

Creating a RedShift Cluster

Details of the red shift cluster for the ETL project.

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2458014--srisaish123@gmail.com @ 5868-7793-0415

Amazon Redshift > Clusters > redshift-cluster-etl

redshift-cluster-etl

Actions Edit Add partner integration Query data

General information

Cluster identifier redshift-cluster-etl	Status Available	Node type dc2.large	Endpoint redshift-cluster-etl.cd9qt6fotrij.us-ea...
Cluster namespace ff748294-5c64-451c-bf65-341c2214ac41	Date created June 14, 2023, 23:05 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-etl.cd9...
Cluster configuration Production	Storage used -		ODBC URL Driver={Amazon Redshift (x64)}; Serv...
	Multi-AZ No		

Cluster performance Query monitoring Schedules Maintenance Properties

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2458014--srisaish123@gmail.com @ 5868-7793-0415

Cluster performance Query monitoring Schedules Maintenance Properties

Database configurations Info

Edit admin credentials Rotate encryption keys Edit

Database name dev	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5439	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADA...	AWS KMS key ID -	
Admin user name awsuser			

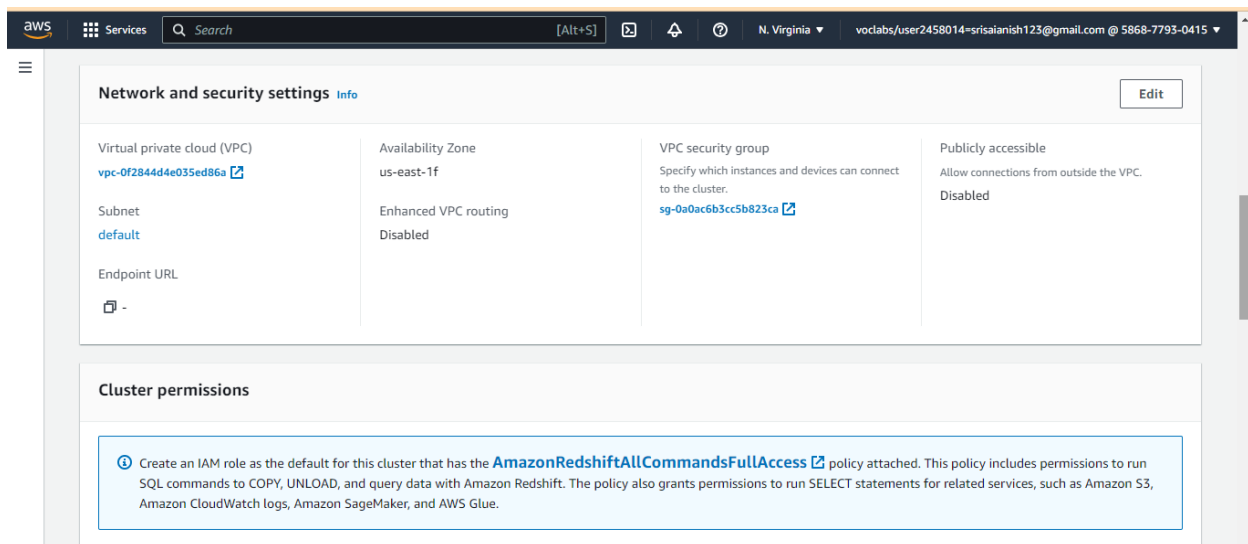
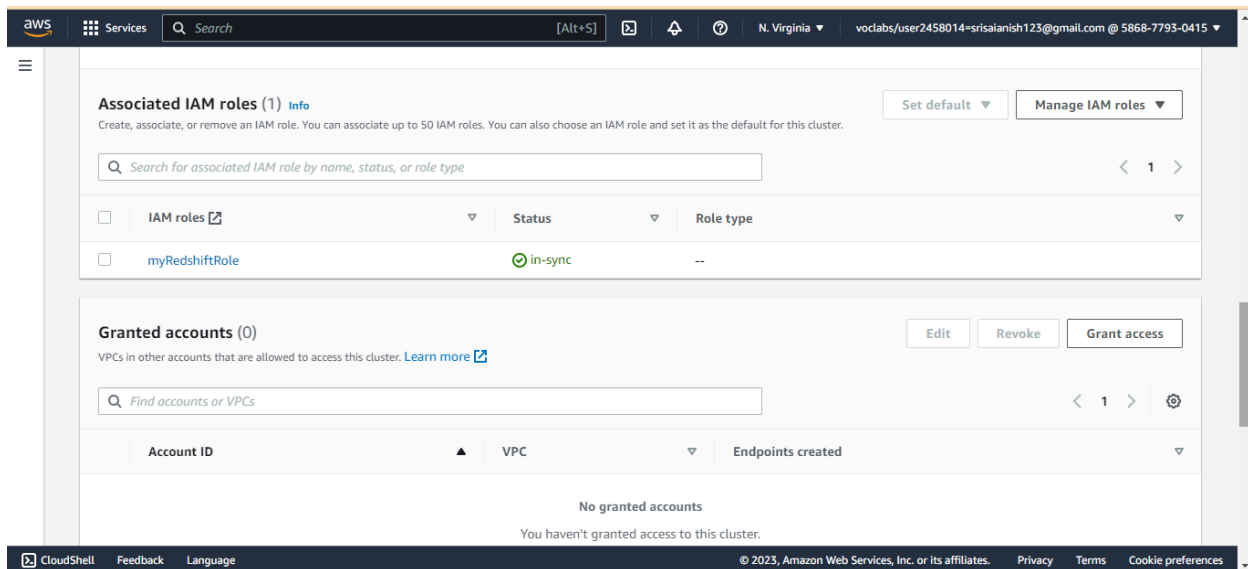
Network and security settings Info

Edit

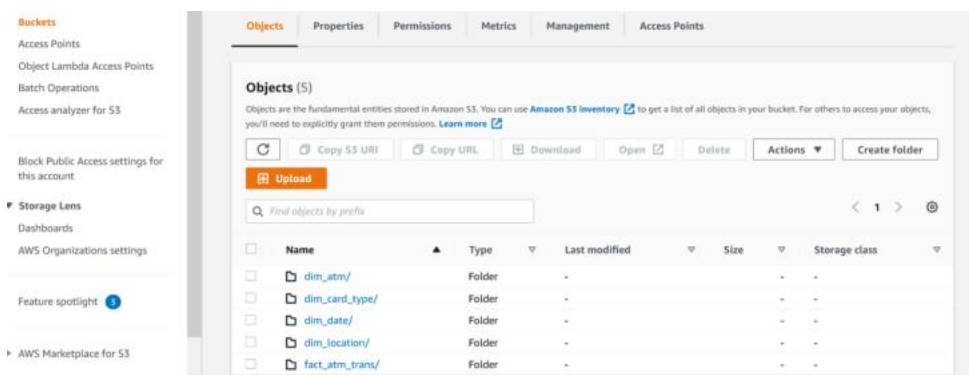
Virtual private cloud (VPC) vpc-0f2844d4e035ed86a	Availability Zone us-east-1f	VPC security group Specify which instances and devices can connect to the cluster. sg-0b9d37c5f832...	Publicly accessible Allow connections from outside the VPC. Disabled
--	---------------------------------	---	--

CloudShell Feedback Language

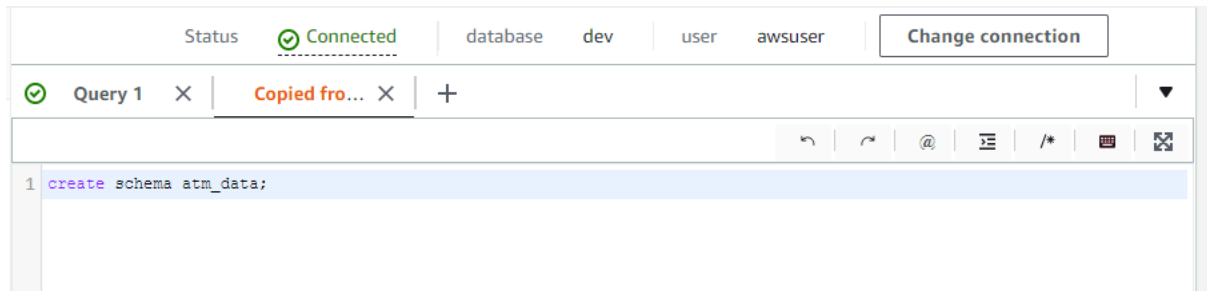
© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences



We create a folders to store the data of dimension and fact tables data in the S3 bucket.

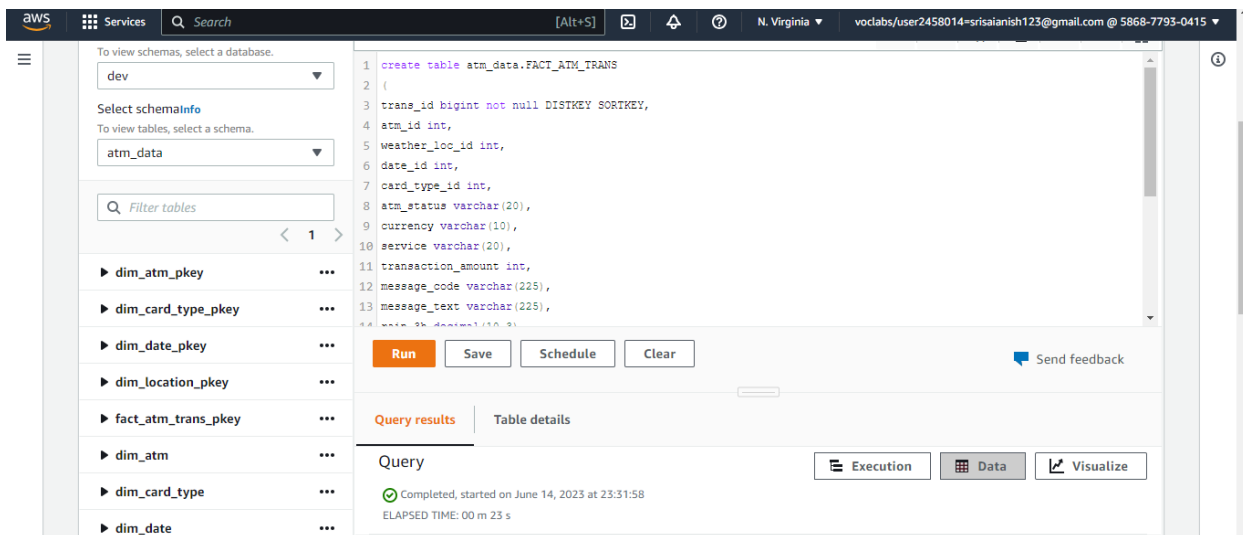


Creating a schema named atm_data to store all the tables

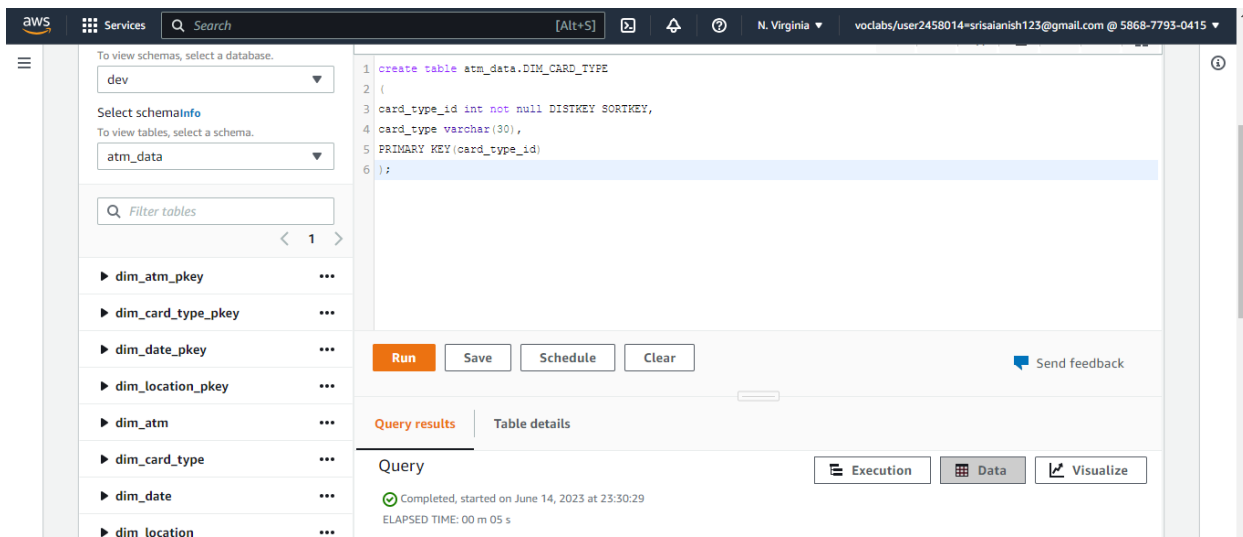


Creating fact and dimension tables

FACT_ATM_TRANS TABLE:



DIM_CARD_TYPE TABLE:



DIM_ATM TABLE:

The screenshot shows the AWS Glue console interface. On the left, the 'To view schemas, select a database.' dropdown is set to 'dev'. Below it, 'Select schema' is set to 'atm_data'. A list of tables is shown, including 'dim_atm_pkey', 'dim_location_pkey', 'dim_atm', and 'dim_location'. The main editor displays the SQL code for creating the 'atm_data.DIM_ATM' table:

```
1 create table atm_data.DIM_ATM
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );
```

Below the code, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link. Below the code editor, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating 'Completed, started on June 14, 2023 at 23:27:42' and 'ELAPSED TIME: 00 m 11 s'. To the right of the 'Query' section are buttons for 'Execution', 'Data', and 'Visualize'.

DIM_DATE TABLE:

The screenshot shows the AWS Glue console interface. On the left, the 'To view schemas, select a database.' dropdown is set to 'dev'. Below it, 'Select schema' is set to 'atm_data'. A list of tables is shown, including 'dim_atm_pkey', 'dim_date_pkey', 'dim_location_pkey', 'dim_atm', 'dim_date', and 'dim_location'. The main editor displays the SQL code for creating the 'atm_data.DIM_DATE' table:

```
1 create table atm_data.DIM_DATE
2 (
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int,
6   month varchar(20),
7   day int,
8   hour int,
9   weekday varchar(20),
10  PRIMARY KEY(date_id)
11 );
```

Below the code, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link. Below the code editor, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating 'Completed, started on June 14, 2023 at 23:29:28' and 'ELAPSED TIME: 00 m 06 s'. To the right of the 'Query' section are buttons for 'Execution', 'Data', and 'Visualize'.

Loading data from S3 bucket into a RedShift cluster

Copying data to DIM_LOCATION TABLE:

The screenshot shows the Amazon Redshift Query Editor interface. On the left, the navigation pane includes 'Amazon Redshift', 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters' (with sub-items 'Reserved nodes' and 'Snapshots'), 'Query editor' (with sub-items 'Query editor v2' and 'Queries and loads'), and 'Datashares'. The main panel displays the 'dev' database and 'public' schema. A query is entered in the editor:

```
1 copy atm_data.dim_location from 's3://etlprojectbucket/dim_location/part-00000-  
db88e974-4ec1-4cd1-b980-8496a258310e-c000.csv'  
2 iam_role 'arn:aws:iam::586877930415:role/myRedshiftRole'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

The query is executed successfully, as indicated by the 'Query 571' status and the 'Completed' message. The execution time is 00 m 16 s. The bottom of the interface shows the AWS logo, 'CloudShell', 'Feedback', 'Language', and copyright information for Amazon Web Services, Inc. or its affiliates.

Copying data to DIM_ATM TABLE:

The screenshot shows the Amazon Redshift Query Editor interface. On the left, the navigation pane includes 'Amazon Redshift', 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters' (with sub-items 'Reserved nodes' and 'Snapshots'), 'Query editor' (with sub-items 'Query editor v2' and 'Queries and loads'), and 'Datashares'. The main panel displays the 'dev' database and 'public' schema. A query is entered in the editor:

```
1 copy atm_data.dim_atm from 's3://etlprojectbucket/dim_atm/part-00000-e0cd3179-d363-  
417b-b8d2-ee85894c479d-c000.csv'  
2 iam_role 'arn:aws:iam::586877930415:role/myRedshiftRole'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

The query is executed successfully, as indicated by the 'Query 645' status and the 'Completed' message. The execution time is 00 m 03 s. The bottom of the interface shows the AWS logo, 'CloudShell', 'Feedback', 'Language', and copyright information for Amazon Web Services, Inc. or its affiliates.

Copying data to DIM_DATE TABLE:

The screenshot shows the Amazon Redshift Query Editor interface. On the left, the navigation pane includes 'Amazon Redshift', 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', and 'Datashares'. The main panel is divided into three sections: 'dev' (selected), 'Select schemaInfo' (showing 'public' as the selected schema), and 'Filter tables' (showing 'No resources' and 'No resources to display'). The right pane displays a SQL query (Query 678) that copies data from an S3 bucket to the 'dim_date' table. The query is as follows:

```
1 copy atm_data.dim_date from 's3://etlprojectbucket/dim_date/part-00000-9e9b7a94-8da6-4c65-895b-bfc6d0a138e7-c000.csv'
2 iam_role 'arn:aws:iam::586877930415:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
6
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is active, showing 'Query 678' with a status of 'Completed, started on June 14, 2023 at 23:51:24' and 'ELAPSED TIME: 00 m 03 s'. The 'Table details' tab is also visible. The bottom of the interface shows the 'CloudShell' button, 'Feedback', 'Language', and copyright information for Amazon Web Services, Inc. or its affiliates.

Copying data to DIM_CARD_TYPE TABLE:

The screenshot shows the Amazon Redshift Query Editor interface. On the left, the navigation pane includes 'Amazon Redshift', 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', and 'Datashares'. The main panel is divided into three sections: 'dev' (selected), 'Select schemaInfo' (showing 'atm_data' as the selected schema), and 'Filter tables' (showing a list of tables: 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', and 'dim_date'). The right pane displays a SQL query (Query 711) that copies data from an S3 bucket to the 'dim_card_type' table. The query is as follows:

```
1 copy atm_data.dim_card_type from 's3://etlprojectbucket/dim_card_type/part-00000-1967544f-8efa-4371-b9f2-8a116ced100d-c000.csv'
2 iam_role 'arn:aws:iam::586877930415:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is active, showing 'Query 711' with a status of 'Completed, started on June 14, 2023 at 23:53:15' and 'ELAPSED TIME: 00 m 13 s'. The 'Table details' tab is also visible. The bottom of the interface shows the 'CloudShell' button, 'Feedback', 'Language', and copyright information for Amazon Web Services, Inc. or its affiliates.

Copying data to FACT_ATM_TRANS TABLE:

The screenshot displays the Amazon Redshift Query Editor interface. On the left, the navigation pane shows the 'Amazon Redshift' section with options for 'Redshift Serverless' and 'Provisioned clusters dashboard'. Below these are links for 'Query editor', 'Query editor v2', 'Queries and loads', and 'Datashares'. The main workspace is divided into three panes. The top-left pane shows the 'dev' database selected, with a 'public' schema chosen. The bottom-left pane displays 'No resources' and 'No resources to display'. The right pane contains a SQL query:

```
1 copy atm_data.fact_atm_trans from 's3://etlprojectbucket/fact_atm_trans/part-00000-740df732-73fd-430c-b3b0-9535bb02dbc7-c000.csv'
2 iam_role 'arn:aws:iam::586877930415:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

 Below the query are buttons for 'Run', 'Save', 'Schedule', and 'Clear', along with a 'Send feedback' link. The bottom section shows 'Query results' and 'Table details' tabs. The 'Query results' tab is active, displaying 'Query 760' with a status of 'Completed, started on June 14, 2023 at 23:55:55' and an 'ELAPSED TIME: 00 m 10 s'.