating_profit	operatingmargin
□	50	1200	60000	30000	50
□	50	1000	50000	15000	30
□	40	1000	40000	14000	35
□	45	850	38250	13387	35
□	60	900	54000	16200	30
□	50	1000	50000	12500	25

Now we will move to Redshift cluster to run the queries

Note: We can directly run queries from the instance but for better visuals we used Redshift's query editor v2

Step 13: Running SQL queries to analyze the sales trends

SQL queries to aggregate sales data by product, region, and time period.

```

1
2  -- aggregate sales data by product
3  SELECT
4    product,
5    SUM(total_sales) AS total_sales
6  FROM sales_data
7    GROUP BY product
8    Order by total_sales DESC;
9
10

```

Result 1 (6)

	product	total_sales
□	Men's Street Footwear	27680769
□	Women's Apparel	23870985
□	Men's Athletic Footwear	20577180
□	Women's Street Footwear	17201563
□	Men's Apparel	16520632
□	Women's Athletic Footwear	14315521

By Region

```

7   --aggregate sales data by region
8 ↴ SELECT
9     product, region ,
10    SUM(total_sales) AS total_sales
11 ↴ FROM sales_data
12   GROUP BY region , product
13   Order by region;
14

```

Result 1 (30)

	product	region	total_sales
□	Men's Apparel	Midwest	2223786
□	Women's Apparel	Midwest	3453008
□	Men's Street Footwear	Midwest	4707360
□	Men's Athletic Footwear	Midwest	2619289
□	Women's Street Footwear	Midwest	1997448
□	Women's Athletic Footwear	Midwest	1673543
□	Men's Street Footwear	Northeast	6841324
□	Men's Athletic Footwear	Northeast	3895862
□	Women's Street Footwear	Northeast	3152823
□	Women's Athletic Footwear	Northeast	2668013
□	Men's Apparel	Northeast	3475037
□	Women's Apparel	Northeast	5045208
□	Men's Apparel	South	2811194
□	Women's Apparel	South	4224937
□	Men's Street Footwear	South	4048261
□	Men's Athletic Footwear	South	3647045
□	Women's Street Footwear	South	3242822
□	Women's Athletic Footwear	South	2629097
□	Men's Apparel	Southeast	3183237
□	Women's Apparel	Southeast	4109786
□	Men's Street Footwear	Southeast	4693836
□	Men's Athletic Footwear	Southeast	3653645
□	Women's Street Footwear	Southeast	3059884
□	Women's Athletic Footwear	Southeast	2674048
□	Men's Apparel	West	4827378
□	Women's Apparel	West	7038046
□	Men's Street Footwear	West	7389988
□	Men's Athletic Footwear	West	6761339

By Time period

```

18 -- aggregate sales data by time period
19 ∵ SELECT
20   product ,
21   TO_CHAR(order_date::DATE,'MONTH' ) as month,
22   sum(total_sales) AS total_sales
23 ∵ FROM sales_data
24   GROUP BY product , month , REGION
25   ORDER BY total_sales desc ;
26

```

Virtual Servers in the Cloud 2

Result 1 (100)

product	month	total_sales
Women's Apparel	MAY	903900
Men's Street Footwear	JULY	894762
Men's Street Footwear	MAY	875565
Women's Apparel	JANUARY	870171
Men's Street Footwear	DECEMBER	865236
Women's Apparel	SEPTEMBER	845374
Men's Athletic Footwear	JANUARY	827784
Men's Athletic Footwear	MAY	798378
Men's Athletic Footwear	AUGUST	796036
Women's Street Footwear	SEPTEMBER	777403
Men's Street Footwear	AUGUST	775264
Men's Street Footwear	APRIL	771903
Women's Apparel	APRIL	757109
Men's Athletic Footwear	SEPTEMBER	753147
Men's Street Footwear	NOVEMBER	734340
Men's Street Footwear	JANUARY	733706
Men's Street Footwear	SEPTEMBER	732989
Women's Apparel	JUNE	727454
Men's Athletic Footwear	JULY	720404
Women's Apparel	JULY	718496
Men's Street Footwear	FEBRUARY	717585
Men's Athletic Footwear	FEBRUARY	717550
Women's Street Footwear	JANUARY	711685
Men's Street Footwear	JANUARY	699308
Women's Apparel	AUGUST	676628
Men's Street Footwear	JULY	674139
Women's Apparel	FEBRUARY	650157
Men's Street Footwear	JUNE	642840
Men's Street Footwear	MARCH	633749
Women's Apparel	DECEMBER	630685
Women's Street Footwear	MAY	627273
Women's Apparel	OCTOBER	623236
Men's Athletic Footwear	APRIL	620850
Men's Street Footwear	DECEMBER	620123
Men's Street Footwear	AUGUST	619224
Men's Street Footwear	SEPTEMBER	618857
Women's Street Footwear	FEBRUARY	617181
Women's Street Footwear	JULY	604788
Men's Apparel	MAY	601740
Women's Street Footwear	APRIL	595061
Men's Apparel	JULY	58730
Men's Street Footwear	HIBC	586740

Result 2 (100)

product	month	total_sales
Women's Street Footwear	JULY	699308
Men's Apparel	MAY	601740
Women's Street Footwear	APRIL	595061
Men's Apparel	JULY	58730
Men's Street Footwear	JUNE	586219
Men's Athletic Footwear	AUGUST	581319
Women's Athletic Footwear	MAY	589951
Men's Street Footwear	AUGUST	588716
Men's Street Footwear	MAY	587409
Men's Street Footwear	DECEMBER	582861
Men's Street Footwear	JULY	581740
Men's Athletic Footwear	JUNE	549573
Women's Athletic Footwear	JANUARY	543866
Men's Apparel	JANUARY	541701
Women's Street Footwear	JUNE	539924
Women's Street Footwear	AUGUST	538454
Women's Apparel	NOVEMBER	538436
Women's Apparel	DECEMBER	532298
Women's Apparel	JULY	525885
Men's Street Footwear	MARCH	524016
Women's Apparel	JULY	522292
Men's Apparel	SEPTEMBER	513620
Men's Street Footwear	OCTOBER	511351
Men's Street Footwear	OCTOBER	508353
Men's Street Footwear	APRIL	506990
Men's Street Footwear	JULY	503231
Women's Athletic Footwear	SEPTEMBER	501163
Men's Apparel	FEBRUARY	497137
Women's Apparel	NOVEMBER	495827
Women's Athletic Footwear	JULY	494946
Women's Apparel	SEPTEMBER	494112
Men's Athletic Footwear	DECEMBER	492676
Men's Athletic Footwear	MARCH	487809
Men's Athletic Footwear	SEPTEMBER	485515
Women's Apparel	APRIL	485057
Men's Street Footwear	AUGUST	484315
Women's Athletic Footwear	FEBRUARY	483639
Women's Street Footwear	OCTOBER	483456
Men's Athletic Footwear	OCTOBER	482089
Men's Athletic Footwear	AUGUST	480367
Men's Apparel	AUGUST	478597
Women's Street Footwear	MARCH	474928

Identify top-selling products by calculating total sales or units sold.

By total sales

```

28
29    -- Identify top-selling products by calculating total sales
30  ✓ select
31    ✓   product
32    ✓     from (
33    ✓       select
34      ✓         product ,
35      ✓         sum(total_sales ) as total_sales
36      ✓       from sales_data
37      ✓       group by product
38      ✓       order by total_sales DESC
39    )
40  ;
41

```

	product
□	Men's Street Footwear
□	Women's Apparel
□	Men's Athletic Footwear
□	Women's Street Footwear
□	Men's Apparel
□	Women's Athletic Footwear

By units sold

```

44  -- Identify top-selling products by calculating units sold.
45
46  select
47    ✓   product as top_selling_products
48    ✓     from (
49      ✓       select
50        ✓         product ,
51        ✓         sum(units_sold) as units_sold
52        ✓       from sales_data
53        ✓       group by product
54        ✓       order by units_sold  DESC
55    ) as tab
56  ;
57

```

	top_selling_products
□	Men's Street Footwear
□	Men's Athletic Footwear
□	Women's Apparel
□	Women's Street Footwear
□	Women's Athletic Footwear
□	Men's Apparel

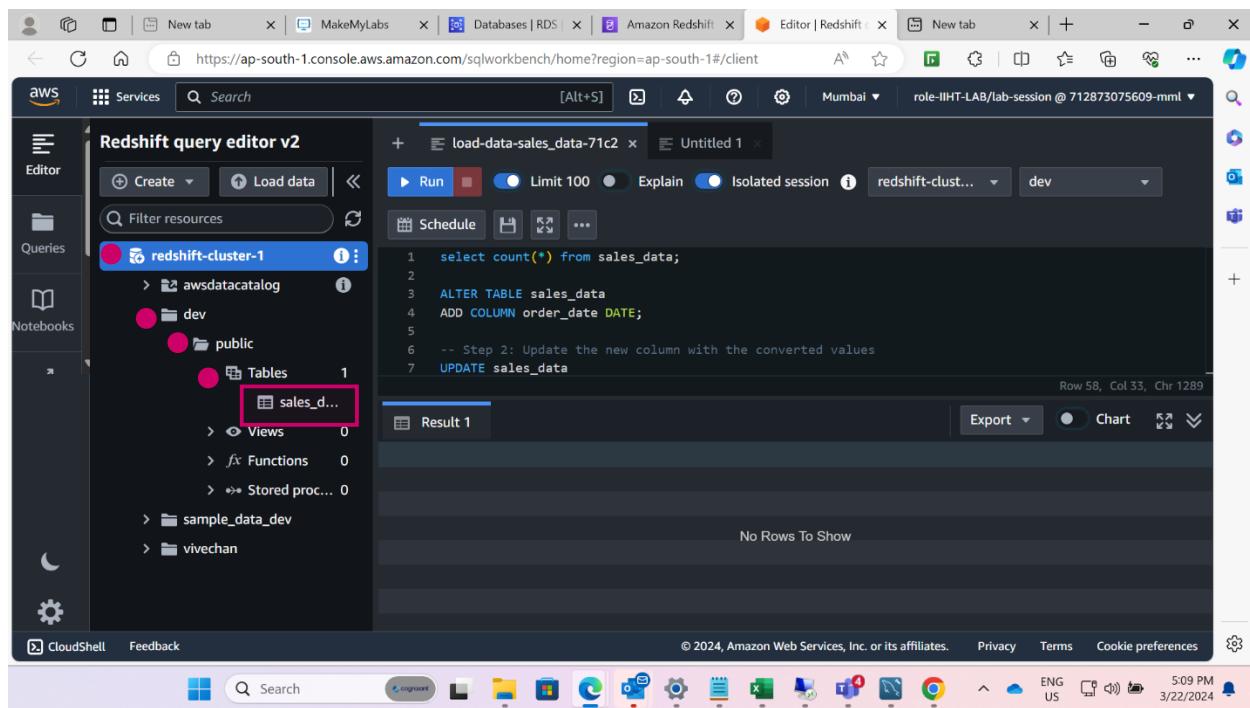
Determine highest revenue-generating regions by analyzing sales revenue per region.

```

55
56      -- Determine the highest revenue generating regions by analyzing sales revenue per region
57
58
59      select
60          region ,
61          sum(total_sales ) as total_sales
62      from sales_data
63      group by region
64      order by total_sales desc ;
65
66
67
68
69

```

	region	total_sales
1	West	36436157
2	Northeast	25078267
3	Southeast	21374436
4	South	20603356
5	Midwest	16674434



The screenshot shows the AWS Redshift Query Editor v2 interface. On the left, the navigation sidebar displays the AWS services, with 'Queries' selected. Under 'Queries', the 'redshift-cluster-1' cluster is expanded, showing its schema. A table named 'sales_d...' is highlighted with a red box. The main workspace contains a query editor window titled 'load-data-sales_data-71c2' with the following SQL code:

```

1  select count(*) from sales_data;
2
3  ALTER TABLE sales_data
4  ADD COLUMN order_date DATE;
5
6  -- Step 2: Update the new column with the converted values
7  UPDATE sales_data

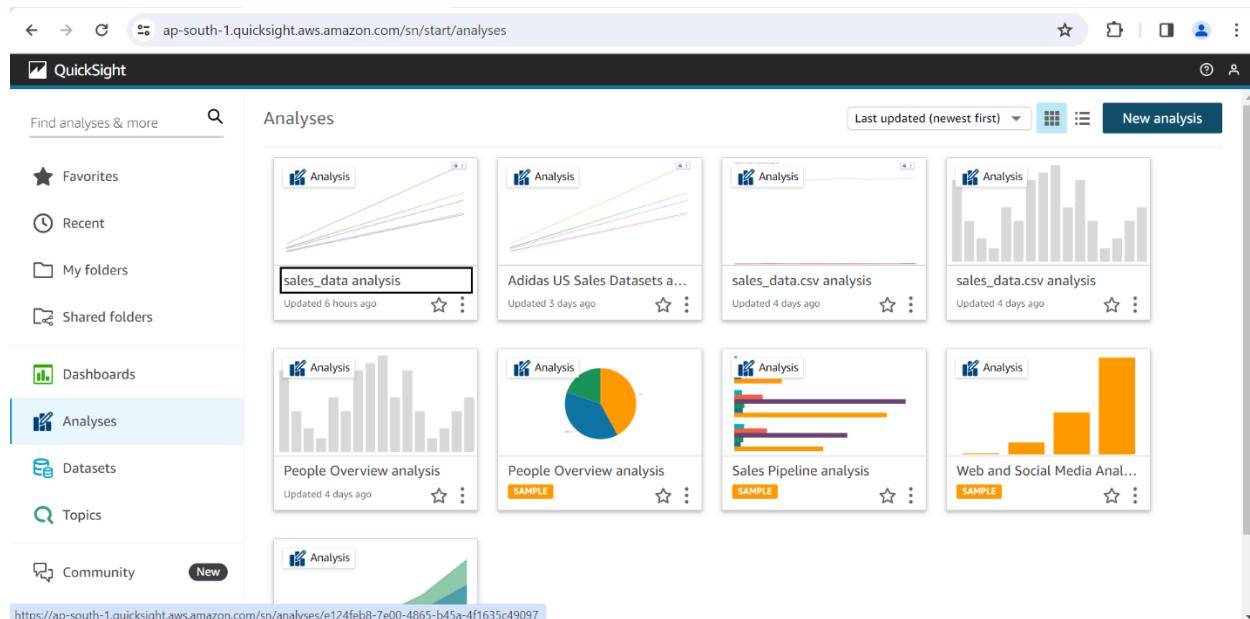
```

The results pane below the editor shows 'No Rows To Show'.

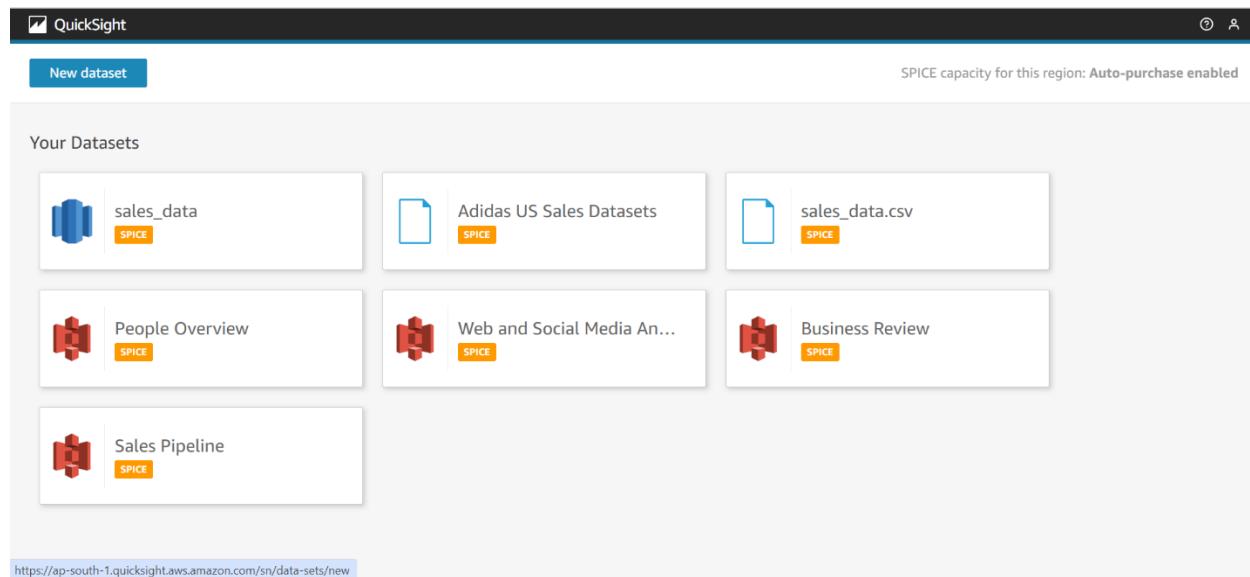
Visualization

For dashboarding we will be using AWS QuickSight Service

We first start importing data to QuickSight from Redshift Database



The screenshot shows the AWS QuickSight interface. On the left, there's a sidebar with navigation links: Favorites, Recent, My folders, Shared folders, Dashboards, Analyses (which is selected), Datasets, Topics, and Community. The main area is titled "Analyses" and displays several card-based visualizations. One card is highlighted with a yellow border: "sales_data analysis" (Updated 6 hours ago). Other cards include "Adidas US Sales Datasets a..." (Updated 3 days ago), "sales_data.csv analysis" (Updated 4 days ago), "People Overview analysis" (Updated 4 days ago), "Sales Pipeline analysis" (Updated 4 days ago), and "Web and Social Media Anal..." (Updated 4 days ago). A "New analysis" button is located at the top right of the analysis grid.



The screenshot shows the AWS QuickSight interface for managing datasets. At the top, there's a "New dataset" button and a note about SPICE capacity: "SPICE capacity for this region: Auto-purchase enabled". Below this, the section "Your Datasets" lists six datasets in cards:

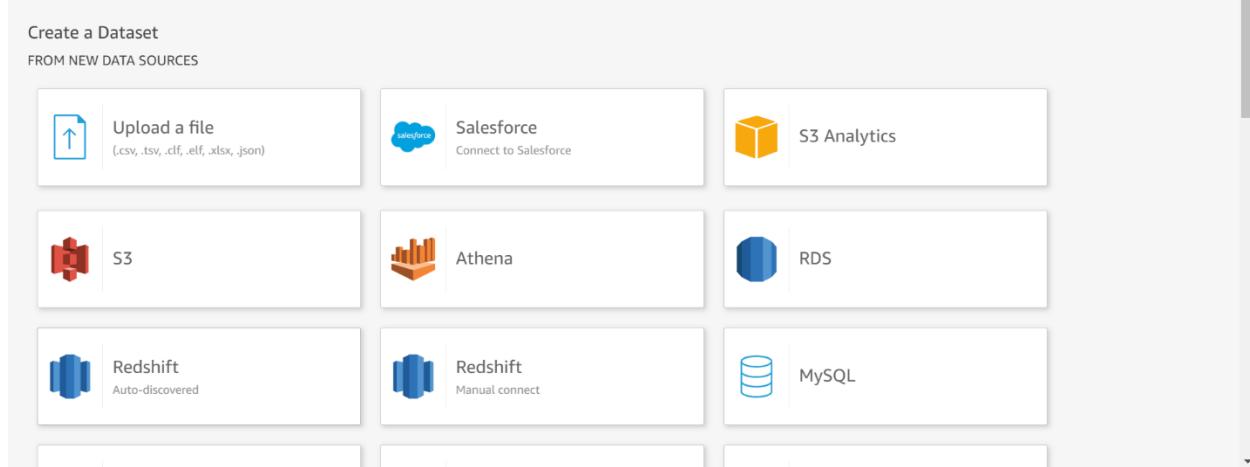
- sales_data** (SPICE)
- Adidas US Sales Datasets** (SPICE)
- sales_data.csv** (SPICE)
- People Overview** (SPICE)
- Web and Social Media An...** (SPICE)
- Sales Pipeline** (SPICE)

QuickSight

Datasets

SPICE capacity for this region: Auto-purchase enabled

Create a Dataset
FROM NEW DATA SOURCES



QuickSight

Datasets

New Redshift data source

Data source name: redshift

Instance ID: redshift-cluster-1

Connection type: Public network

Database name: dev

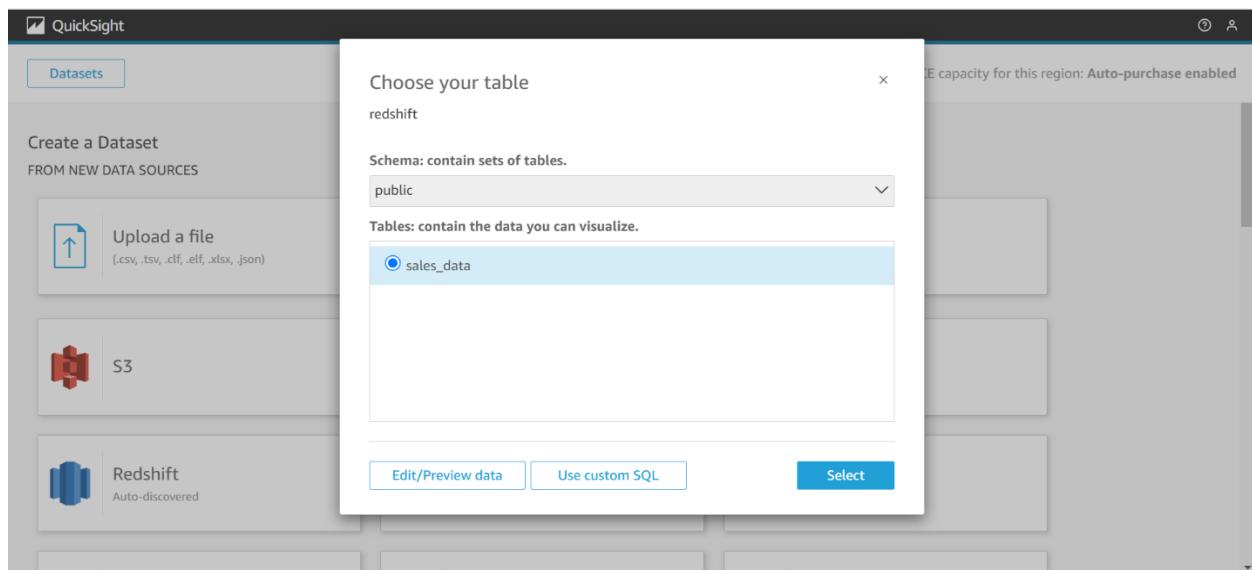
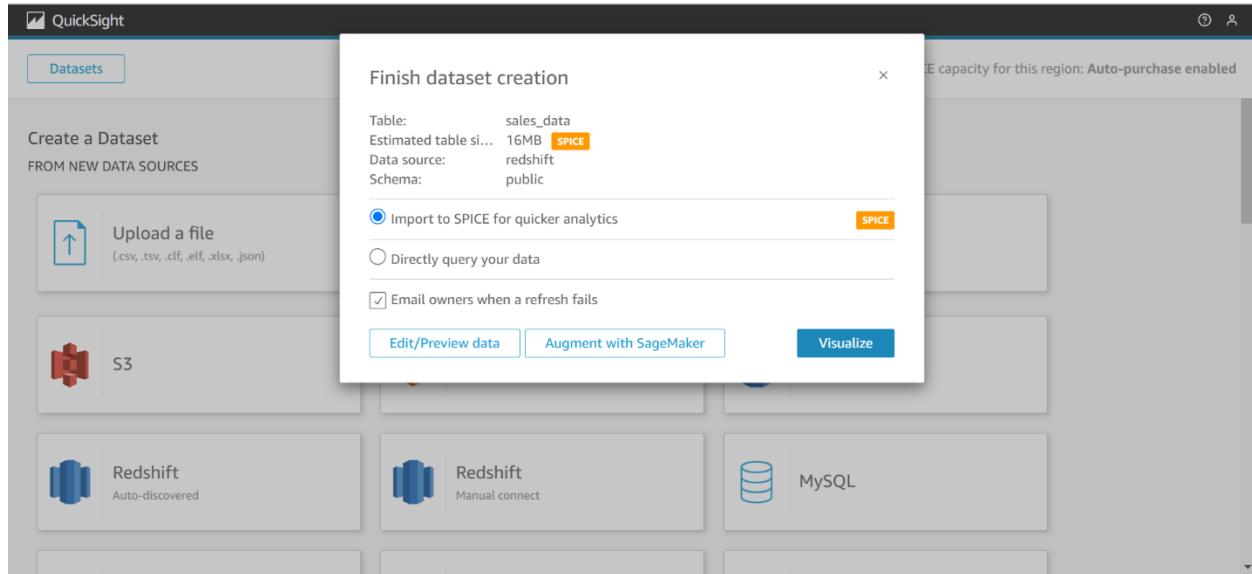
Username: awsuser

Password: *****

Validated SSL is enabled

Create data source

CE capacity for this region: Auto-purchase enabled



QuickSight | sales_data analysis

File Edit Data Insert Sheets Objects Search

Data

Dataset: SPICE sales_data 100%

Search fields:

+ CALCULATED FIELD

- city
- # operatingmargin
- # operatingprofit
- order_date
- # priceperunit
- product
- region
- retailer
- # retailerid
- salesmethod
- state
- # total_sales
- units_sold

Visuals

+ ADD CHANGE VISUAL TYPE

Line chart Bar chart Scatter plot Heatmap Box plot Treemap Funnel chart Gantt chart Map

Sheet 1

Import complete:
100% success
9,648 rows were imported to SPICE
0 rows were skipped

AutoGraph
Add 1 or more fields to build a visual.

MakeMyLabs sales_data analysis AWS VPC Tutorial teradata in kannad 1C OneCognizant Timesheet Landing

QuickSight | sales_data analysis

File Edit Data Insert Sheets Objects Search FIT TO WIDTH PUBLISH NEW LOOK

Data

Dataset: SPICE sales_data 100%

Search fields:

+ CALCULATED FIELD

- city
- # operatingmargin
- # operatingprofit
- order_date
- # priceperunit
- product
- region
- retailer
- # retailerid
- salesmethod
- state
- = sum_max_unit
- # total_sales

Visuals

+ ADD CHANGE VISUAL TYPE Line chart

X AXIS: order_date (QUARTER) Add a dimension

VALUE: units_sold (Sum) Add a measure

COLOR: product Add a dimension

SMALL MULTIPLES: Add a dimension

Sheet 1 Sheet 2

No units sold by month

Sum of Units Sold by Order Date and Product

Sum of Total Sales by Region and Product

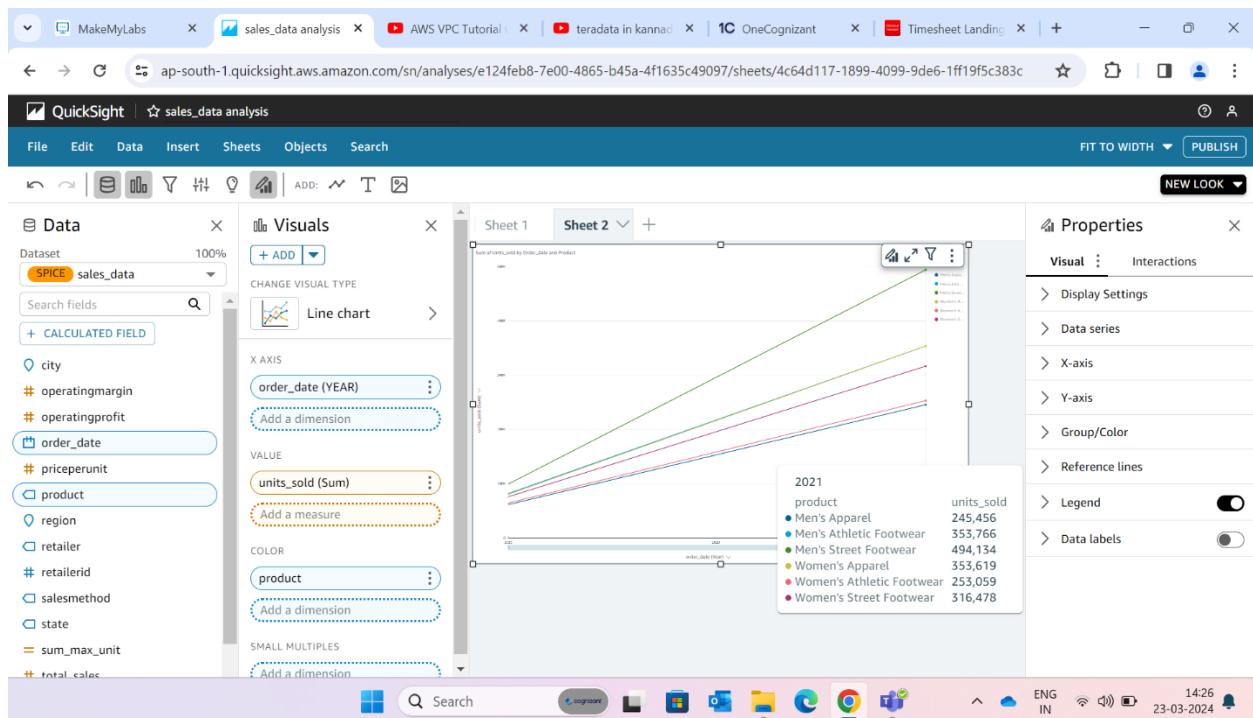
Sum of Total Sales by Product

Properties

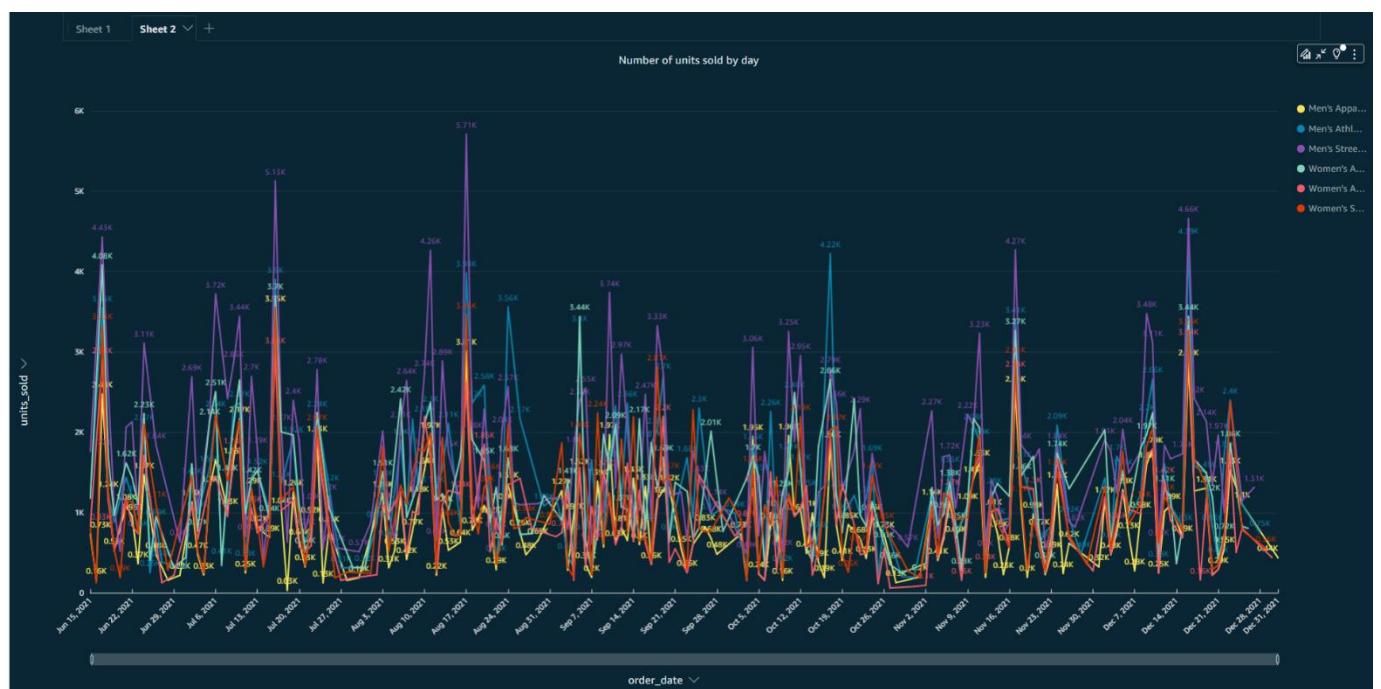
Visual: Interactions

- > Display Settings
- > Data series
- > X-axis
- > Y-axis
- > Group/Color
- > Reference lines
- > Legend
- > Data labels

ENG IN 14:26 23-03-2024



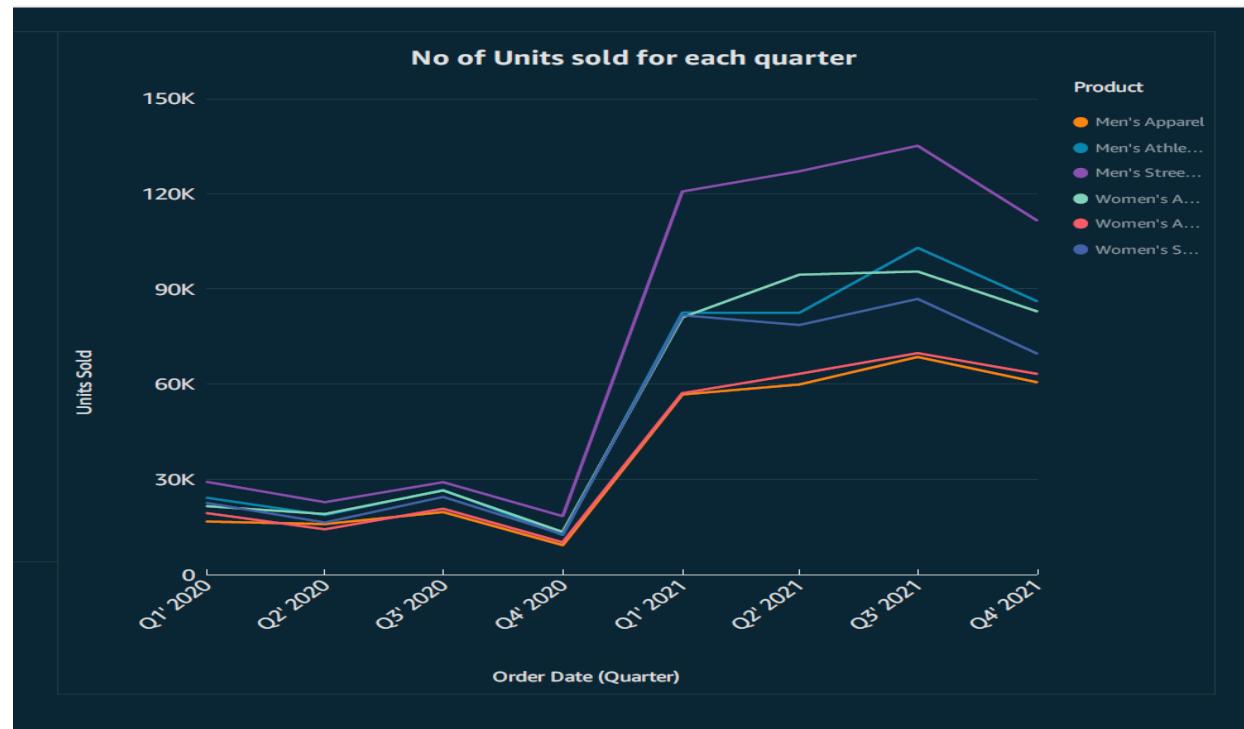
Plotting no of units sold for each day



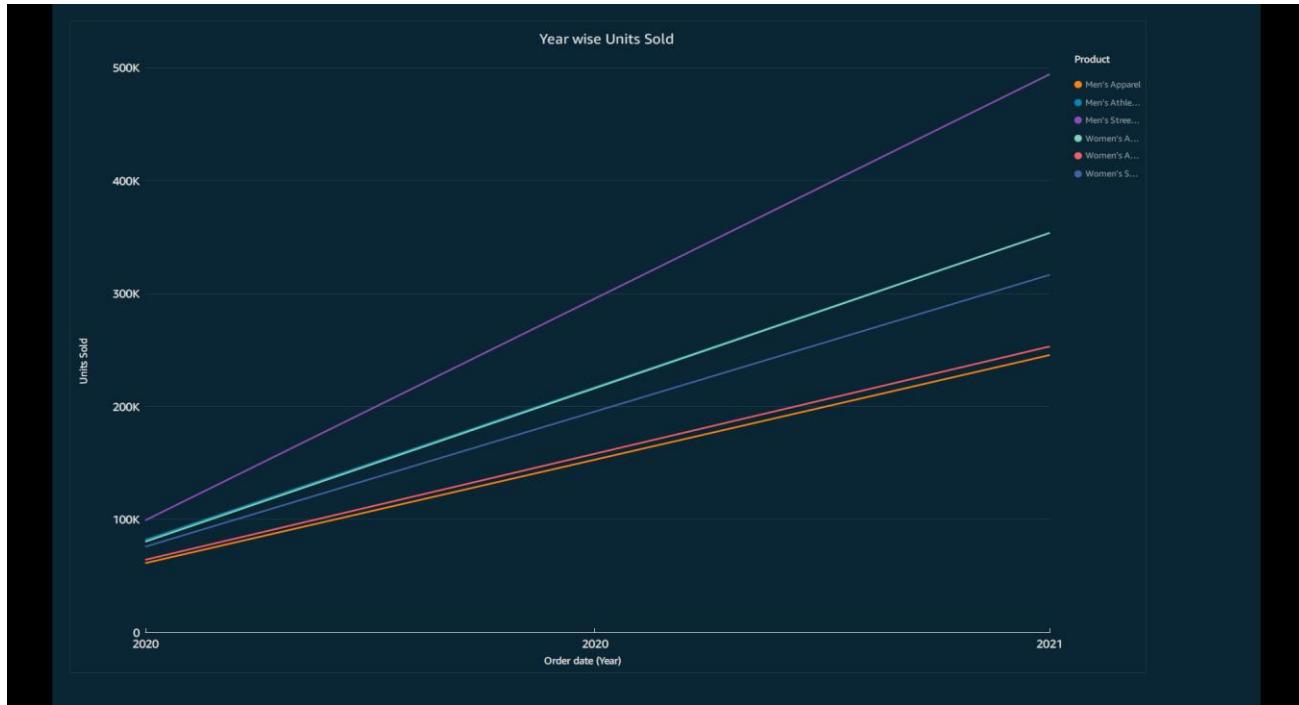
Plotting no of units sold for each month



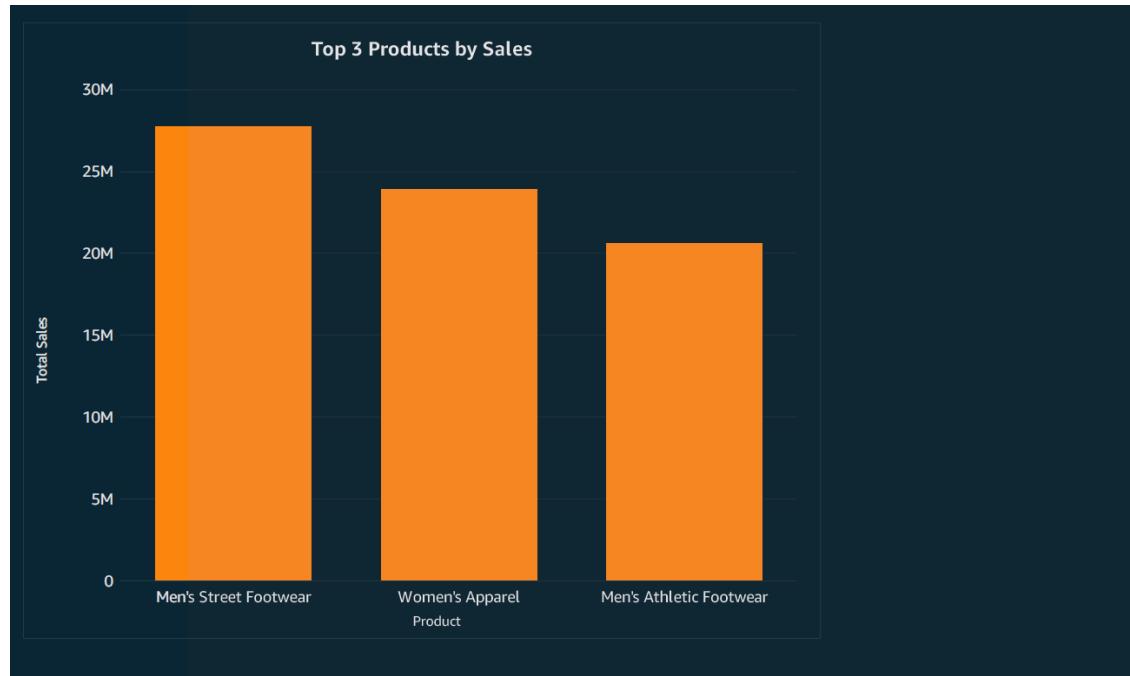
Quarterly wise units sold



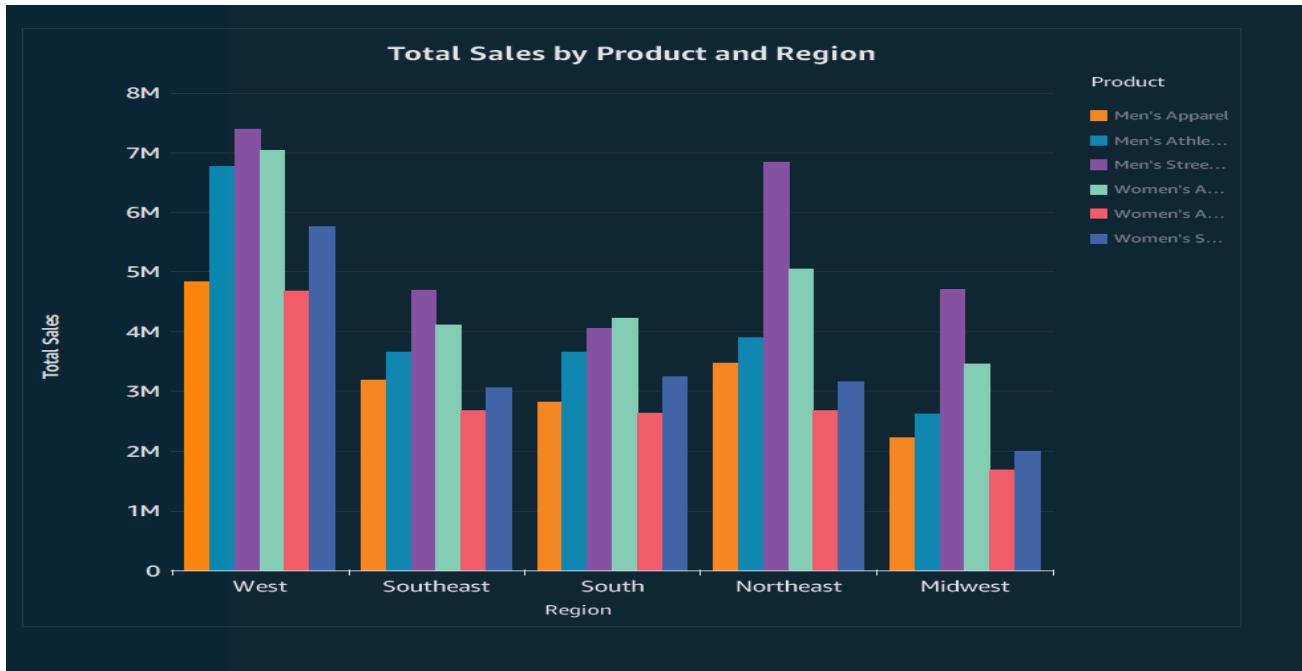
Year wise units sold



Top 3 products by sales



Region wise product sales



Conclusion

1. Data Warehousing with Amazon Redshift: By leveraging Amazon Redshift, we successfully centralized and managed the sales data. Redshift's scalability and performance allowed us to handle large volumes of information efficiently. We transformed raw data into a structured format, enabling complex queries and aggregations. The data warehouse served as the backbone for our analysis, providing a solid foundation for further insights.
2. Visual Insights with Amazon Quick Sight: Amazon Quick Sight played a pivotal role in turning data into actionable insights. We created interactive dashboards and visualizations that allowed stakeholders to explore sales trends, product performance, and customer behavior. Quick Sight's user-friendly interface empowered business users to make informed decisions without relying on technical expertise. Whether tracking revenue, monitoring inventory, or identifying growth opportunities, Quick Sight provided a responsive platform for data exploration.

“Armed with Redshift and Quick Sight, the sales team can gain valuable insights. They can identify top-performing products, can analyze regional sales patterns, and optimize inventory management. By embracing AWS tools, they can elevate their sales analysis capabilities, driving smarter business decisions and enhancing overall performance”.