

## Module 7

### Redshift Assignment

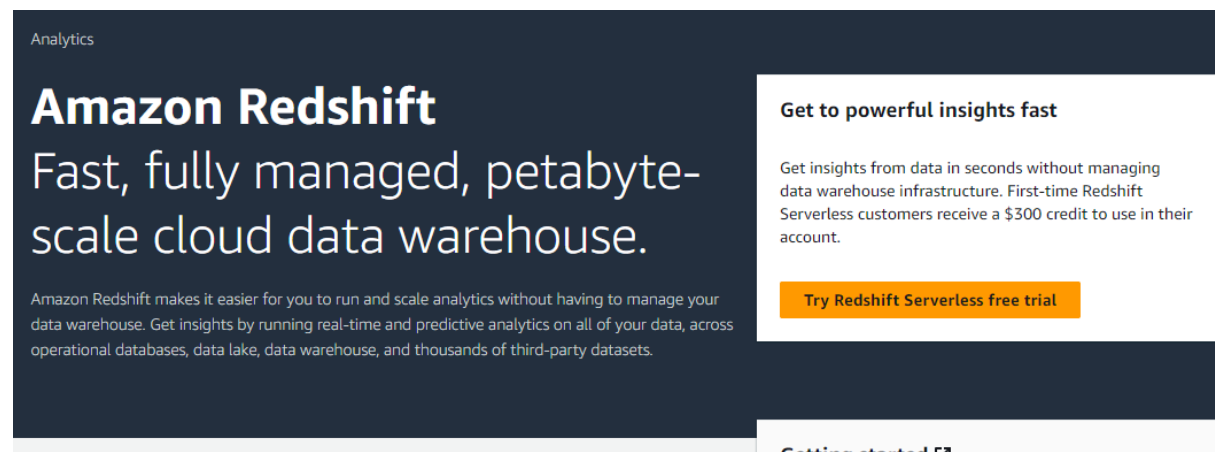
#### **Problem Statement:**

You work for XYZ Corporation. Their application requires a database service that can store data which can be retrieved if required. Implement suitable service for the same. While migrating, you are asked to perform the following tasks:

1. Create a Redshift data warehouse.
2. Using the query editor:
  - a. Load some data
  - b. Query the data

#### **Solutions:**

Open Amazon Redshift

A promotional banner for Amazon Redshift. The background is dark blue. On the left, the text "Analytics" is in small white font. Below it, "Amazon Redshift" is in large white font, followed by "Fast, fully managed, petabyte-scale cloud data warehouse." in a slightly smaller white font. Below that, a paragraph of smaller white text describes the service. On the right, there is a white box with the heading "Get to powerful insights fast" in bold. Below the heading, a paragraph of smaller text describes the benefits. At the bottom of this box is an orange button with the text "Try Redshift Serverless free trial". Below the white box, there is a dark blue bar with the text "Getting started" in white, followed by a small icon of a document with a checkmark.

Analytics

# Amazon Redshift


Fast, fully managed, petabyte-scale cloud data warehouse.

Amazon Redshift makes it easier for you to run and scale analytics without having to manage your data warehouse. Get insights by running real-time and predictive analytics on all of your data, across operational databases, data lake, data warehouse, and thousands of third-party datasets.

**Get to powerful insights fast**

Get insights from data in seconds without managing data warehouse infrastructure. First-time Redshift Serverless customers receive a \$300 credit to use in their account.

**Try Redshift Serverless free trial**

**Getting started** 

# Create workgroup

Amazon Redshift Serverless

Serverless dashboard [Info](#)

↺

Query data [↗](#)

Create workgroup

Namespace overview [Info](#)

Filter namespace

All namespaces ▼

Namespace data from your account

Total snapshots

0

Datashares in my account

0

Datashares requiring authorization

0

Datashares from other accounts

0

Datashares requiring association

0

## Create workgroup

Workgroup

Workgroup is a collection of compute resources from which an endpoint is created. Compute properties include network and security settings.

Workgroup name

This is a unique name that defines the workgroup.

test-wg

The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).

Capacity [Info](#)

Set the base capacity used for processing your data warehouse workloads. Capacity is measured in Redshift processing units (RPUs). Setting the RPU value higher increases capability and improves performance.

Base capacity

The default value is 128 RPUs. To change the base capacity, choose another RPU value.

128 ▼

Range must be 8-512 in increments of 8.

## Network and security

### Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this database.

vpc-09e421fda8e59206f ▼

### VPC security groups

This VPC security group defines which subnets and IP ranges can be used in the VPC.

Choose one or more security groups ▼

sg-067ce7ba4cc4db2fb ✕

### Subnet

The subnet in the chosen VPC that is associated with the specified database.

Choose three or more subnet IDs ▼

subnet-0f757c191c1351ae0 ✕

subnet-0cf105f46bf03a18f ✕

subnet-09000fba630e2f3ca ✕

## Create Namespace for workgroup

## Choose namespace

### Namespace

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

☒ Create a new namespace

☐ Add to an existing namespace

### Namespace

This is a unique name that defines the namespace.

test-ns

The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).

# Review the data and Create

## Step 1: Create workgroup

Edit

### Workgroup

Workgroup is a collection of compute resources from which an endpoint is created. Compute properties include network and security settings.

Workgroup name  
test-wg

### Capacity

Set the base capacity used for processing your data warehouse workloads. Capacity is measured in Redshift processing units (RPU). Setting the RPU value higher increases capability and improves performance.

Base capacity  
128

### Network and security

Virtual private cloud (VPC)  
[vpc-09e421fda8e59206f](#)

VPC security group  
[sg-067ce7ba4cc4db2fb](#)

Subnet  
subnet-0f757c191c1351ae0,  
subnet-0cf105f46bf03a18f,  
subnet-09000fba630e2f3ca,

## Step 2: Choose namespace

Edit

### Namespace

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

Target namespace  
test-ns

Database name and password

Database name  
dev

Admin user credentials  
admin

Permissions

Default IAM role  
Not applied

IAM roles  
Not applied

Encryption and security

AWS KMS encryption  
AWS owned key

Audit logging  
Off

Cancel

Previous

Create

Namespace/ Workgroup is being created

Namespaces / Workgroups <a href="#">Info</a>			
Namespace	Status	Workgroup	Status
test-ns	✔ Available	test-wg	✔ Available

Queries metrics

Workgroup metrics from your account

↺

test-wg ▼

Last hour ▼

View details

## Click on the Query data

Amazon Redshift Serverless > Workgroup configuration > test-wg

test-wg info Actions ▾ Query data [↗](#)

**General information** ↻

Workgroup test-wg	Date created April 20, 2024, 23:17 (UTC+05:30)	Endpoint <a href="#">Not Applicable</a>
Namespace test-ns	Status <span>✓ Available</span>	JDBC URL <a href="#">Not Applicable</a>
Workgroup ARN <a href="#">arn:aws:redshift-serverless:ap-southeast-1:730335274013:workgroup/67cdd9c0-38a9-4dbb-8c16-87da1f9923db</a>	Custom domain name -	ODBC URL <a href="#">Not Applicable</a>
	Patch version <a href="#">patch 180 <a href="#">↗</a></a>	

## Select Federated user and dev Database > Create connection

Redshift query editor v2

+ Create ▾ ⌂ Load data ⏪

↻

> **Serverless: test-wg** ⋮

**Connect to test-wg** ✕

☐ IAM Identity Center  
Connect to Amazon Redshift with your single sign-on credentials from your identity provider (IdP). Your cluster or workgroup must be enabled for IAM Identity Center.

☒ Other ways to connect [Learn more \[↗\]\(#\)](#)

☒ Federated user  
The query editor v2 generates a temporary password to connect to the database.

☐ Database user name and password  
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

☐ AWS Secrets Manager  
Choose a secret with credentials that are associated with the namespace or that you created in AWS Secrets Manager. Only secrets tagged with a key starting with 'Redshift' are listed.

**Database**

The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

Cancel Create connection

To Load Data First create S3 Bucket

So that the data can be loaded through S3 bucket

[Amazon S3](#) > [Buckets](#) > Create bucket

## Create bucket Info

Buckets are containers for data stored in S3.

### General configuration

AWS Region  
Asia Pacific (Singapore) ap-southeast-1

Bucket name Info

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - *optional*  
Only the bucket settings in the following configuration are copied.

Format: s3://bucket/prefix

Give name and create Bucket

General purpose buckets (3) Info All AWS Regions

Copy ARN Empty Create bucket

< 1 >

	Name ▲	AWS Region ▼	IAM Access Analyzer	Creation date ▼
<input type="radio"/>	<a href="#">cf-templates-1k1y4li1me7pt-us-east-1</a>	US East (N. Virginia) us-east-1	<a href="#">View analyzer for us-east-1</a>	April 9, 2024, 22:59:11 (UTC+05:30)
<input type="radio"/>	<a href="#">elasticbeanstalk-us-east-1-730335274013</a>	US East (N. Virginia) us-east-1	<a href="#">View analyzer for us-east-1</a>	March 28, 2024, 07:55:12 (UTC+05:30)
<input type="radio"/>	<a href="#">test-bucket-redshift-ruchi</a>	Asia Pacific (Singapore) ap-southeast-1	<a href="#">View analyzer for ap-southeast-1</a>	April 21, 2024, 00:21:45 (UTC+05:30)

Then Upload the Csv file

< 1

▲	Type	▼	Last modified	▼	Size	▼	Storage class
---	------	---	---------------	---	------	---	---------------

No objects

You don't have any objects in this bucket.

📁 Upload

Add and upload the file here

Upload

Info

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](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (0)

Remove

Add files

Add folder

All files and folders in this table will be uploaded.

🔍 Find by name

< 1 >

☐	Name	▼	Folder	▼	Type
No files or folders					
You have not chosen any files or folders to upload.					

Files and folders (4 Total, 1.7 MB)

Remove

Add files

Add folder

All files and folders in this table will be uploaded.

🔍 Find by name

< 1 >

☐	Name	▼	Folder	▼	Type
☐	employee_data.csv		-		text/csv
☐	employee_engagement_survey_dat...		-		text/csv
☐	recruitment_data.csv		-		text/csv
☐	training_and_development_data.csv		-		text/csv

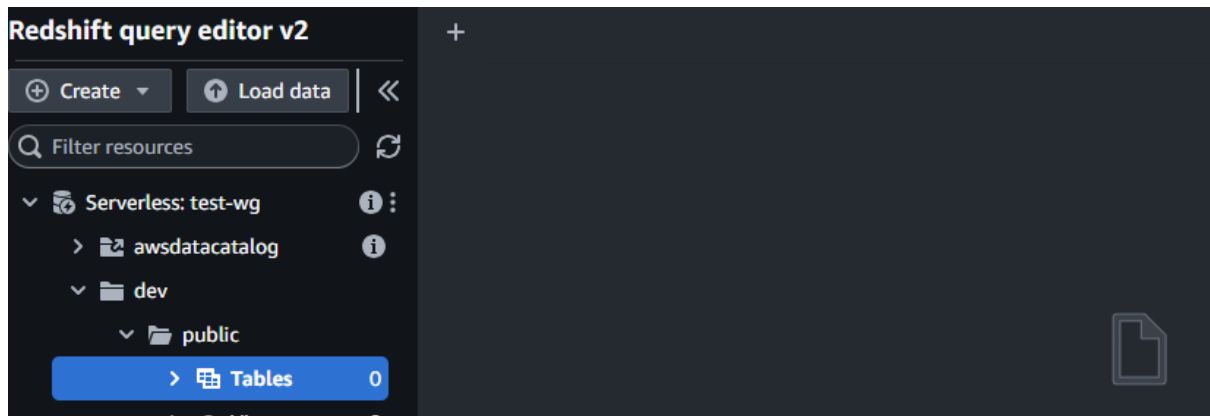


**Files and folders** (4 Total, 1.7 MB)

Find by name

Name	Folder	Type	Size	Status	Error
employee_d...	-	text/csv	761.7 KB	✓ Succeeded	-
employee_e...	-	text/csv	67.5 KB	✓ Succeeded	-
recruitment...	-	text/csv	680.5 KB	✓ Succeeded	-
training_an...	-	text/csv	262.1 KB	✓ Succeeded	-

Now go to Redshift query editor and Load the data



Select Load from S3 bucket, then Browse and select the Csv files

Load data

Data source

Load from S3 bucket

Load from local file

S3 URI

s3://bucket/prefix/object

Browse S3

S3 file location

Manifest file

File format

CSV

File options

No compression

Delimiter character

,

Specifies the single ASCII character that is used to separate fields in the input file, such as a pipe character (|), a comma (,), or a tab (\t).

Ignore header rows

1

Treats the specified number\_rows as a file header and doesn't load them. Use this option to skip file headers in all files in a parallel load.

Select the File key and give Next

S3 buckets > test-bucket-redshift-ruchi

Objects

Key	Last modified	Size
<input type="checkbox"/> employee_data.csv	April 21, 2024, 00:29:53 (UTC+05:30)	761.72 KB
<input type="checkbox"/> employee_engagement_survey	April 21, 2024, 00:29:54 (UTC+05:30)	67.47 KB
<input type="checkbox"/> recruitment_data.csv	April 21, 2024, 00:29:56 (UTC+05:30)	680.54 KB
<input checked="" type="checkbox"/> sampledatainsurance.csv	April 21, 2024, 00:58:03 (UTC+05:30)	32.58 KB
<input type="checkbox"/> training_and_development_dat	April 21, 2024, 00:29:57 (UTC+05:30)	262.09 KB

Choose Load new table and select IAM role

The screenshot shows the 'Load new table' configuration screen in the AWS Redshift console. At the top, there are two radio buttons: 'Load existing table' (unselected) and 'Load new table' (selected). Below these are four dropdown menus: 'Cluster or workgroup' (Serverless: test...), 'Database' (dev), 'Schema' (public), and 'Table' (Insurance). The 'IAM role' field is highlighted with a blue box and contains the value 'arn:aws:iam::730335274013:role/test-role-redshift-ruchi'. Below the IAM role field, there are tabs for 'Columns' and 'Table details'. At the bottom, there is a button labeled 'Add column' and a section for 'Column options'.

If IAM Role not present Create a new Role for Redshift

The screenshot shows the 'Roles' page in the AWS IAM console. The page title is 'Roles (14) Info'. Below the title, there is a search bar and a table of roles. The table has columns for 'Role name', 'Trusted entities', and 'Last activity'. The first two roles listed are 'AWSBackupDefaultServiceRole' and 'AWSServiceRoleForAmazonElasticFilesystem'. The 'Create role' button is highlighted in orange.

Role name	Trusted entities	Last activity
<a href="#">AWSBackupDefaultServiceRole</a>	AWS Service: backup	9 days ago
<a href="#">AWSServiceRoleForAmazonElasticFilesystem</a>	AWS Service: elasticfilesystem (Servi	27 days ago

Choose the Service as Redshift and Click Next

### Use case

Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Service or use case

Redshift ▼

Choose a use case for the specified service.

Use case

☒ Redshift - Customizable

Allows Redshift clusters to call AWS services on your behalf.

☐ Redshift

Allows Redshift clusters to call AWS services on your behalf.

☐ Redshift - Scheduler

Allow Redshift Scheduler to call Redshift on your behalf.

Cancel

Next

### Permissions policies (1/921) [Info](#)

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

×





Filter by Type

All types
▼

9 matches

< 1 >

⚙️

<input type="checkbox"/>	Policy name <a href="#">↗</a>	Type	Description
<input type="checkbox"/>	 <a href="#">AmazonDMSRedshiftS3Role</a>	AWS managed	Provides access to manage S3 settings...
<input checked="" type="checkbox"/>	 <a href="#">AmazonS3FullAccess</a>	AWS managed	Provides full access to all buckets via t...
<input type="checkbox"/>	 <a href="#">AmazonS3ObjectLambda...</a>	AWS managed	Provides AWS Lambda functions permi...
<input type="checkbox"/>	 <a href="#">AmazonS3OutpostsFullA...</a>	AWS managed	Provides full access to Amazon S3 on ...

## Name, review, and create

### Role details

#### Role name

Enter a meaningful name to identify this role.

Maximum 64 characters. Use alphanumeric and '+,=, @, -, \_' characters.

#### Description


Add a short explanation for this role.

Maximum 1000 characters. Use letters (A-Z and a-z), numbers (0-9), tabs, new lines, or any of the following characters: \_+=, @-/[[]!#\$%^&\*():"'<>`

## Step 2: Add permissions

[Edit](#)

### Permissions policy summary

Policy name 	Type	Attached as
<a href="#">AmazonRedshiftFullAccess</a>	AWS managed	Permissions policy
<a href="#">AmazonS3FullAccess</a>	AWS managed	Permissions policy
<a href="#">AmazonS3ReadOnlyAccess</a>	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.

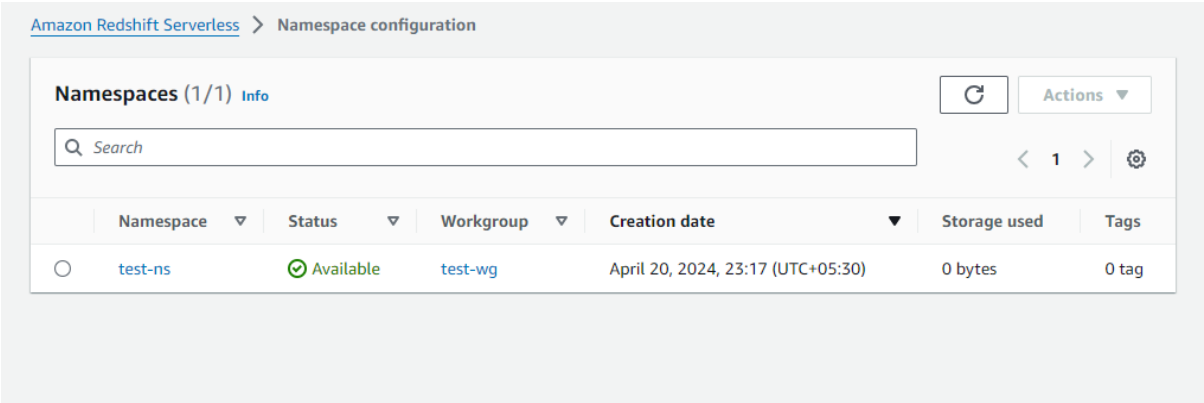
No tags associated with the resource.

[Add new tag](#)

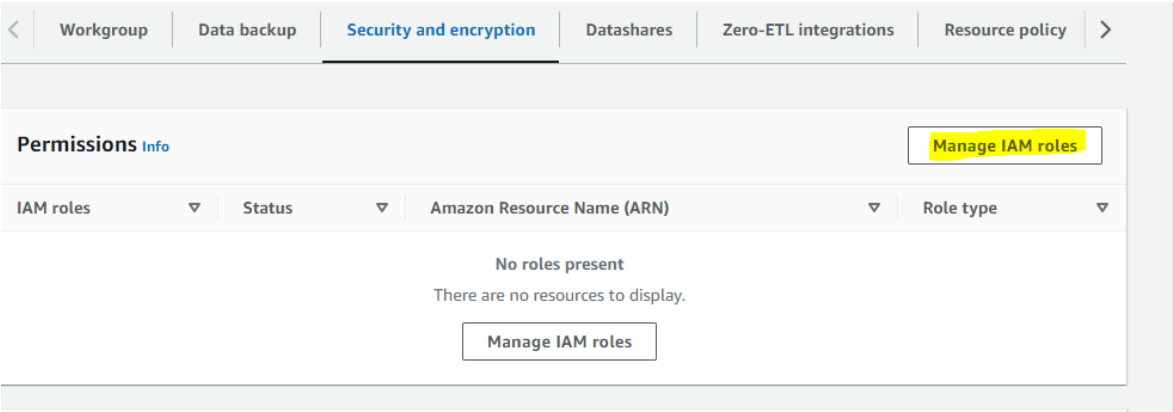
You can add up to 50 more tags.

[Cancel](#)[Previous](#)[Create role](#)

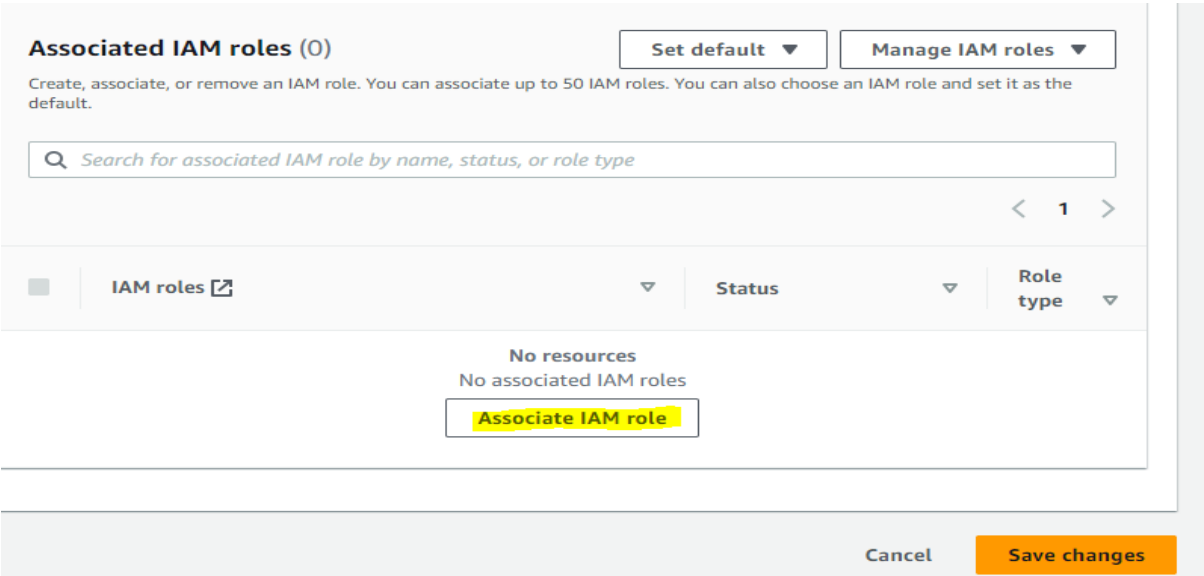
Create the Name configuration under Redshift to Manage the role for Redshift



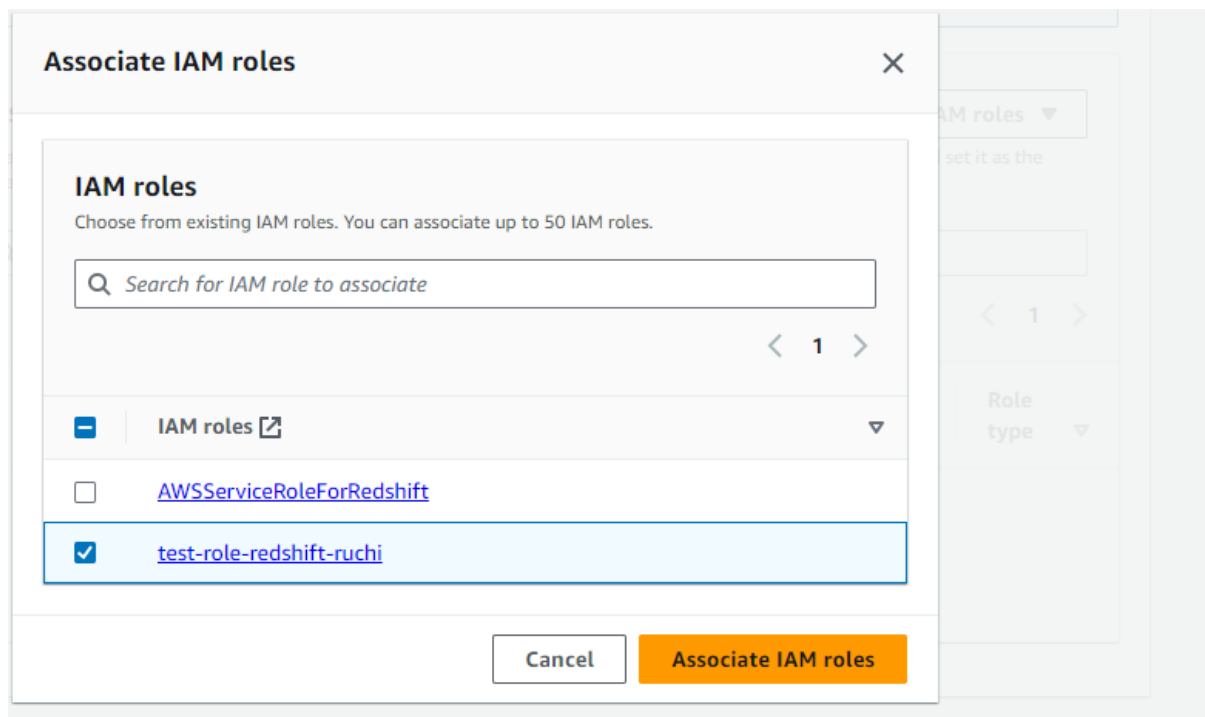
Click on security & encryption tab and click on Manage IAM Role



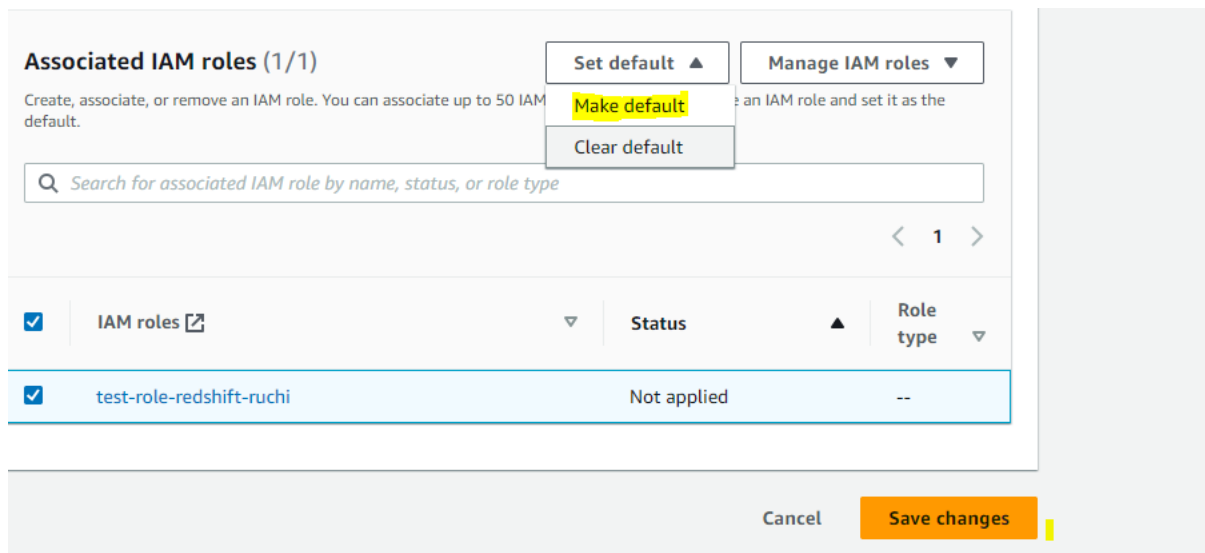
Choose Associate IAM role



Select the created IAM role



Set the default as Make default and save the changes



Associated IAM role is created

<

Workgroup

Data backup

Security and encryption

Datashares

Zero-ETL integrations

Resource policy

>

Permissions

Info

Manage IAM roles

IAM roles	Status	Amazon Resource Name (ARN)	Role type
<a href="#">test-role-redshift-ruchi</a>	<span>in-sync</span>	<code>arn:aws:iam::730335274013:role/test-role-redshift-ruchi</code>	Default

Namespace configuration is Created

Amazon Redshift Serverless

>

Namespace configuration

Namespaces (1/1)

Info

Actions

Search

< 1 >

	Namespace	Status	Workgroup	Creation date	Storage used	Tags
<input type="radio"/>	<a href="#">test-ns</a>	<span>Available</span>	<a href="#">test-wg</a>	April 20, 2024, 23:17 (UTC+05:30)	0 bytes	0 tag



Now IAM Role will reflect then mention the following configuration and create table

☐ Load existing table  
Load data into an existing table

☒ Load new table  
Create table with detected schema

Cluster or workgroup: Serverless: test... Database: dev Schema: public Table: Insurance

IAM role: `arn:aws:iam::730335274013:role/test-role-redshift-ruchi`

**Columns** Table details

Selected 53 file: sampledatainsurance.csv [+ Add column](#)

Column name	Data type	Encoding	
Policy	INTEGER	No selection	
Expiry	VARCHAR	No selection	
Location	VARCHAR	No selection	
State	VARCHAR	No selection	
Region	VARCHAR	No selection	
InsuredValue	VARCHAR	No selection	
Construction	VARCHAR	No selection	
BusinessType	VARCHAR	No selection	

**Column options**

Default value

☐ Custom

☐ Empty string

☐ NULL

☒ No default value

Size:

☐ Primary key

☐ Unique key

☐ Not NULL

☐ Automatically increment

[Back](#) [Restore to defaults](#) [Cancel](#) [Create table](#)

Click on Load Data

Cluster or workgroup: Serverless: test-wg Database: dev Schema: public Table: Insurance

IAM role: arn:aws:iam: Insurance table is created successfully. X

Table summary

Column name	Data type	Encoding
Policy	INTEGER	No selection
Expiry	VARCHAR	No selection
Location	VARCHAR	No selection
State	VARCHAR	No selection
Region	VARCHAR	No selection
InsuredValue	VARCHAR	No selection
Construction	VARCHAR	No selection
BusinessType	VARCHAR	No selection
Earthquake	VARCHAR	No selection
Flood	VARCHAR	No selection

Back Cancel Load data

Query is reflecting

Redshift query editor v2

Filter resources

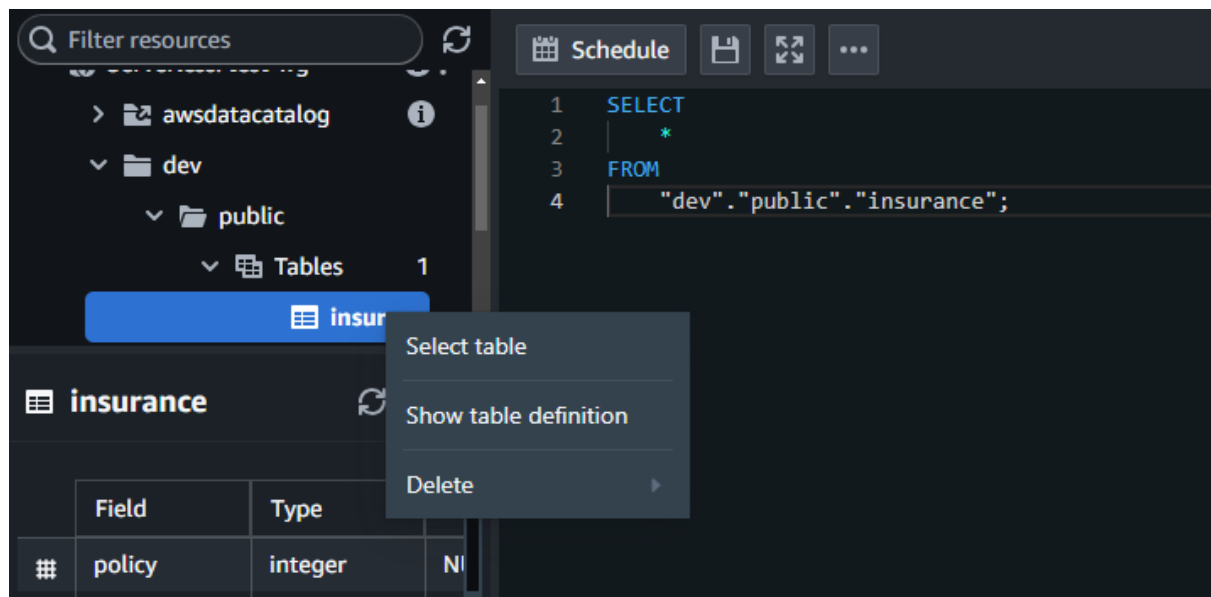
Serverless: test-wg

- awsdatacatalog
- dev
  - public
    - Tables 1
      - insurance
    - Views 0
    - Functions 0
    - Stored proc... 0

1 .274013:role/test-role-redshift-ruchi' FORMAT AS CSV DELIMITER ',' QUOTE '' IGNOREHEADER 1 REGION AS 'ap-south

```
COPY dev.public.Insurance FROM 's3://test-bucket-redshift-ruchi/sampledatabinsurance.csv' IAM_ROLE 'arn:aws:iam::730335274013:role/test-role-redshift-ruchi' FORMAT AS CSV DELIMITER ',' QUOTE '' IGNOREHEADER 1 REGION AS 'ap-southeast-1'
```

Right click the insurance, select Table and Then Run the query



The data will be loaded after running the query

