Creation of a Redshift Cluster
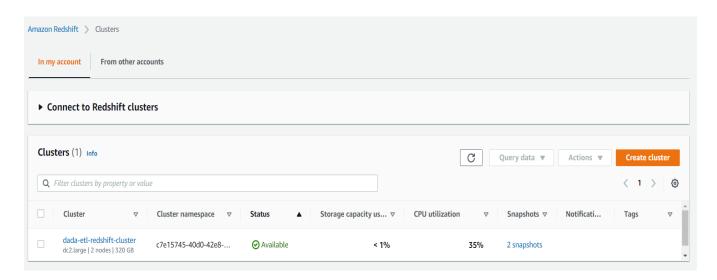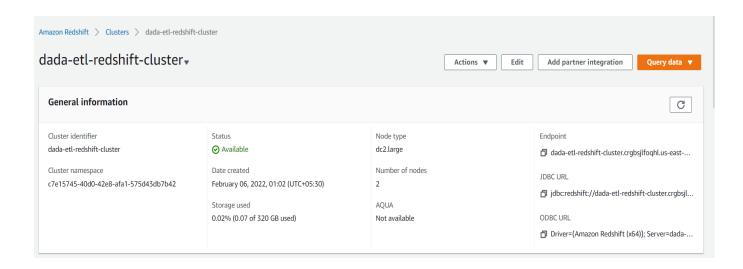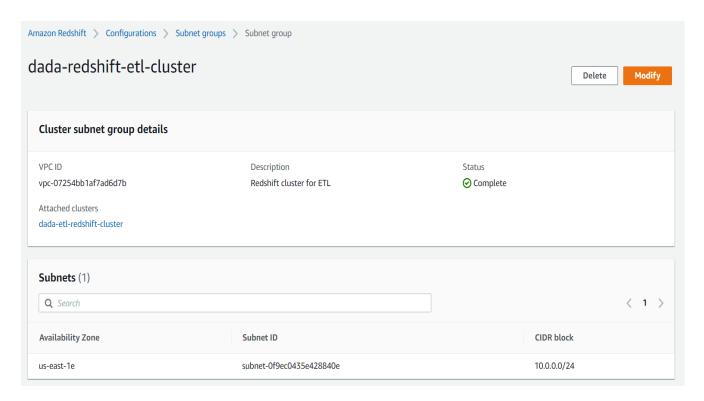
Screenshots of the configuration of the Redshift cluster that you have created:

Amazon Redshift > Clusters

In my account | From other accounts

▶ **Connect to Redshift clusters**

Clusters (1) Info

Query data ▼ | Actions ▼ | **Create cluster**

Filter clusters by property or value

< 1 >

	Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapshots	Notificati...	Tags
	dada-etl-redshift-cluster dc2.large | 2 nodes | 320 GB	c7e15745-40d0-42e8-...	⊘ Available	< 1%	35%	2 snapshots		

Amazon Redshift > Clusters > dada-etl-redshift-cluster

dada-etl-redshift-cluster ▾

Actions ▼ | Edit | Add partner integration | Query data ▼

General information

Cluster identifier dada-etl-redshift-cluster	Status ⊘ Available	Node type dc2.large	Endpoint ⧉ dada-etl-redshift-cluster.crgbsjlfoqhl.us-east-...
Cluster namespace c7e15745-40d0-42e8-afa1-575d43db7b42	Date created February 06, 2022, 01:02 (UTC+05:30)	Number of nodes 2	JDBC URL ⧉ jdbc:redshift://dada-etl-redshift-cluster.crgbsjl...
	Storage used 0.02% (0.07 of 320 GB used)	AQUA Not available	ODBC URL ⧉ Driver={Amazon Redshift (x64)}; Server=dada-...

Amazon Redshift > Configurations > Subnet groups > Subnet group

dada-redshift-etl-cluster

Delete   Modify

Cluster subnet group details

VPC ID	Description	Status
vpc-07254bb1af7ad6d7b	Redshift cluster for ETL	⊘ Complete

Attached clusters
dada-etl-redshift-cluster

Subnets (1)

🔍 Search   ‹ 1 ›

Availability Zone	Subnet ID	CIDR block
us-east-1e	subnet-0f9ec0435e428840e	10.0.0.0/24

Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

CREATE SCHEMA etl;

CREATE TABLE etl.loc (
        location_id INTEGER NOT NULL
        ,atm_location VARCHAR(50)
        ,atm_streetname VARCHAR(255)
        ,atm_street_number INTEGER
        ,atm_zipcode INTEGER
        ,atm_lat NUMERIC(10, 3)
        ,atm_lon NUMERIC(10, 3)
        ,PRIMARY KEY (location_id)
        );

```sql
CREATE TABLE etl.atm (
        atm_prim_id INTEGER NOT NULL
        ,atm_id VARCHAR(20)
        ,atm_manufacturer VARCHAR(50)
        ,location_id INTEGER
        ,PRIMARY KEY (atm_prim_id)
        ,FOREIGN KEY (location_id) REFERENCES etl.loc(location_id)
        );


CREATE TABLE etl.DATE (
        year INTEGER
        ,month VARCHAR(20)
        ,day INTEGER
        ,hour INTEGER
        ,weekday VARCHAR(20)
        ,full_date_time TIMESTAMP
        ,date_id INTEGER
        ,PRIMARY KEY (date_id)
        );


CREATE TABLE etl.card (
        card_type VARCHAR(23)
        ,card_type_id INTEGER
        ,PRIMARY KEY (card_type_id)
        );
```

```sql
CREATE TABLE etl.FACT_ATM_TRANS (
        trans_id BIGINT NOT NULL
        ,atm_prim_id INTEGER
        ,location_id INTEGER
        ,date_id INTEGER
        ,card_type_id INTEGER
        ,atm_status VARCHAR(20)
        ,currency VARCHAR(10)
        ,service VARCHAR(20)
        ,transaction_amount INTEGER
        ,message_code VARCHAR(255)
        ,message_text VARCHAR(255)
        ,rain_3h NUMERIC(10, 3)
        ,clouds_all INTEGER
        ,weather_id INTEGER
        ,weather_main VARCHAR(50)
        ,weather_description VARCHAR(255)
        ,PRIMARY KEY (trans_id)
        ,FOREIGN KEY (atm_prim_id) REFERENCES etl.atm(atm_prim_id)
        ,FOREIGN KEY (location_id) REFERENCES etl.loc(location_id)
        ,FOREIGN KEY (date_id) REFERENCES etl.DATE (date_id)
        ,FOREIGN KEY (card_type_id) REFERENCES etl.card(card_type_id)
        );
```
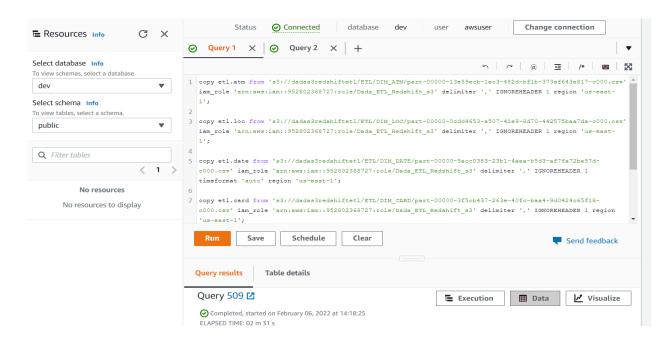
Select database Info
To view schemas, select a database.

dev ▼

Select schema Info
To view tables, select a schema.

etl ▼

🔍 Filter tables

‹ 1 ›

▶ atm_pkey ⋯
▶ card_pkey ⋯
▶ date_pkey ⋯
▶ fact_atm_trans_pkey ⋯
▶ loc_pkey ⋯
▶ atm ⋯
▶ card ⋯
▶ date ⋯
▶ fact_atm_trans ⋯
▶ loc ⋯

Loading data into a Redshift cluster from Amazon S3 bucket

Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

copy etl.atm from 's3://dadas3redshiftetl/ETL/DIM_ATM/part-00000-13e59ecb-1ec3-482d-bf1b-373ef643e817-c000.csv' iam_role 'arn:aws:iam::952802368727:role/Dada_ETL_Redshift_s3' delimiter ',' IGNOREHEADER 1 region 'us-east-1';

copy etl.loc from 's3://dadas3redshiftetl/ETL/DIM_LOC/part-00000-0cdd4653-a507-41e9-8d70-442575baa7da-c000.csv' iam_role 'arn:aws:iam::952802368727:role/Dada_ETL_Redshift_s3' delimiter ',' IGNOREHEADER 1 region 'us-east-1';

copy etl.date from 's3://dadas3redshiftetl/ETL/DIM_DATE/part-00000-5acc0383-23b1-4aea-b5d3-af7fa72be57d-c000.csv' iam_role 'arn:aws:iam::952802368727:role/Dada_ETL_Redshift_s3' delimiter ',' IGNOREHEADER 1 timeformat 'auto' region 'us-east-1';

copy etl.card from 's3://dadas3redshiftetl/ETL/DIM_CARD/part-00000-3f5cb437-263e-40fc-baa4-9d0424c65f18-c000.csv' iam_role 'arn:aws:iam::952802368727:role/Dada_ETL_Redshift_s3' delimiter ',' IGNOREHEADER 1 region 'us-east-1';

copy etl.FACT_ATM_TRANS from 's3://dadas3redshiftetl/ETL/FACT_ATM_TRANS/part-00000-b63eac25-f44b-4363-a568-5bca6f00f734-c000.csv' iam_role 'arn:aws:iam::952802368727:role/Dada_ETL_Redshift_s3' delimiter ',' IGNOREHEADER 1 region 'us-east-1' TRUNCATECOLUMNS CSV;