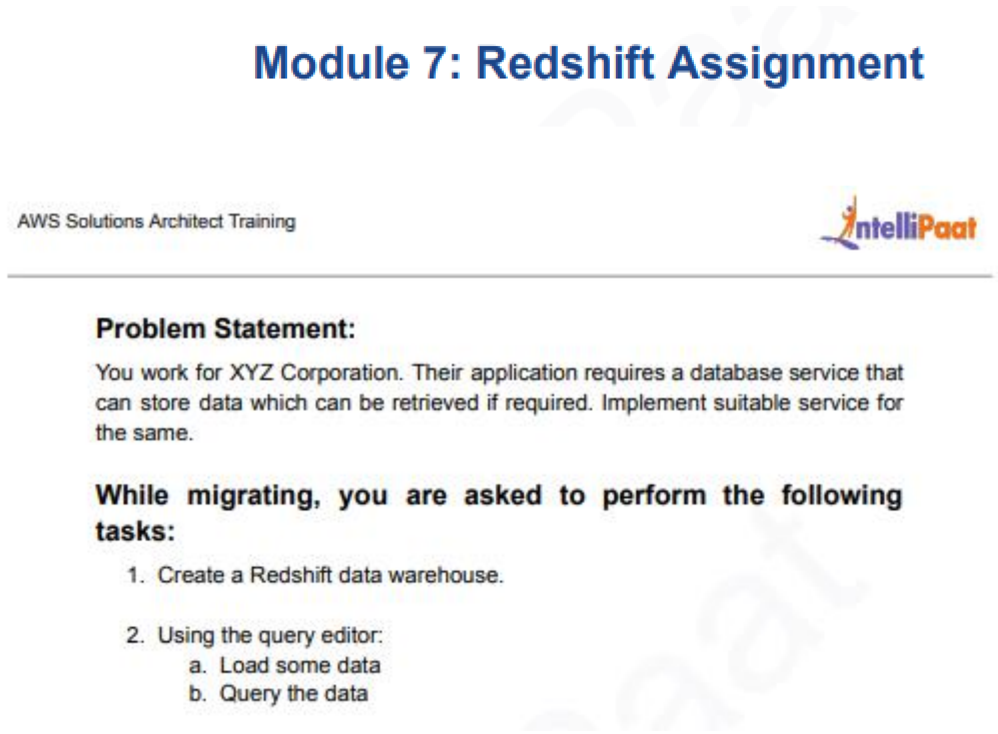


Module 7: Redshift Assignment

AWS Solutions Architect Training

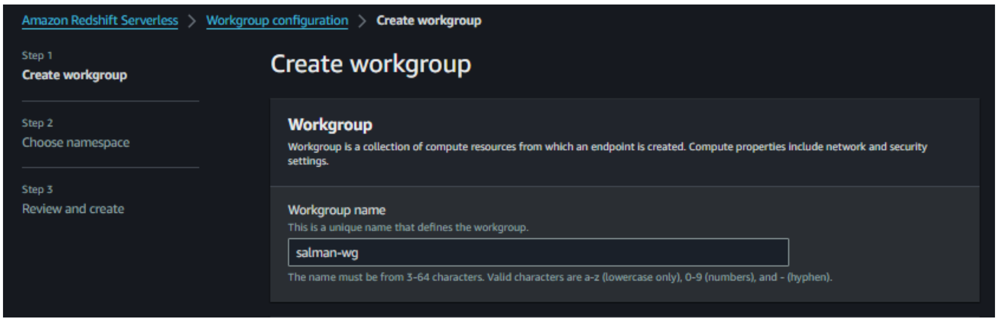
Problem Statement:

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

Solution: 1) Create workgroup and namespace Name of the workgroup: salman-wg



Amazon Redshift Serverless > Workgroup configuration > Create workgroup
Step 1
Create workgroup
Step 2
Choose namespace
Step 3
Review and create

Create workgroup

Workgroup

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

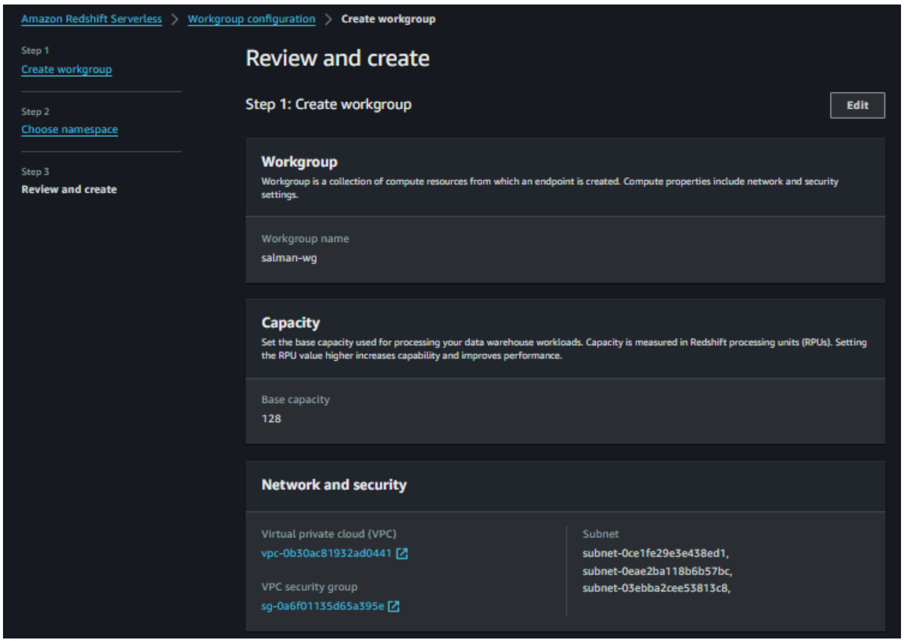
Workgroup name

This is a unique name that defines the workgroup.

salman-wg

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

Namespace: salman-ns



Amazon Redshift Serverless > Workgroup configuration > Create workgroup
Step 1
Create workgroup
Step 2
Choose namespace
Step 3
Review and create
Edit

Review and create

Step 1: 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
salman-wg

Capacity

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
128

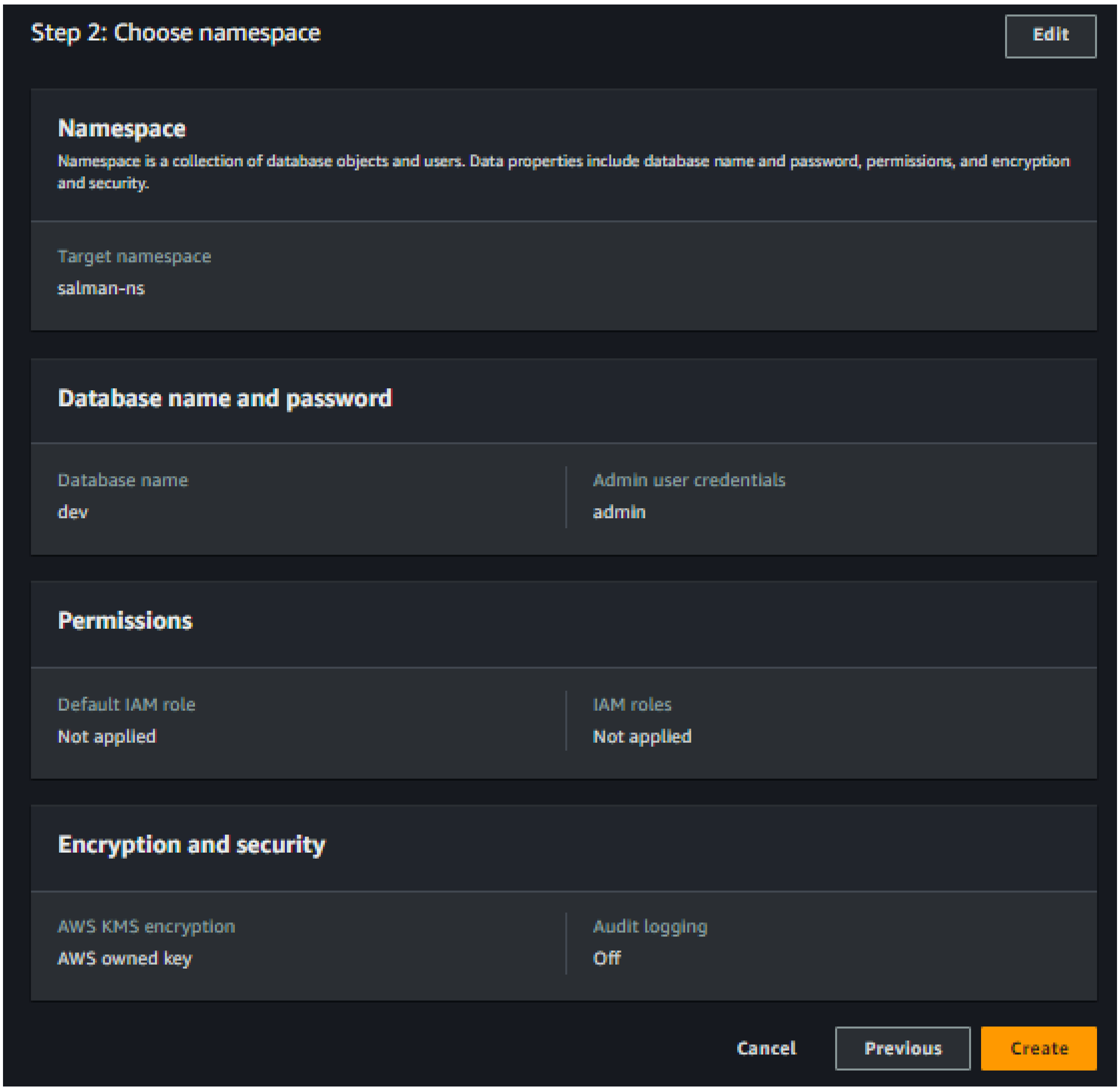
Network and security

Virtual private cloud (VPC)
vpc-0b30ac81932ad0441

VPC security group
sg-0a6f01135d65a395e

Subnet
subnet-0ce1fe29e3e438ed1,
subnet-0eae2ba118b6b57bc,
subnet-03ebba2cee53813c8,

2) Review the workgroup and namespace and create it.



Step 2: Choose namespace

Namespace

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

Target namespace: salman-ns

Database name and password

Database name: dev

Admin user credentials: admin

Permissions

Default IAM role: Not applied

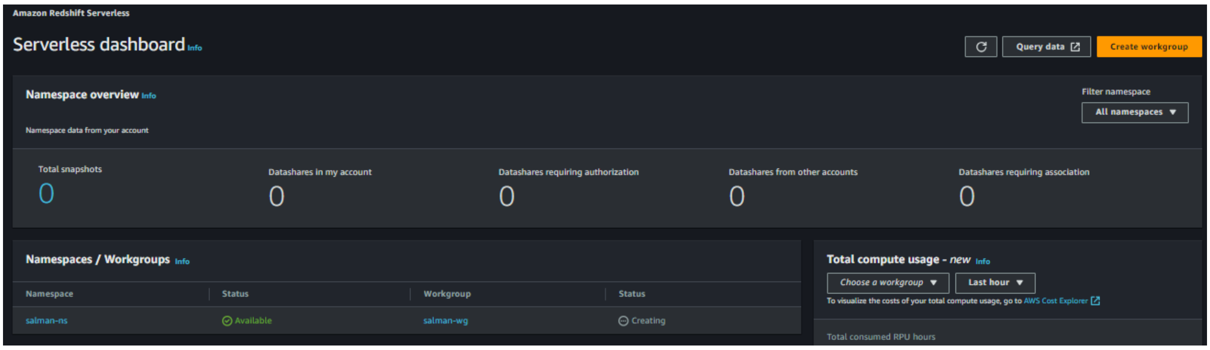
IAM roles: Not applied

Encryption and security

AWS KMS encryption: AWS owned key

Audit logging: Off

3) Workgroup and namespace created successfully.



Namespace overview

Namespace data from your account

Filter namespace: All namespaces

Total snapshots	Datashares in my account	Datashares requiring authorization	Datashares from other accounts	Datashares requiring association
0	0	0	0	0

Namespaces / Workgroups

Namespace	Status	Workgroup	Status
salman-ns	Available	salman-wg	Creating

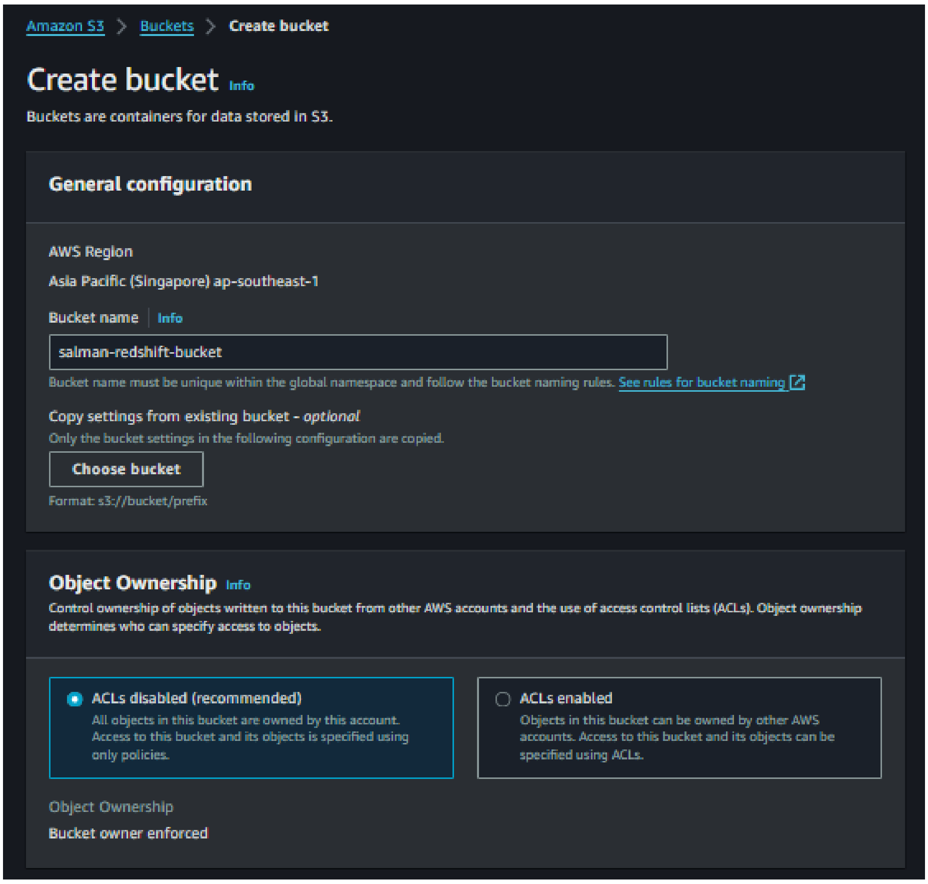
Total compute usage - *new*

To visualize the costs of your total compute usage, go to AWS Cost Explorer

Total consumed RPU hours

4) Now we will create the s3 bucket and other options we will choose as default and upload the csv file in our bucket.

Name: salman-redshift-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

salman-redshift-bucket

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

Object Ownership Info

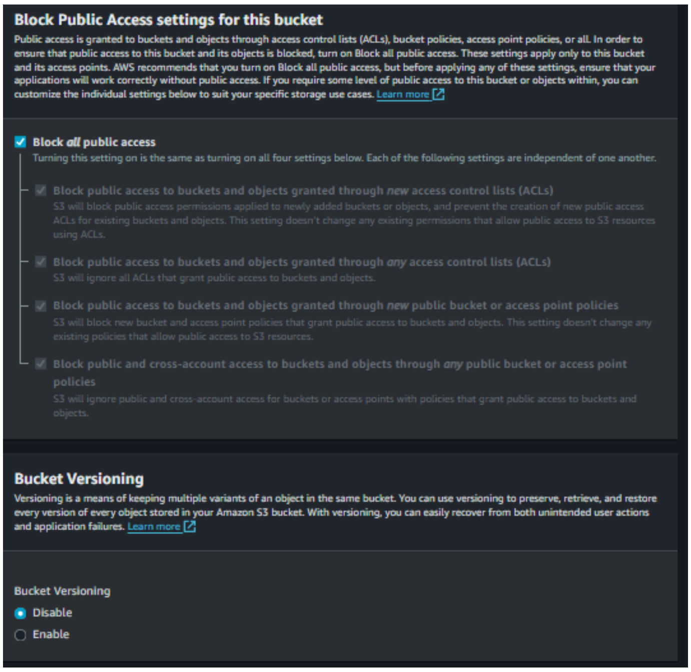
Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

● ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

○ ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced



Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

☑ Block *all* public access
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

- ☑ Block public access to buckets and objects granted through *new* access control lists (ACLs)
  S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
- ☑ Block public access to buckets and objects granted through *any* access control lists (ACLs)
  S3 will ignore all ACLs that grant public access to buckets and objects.
- ☑ Block public access to buckets and objects granted through *new* public bucket or access point policies
  S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
- ☑ Block public and cross-account access to buckets and objects through *any* public bucket or access point policies
  S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

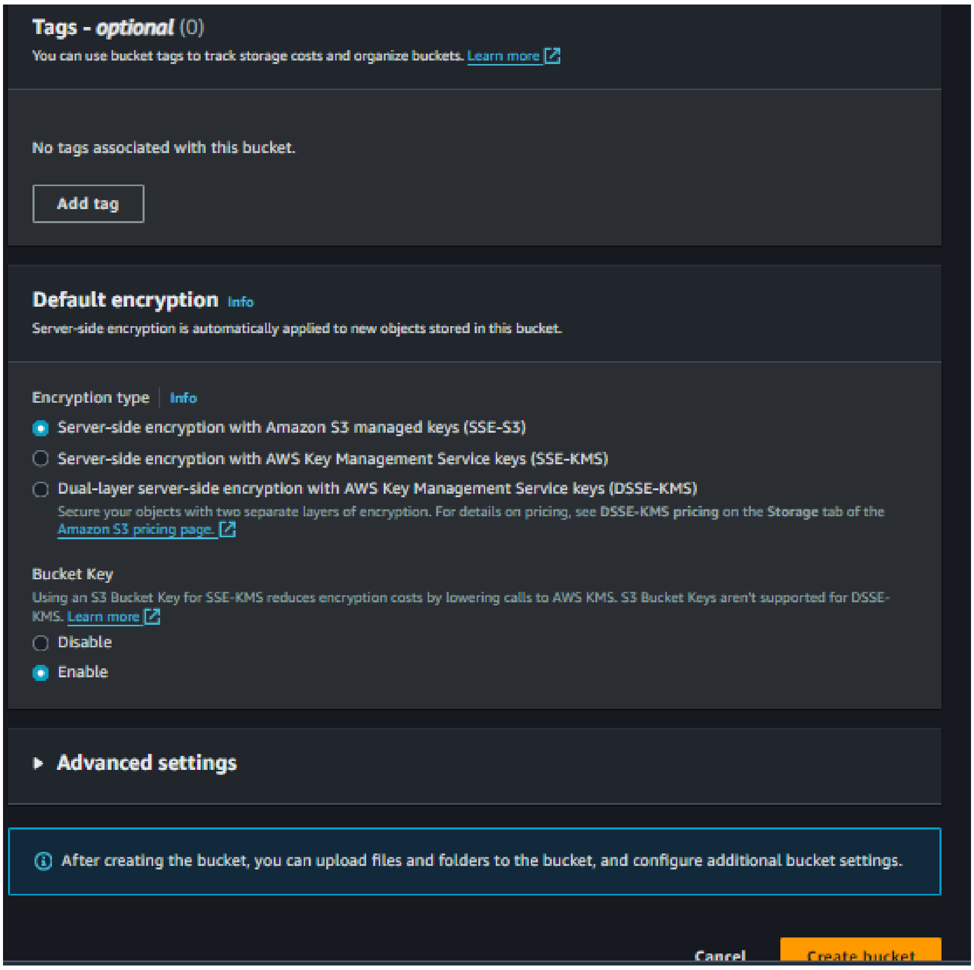
Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. Learn more

Bucket Versioning

● Disable

○ Enable



Tags - *optional* (0)

You can use bucket tags to track storage costs and organize buckets. Learn more

No tags associated with this bucket.

Default encryption Info

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type | Info

● Server-side encryption with Amazon S3 managed keys (SSE-S3)

○ Server-side encryption with AWS Key Management Service keys (SSE-KMS)

○ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)
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.

Bucket Key

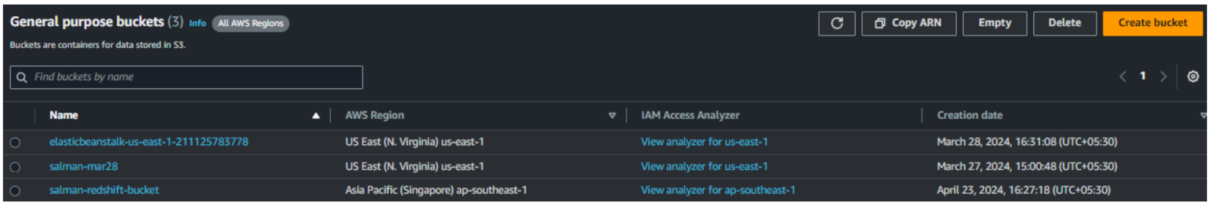
Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. Learn more

○ Disable

● Enable

▸ Advanced settings

ⓘ After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

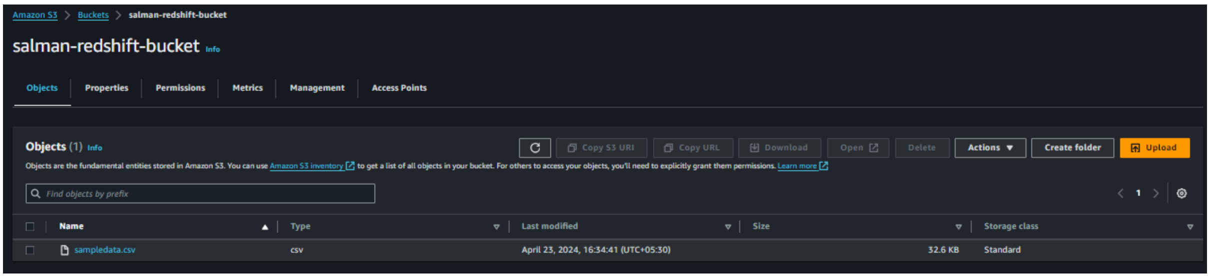


General purpose buckets (3) Info All AWS Regions

Buckets are containers for data stored in S3.

	Name	AWS Region	IAM Access Analyzer	Creation date
○	elasticbeanstalk-us-east-1-211125783778	US East (N. Virginia) us-east-1	View analyzer for us-east-1	March 28, 2024, 16:31:08 (UTC+05:30)
○	salman-mar28	US East (N. Virginia) us-east-1	View analyzer for us-east-1	March 27, 2024, 15:00:48 (UTC+05:30)
○	salman-redshift-bucket	Asia Pacific (Singapore) ap-southeast-1	View analyzer for ap-southeast-1	April 23, 2024, 16:27:18 (UTC+05:30)

5) Upload csv file in our bucket.



Amazon S3 > Buckets > salman-redshift-bucket

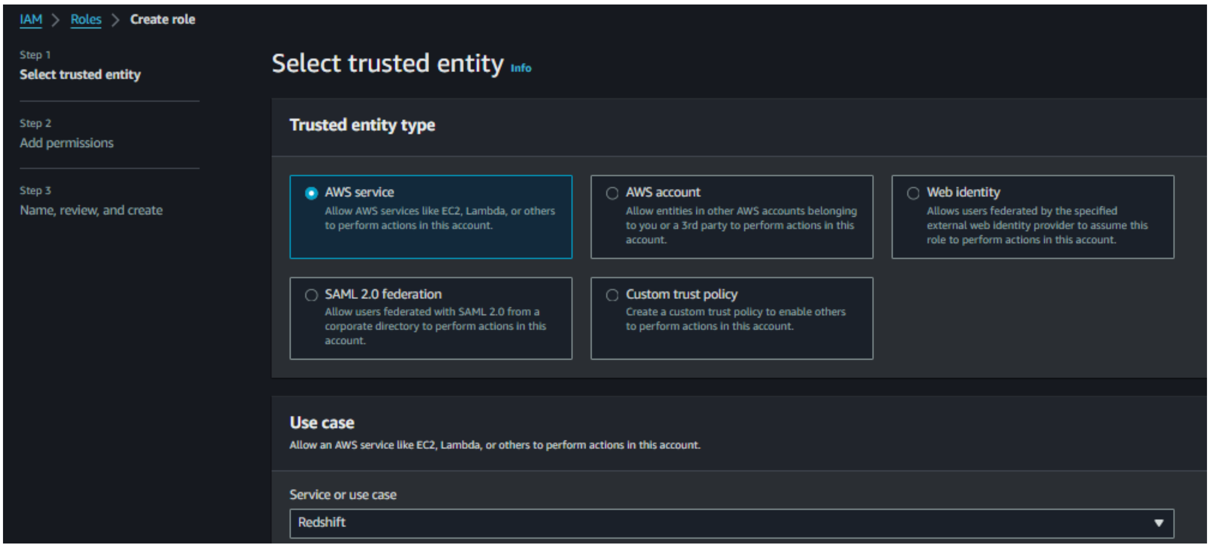
salman-redshift-bucket Info

Objects (1) Info

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. Learn more

□	Name	Type	Last modified	Size	Storage class
□	sampledata.csv	csv	April 23, 2024, 16:34:41 (UTC+05:30)	32.6 KB	Standard

6) Create IAM role and in service or use case we will choose Redshift



IAM > Roles > Create role

Step 1
Select trusted entity

Step 2
Add permissions

Step 3
Name, review, and create

Select trusted entity Info

Trusted entity type

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

○ AWS account
Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.

○ Web identity
Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.

○ SAML 2.0 federation
Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.

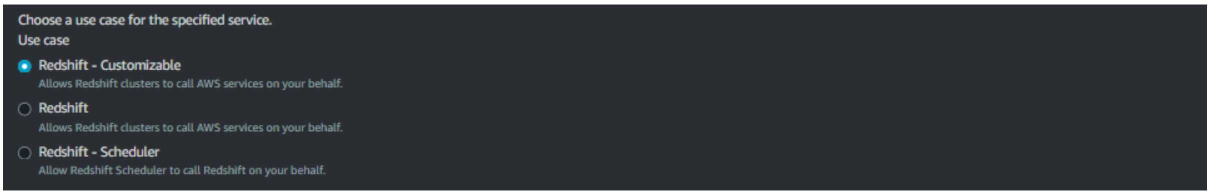
○ Custom trust policy
Create a custom trust policy to enable others to perform actions in this account.

Use case

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

Service or use case
Redshift ▼

7) & Redshift-Customizable.



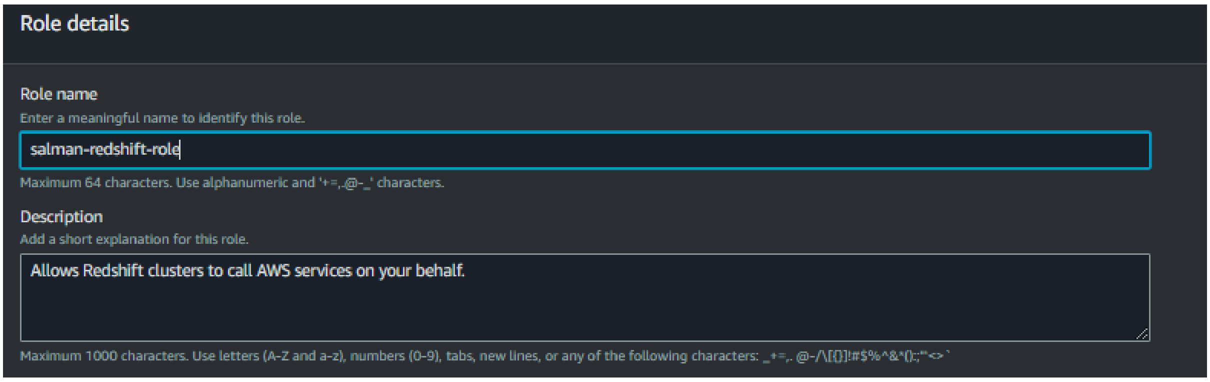
Choose a use case for the specified service.
Use case

● 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.

8) Give a role as salman-redshift-role.

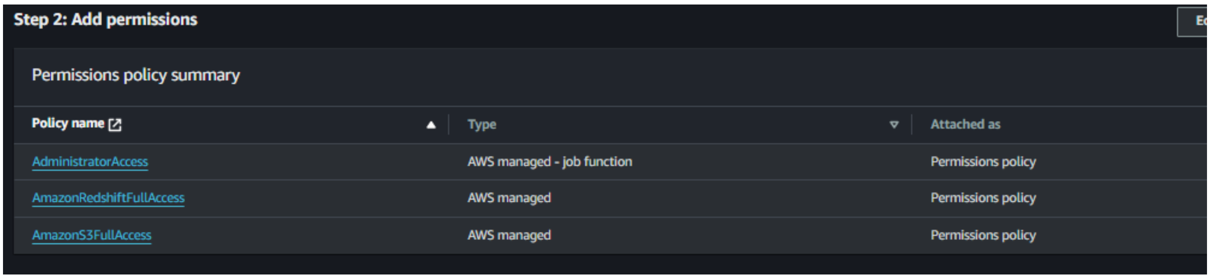


Role details

Role name
Enter a meaningful name to identify this role.
salman-redshift-role
Maximum 64 characters. Use alphanumeric and '+=,.@-_' characters.

Description
Add a short explanation for this role.
Allows Redshift clusters to call AWS services on your behalf.
Maximum 1000 characters. Use letters (A-Z and a-z), numbers (0-9), tabs, new lines, or any of the following characters: _+=,. @-/\[{}]!#$%^&*();:'"<>`

permissions

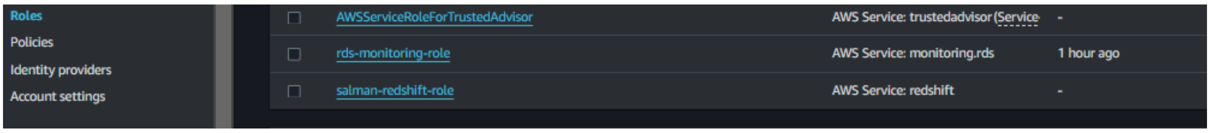


Step 2: Add permissions

Permissions policy summary

Policy name	Type	Attached as
AdministratorAccess	AWS managed - job function	Permissions policy
AmazonRedshiftFullAccess	AWS managed	Permissions policy
AmazonS3FullAccess	AWS managed	Permissions policy

Successfully created

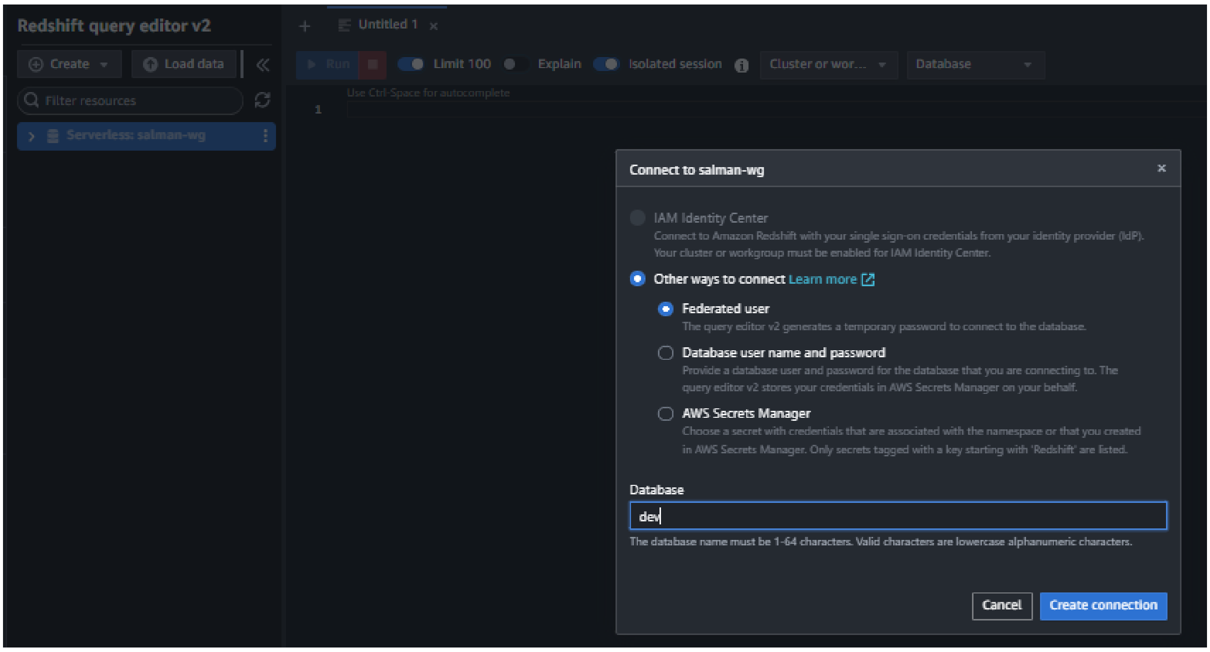


Roles
Policies
Identity providers
Account settings

□	AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service	-
□	rds-monitoring-role	AWS Service: monitoring.rds	1 hour ago
□	salman-redshift-role	AWS Service: redshift	-

9) We will go to amazon redshift and click on query editor.

10) Click on Serverless:salman-wg and connect to salman-wg and choose Federated user and click on create connection.



Connect to salman-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

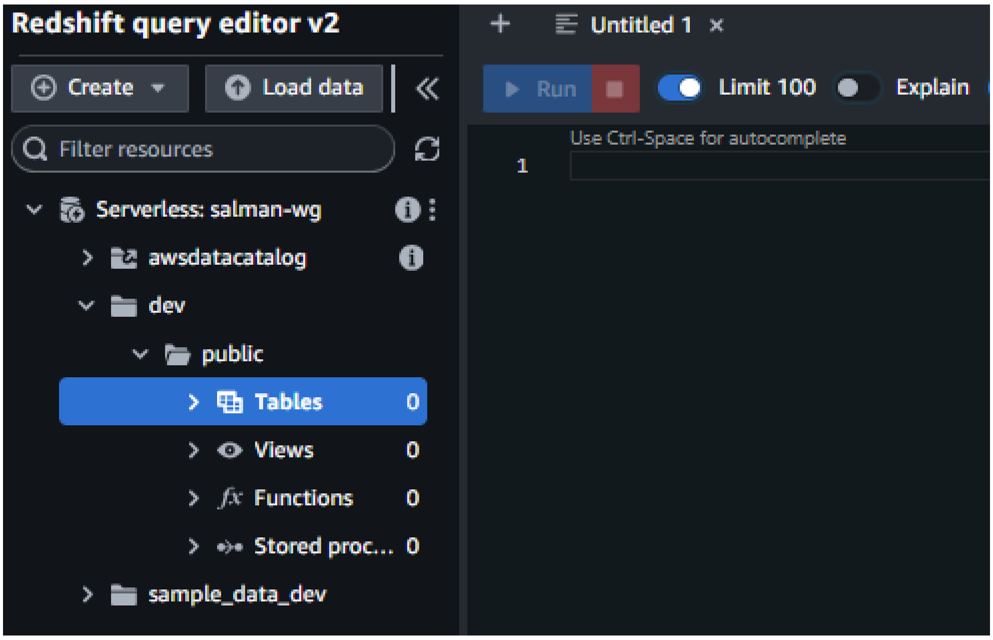
● Federated user
The query editor v2 generates a temporary password to connect to the database.

○ 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.

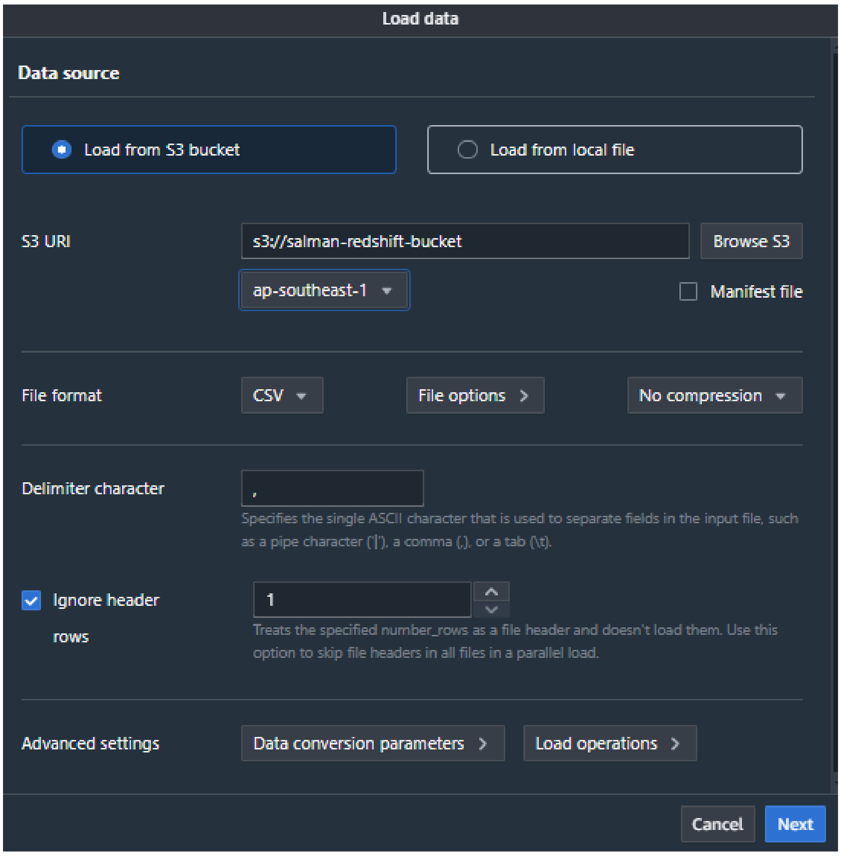
Database: dev
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

11) Click on Dev and public and right now we have 0 tables in public. Click on load data.



Serverless: salman-wg
- awsdatacatalog
- dev
  - public
    - Tables 0
    - Views 0
    - Functions 0
    - Stored proc... 0
- sample_data_dev

12. We will load data from our s3 bucket. Choose load from s3 bucket, region-us-east-1.



Load data

Data source

● Load from S3 bucket ○ Load from local file

S3 URI: s3://salman-redshift-bucket

ap-southeast-1 □ Manifest file

File format: CSV | File options | No compression

Delimiter character: ,

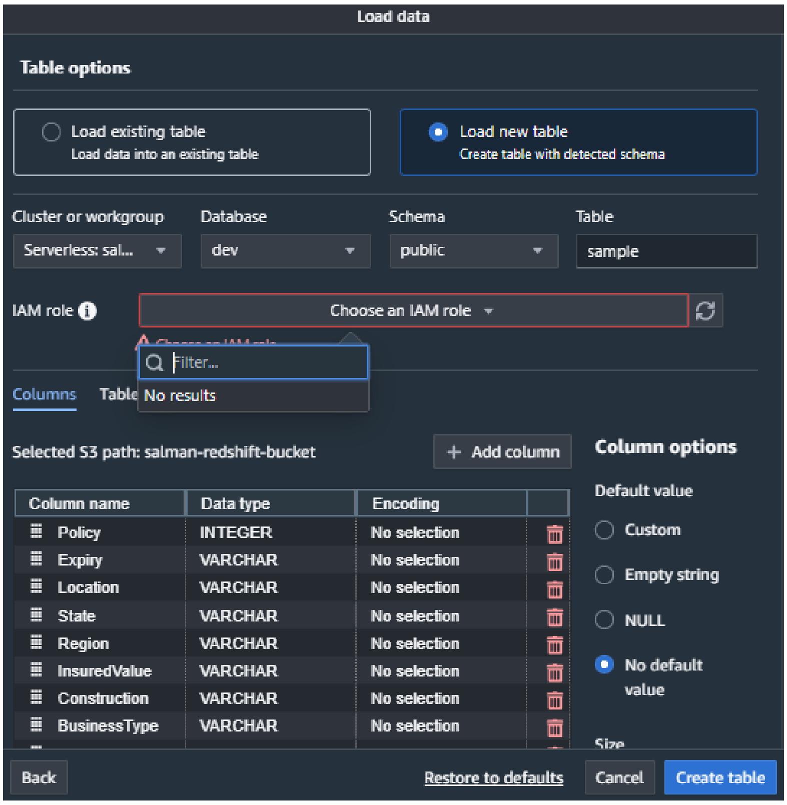
Specifies the single ASCII character that is used to separate fields in the input file, such as a pipe character (|), a comma (,), or a tab (\t).

☑ 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.

Advanced settings: Data conversion parameters | Load operations

13) Next and click on load new table. In schema choose public and table name is "sample" and choose IAM Role which we created.



Load data

Table options

○ Load existing table
Load data into an existing table

● Load new table
Create table with detected schema

Cluster or workgroup: Serverless: sal...

Database: dev

Schema: public

Table: sample

IAM role: Choose an IAM role

Filter...

No results

Selected S3 path: salman-redshift-bucket

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

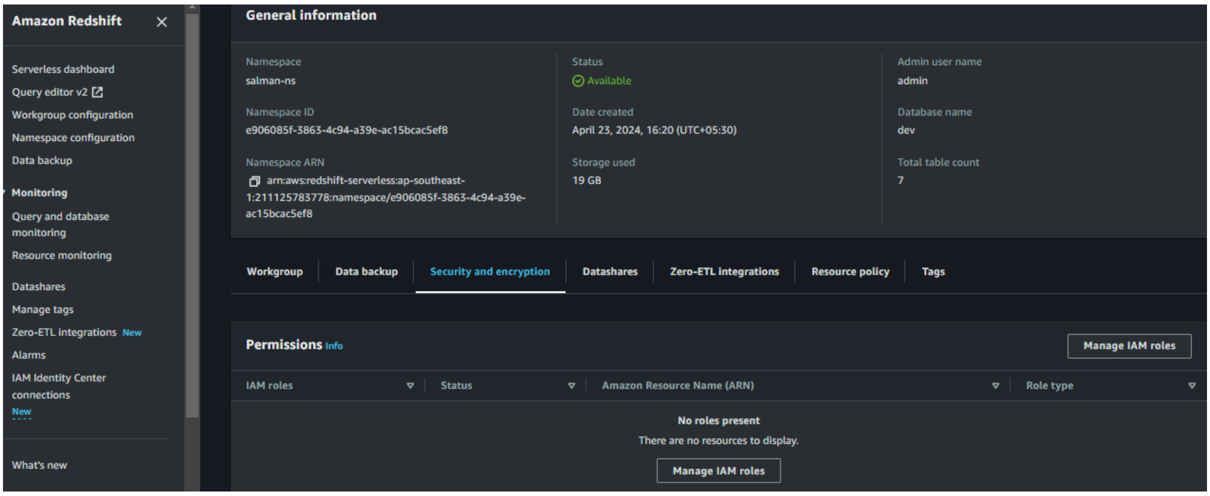
○ NULL

● No default value

Size

Note: Created IAM Role still not available so we need to associate our IAM Role in namespace configuration.

14) Go to namespace configuration
15) Click on security & encryption tab and click on Manage IAM Role



General information

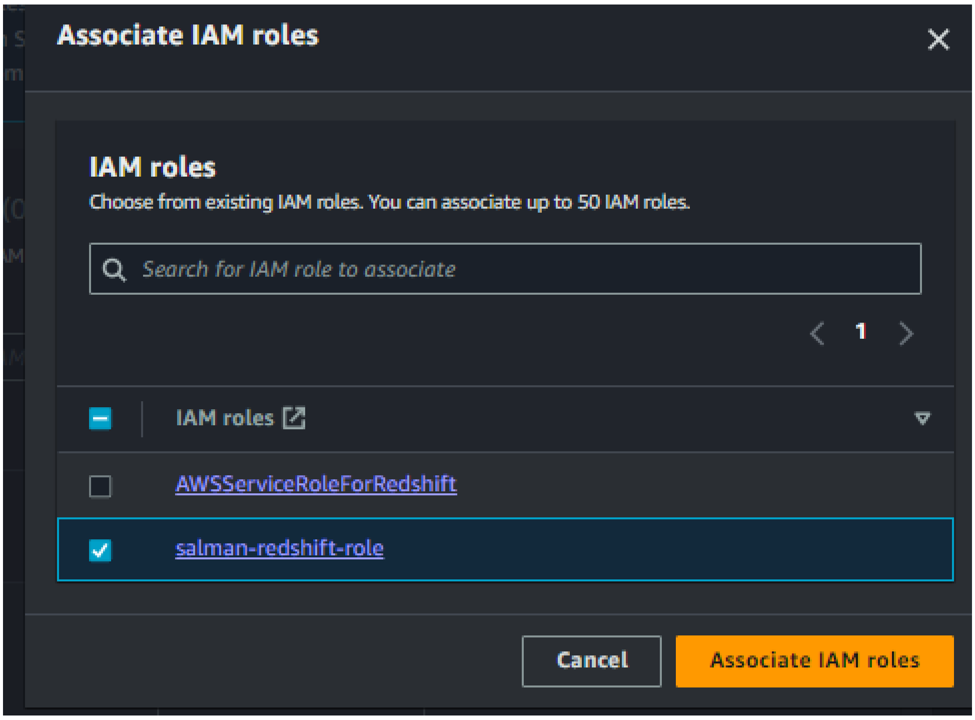
Namespace: salman-ns
Status: Available
Admin user name: admin
Namespace ID: e906085f-3863-4c94-a39e-ac15bcac5ef8
Date created: April 23, 2024, 16:20 (UTC+05:30)
Database name: dev
Namespace ARN: arn:aws:redshift-serverless:ap-southeast-1:211125783778:namespace/e906085f-3863-4c94-a39e-ac15bcac5ef8
Storage used: 19 GB
Total table count: 7

Permissions Info

IAM roles	Status	Amazon Resource Name (ARN)	Role type

No roles present
There are no resources to display.

16) Click on associate IAM role and make it default.
17) Choose IAM Role which we created and click on associate IAM Role.

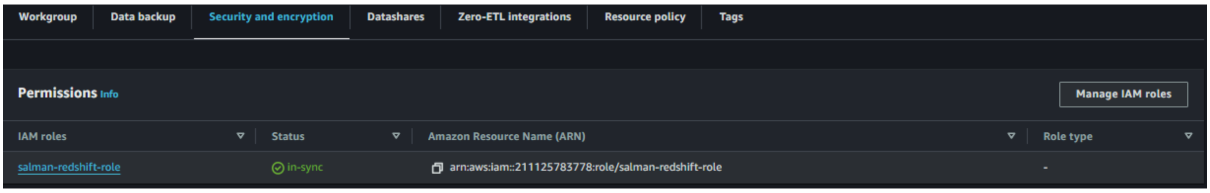


Associate IAM roles

IAM roles
Choose from existing IAM roles. You can associate up to 50 IAM roles.

Search for IAM role to associate

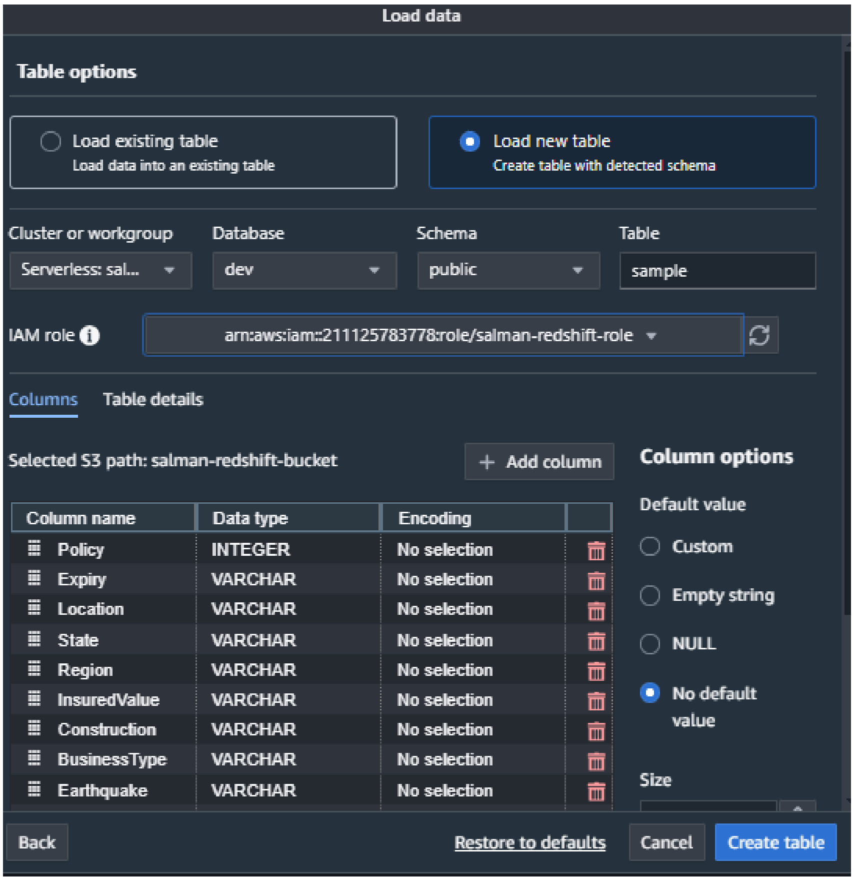
	IAM roles
□	AWSServiceRoleForRedshift
☑	salman-redshift-role



Permissions Info

IAM roles	Status	Amazon Resource Name (ARN)	Role type
salman-redshift-role	in-sync	arn:aws:iam::211125783778:role/salman-redshift-role	-

18) Now our IAM role is reflecting when loading the data.



Load data

Table options

○ Load existing table
Load data into an existing table

● Load new table
Create table with detected schema

Cluster or workgroup: Serverless: sal...
Database: dev
Schema: public
Table: sample

IAM role: arn:aws:iam::211125783778:role/salman-redshift-role

Selected S3 path: salman-redshift-bucket

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

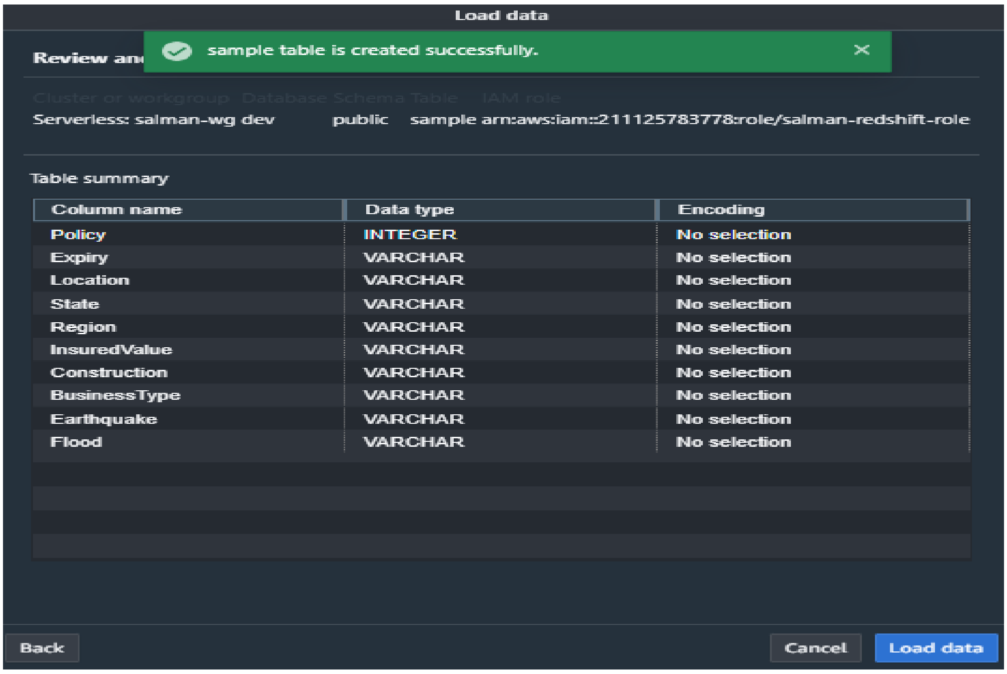
Column options

Default value
○ Custom
○ Empty string
○ NULL
● No default value

Size

19) We will update the IAM role and click on create table and our insurance table created successfully.

20) Review and load Data and our data is now loaded successfully.



Load data

Review an

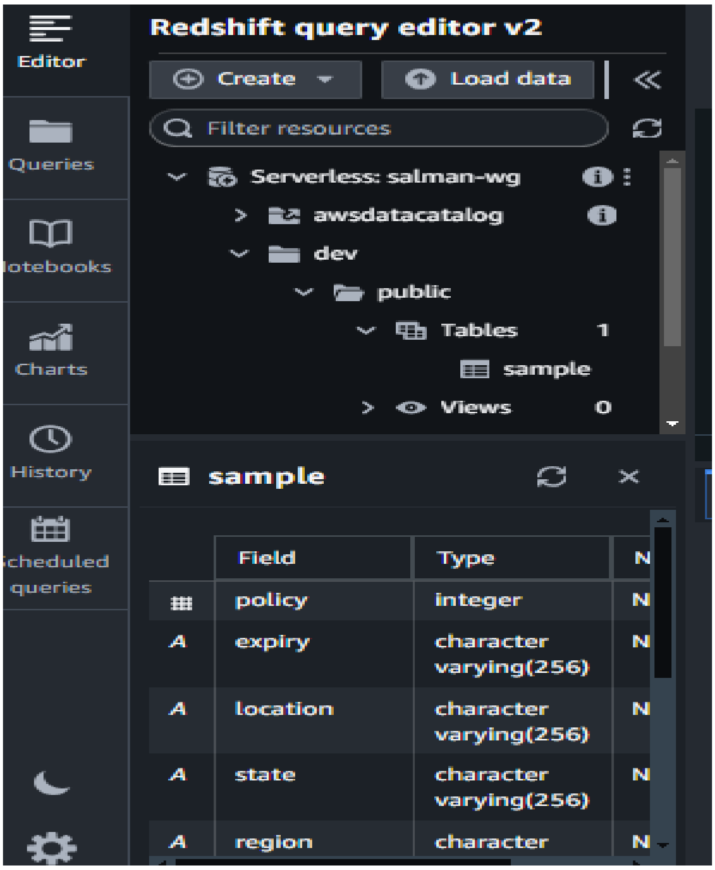
sample table is created successfully.

Cluster or workgroup	Database	Schema	Table	IAM role
Serverless: salman-wg	dev	public	sample	arn:aws:iam::211125783778:role/salman-redshift-role

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

21) Now our table is showing under public.

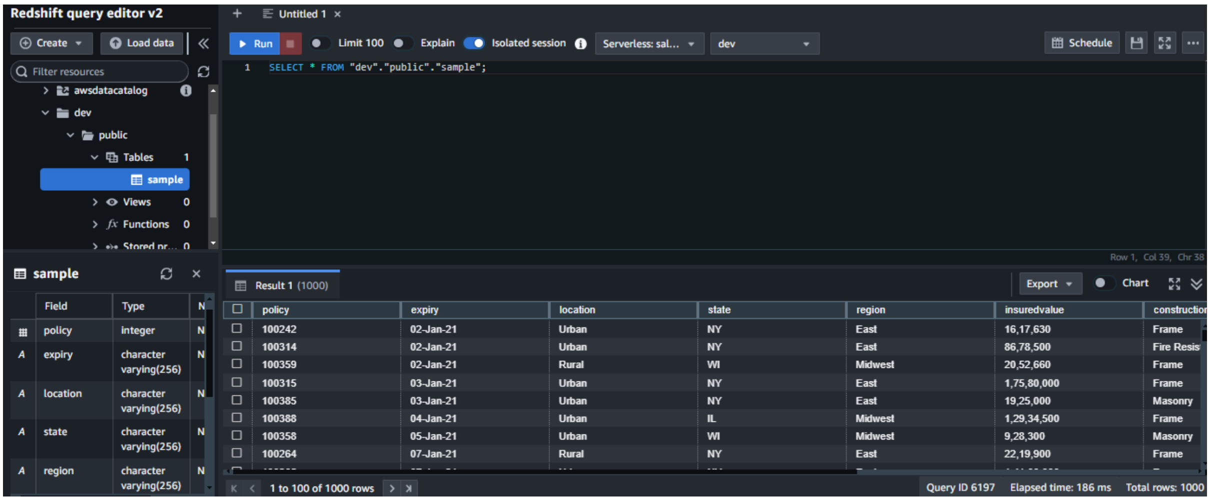


Field	Type	N
policy	integer	N
expiry	character varying(256)	N
location	character varying(256)	N
state	character varying(256)	N
region	character	N

22) We will right click on the table and click on select table.

23) We will uncheck the limit and click on Run.

24) It is now showing all our 500 rows data.



```
SELECT * FROM "dev"."public"."sample";
```

Field	Type	N
policy	integer	N
expiry	character varying(256)	N
location	character varying(256)	N
state	character varying(256)	N
region	character varying(256)	N

policy	expiry	location	state	region	insuredvalue	constructio
100242	02-Jan-21	Urban	NY	East	16,17,630	Frame
100314	02-Jan-21	Urban	NY	East	86,78,500	Fire Resis
100359	02-Jan-21	Rural	WI	Midwest	20,52,660	Frame
100315	03-Jan-21	Urban	NY	East	1,75,80,000	Frame
100385	03-Jan-21	Urban	NY	East	19,25,000	Masonry
100388	04-Jan-21	Urban	IL	Midwest	1,29,34,500	Frame
100358	05-Jan-21	Urban	WI	Midwest	9,28,300	Masonry
100264	07-Jan-21	Rural	NY	East	22,19,900	Frame

1 to 100 of 1000 rows

Query ID 6197 Elapsed time: 186 ms Total rows: 1000