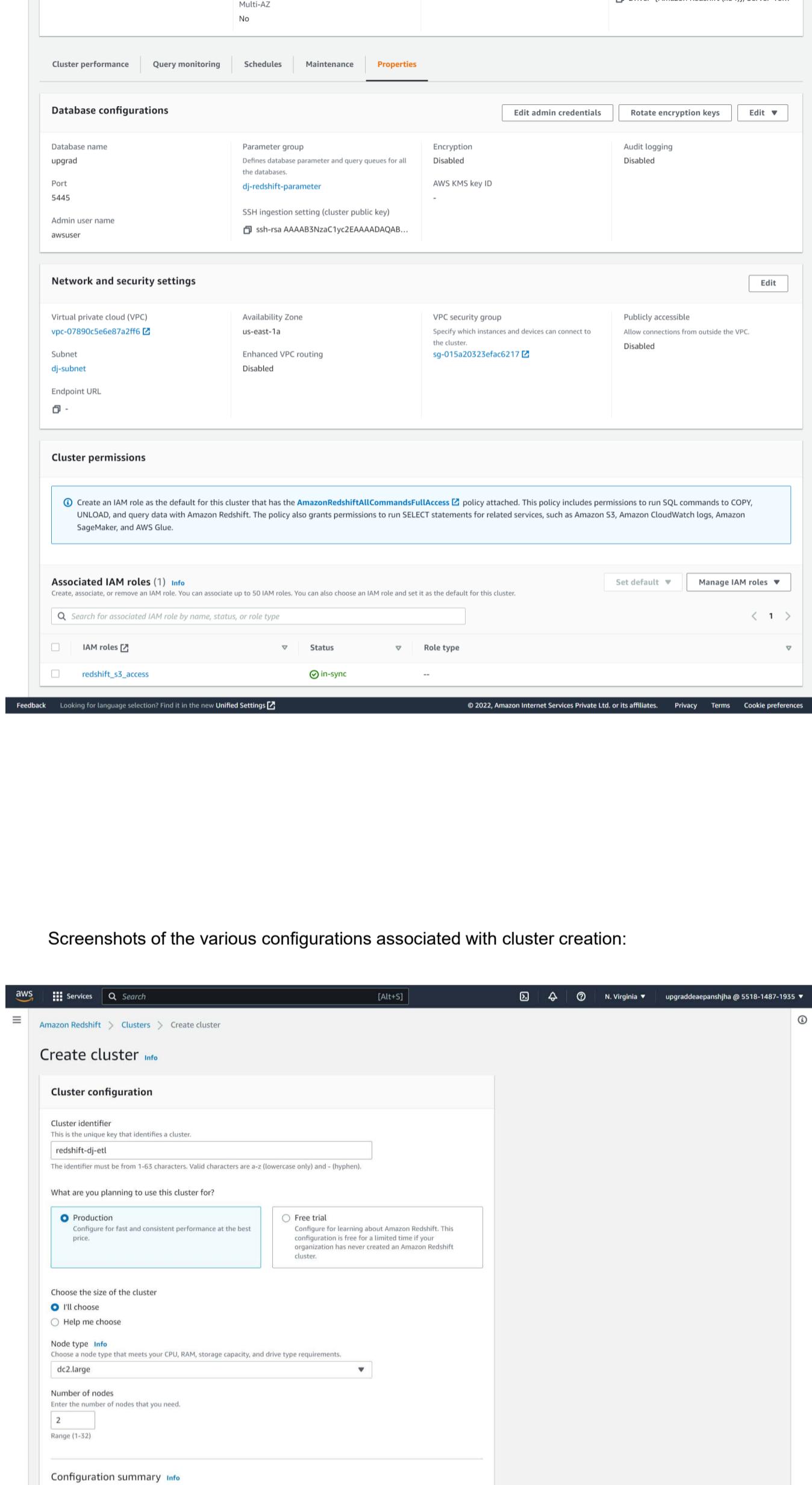


Creation of a Redshift Cluster

Screenshots of the configuration of the Redshift cluster that you have created:

Screenshot of the type of machine used along with number of nodes:



The screenshot shows the AWS Amazon Redshift cluster configuration page for a cluster named "redshift-dj-etl".

General information:

Cluster identifier: redshift-dj-etl	Status: Available	Node type: dc2.large	Endpoint: redshift-dj-etl.cnyvkjs9bcu.us-east-1.redshift.amazonaws.com
Cluster namespace: 99322027-dd9c-484f-9c7b-4d267ad4a698	Date created: December 15, 2022, 13:23 (UTC+05:30)	Number of nodes: 2	JDBC URL: jdbc:redshift://redshift-dj-etl.cnyvkjs9bcu...
Cluster configuration: Production	Storage used: -		ODBC URL: Driver={Amazon Redshift (x64)}; Server=re...
	Multi-AZ: No		

Database configurations:

Database name: upgrad	Parameter group: dj-redshift-parameter	Encryption: Disabled	Audit logging: Disabled
Port: 5445		AWS KMS key ID: -	
Admin user name: awssuser	SSH ingestion setting (cluster public key): ssh-rsa AAAAB3NzaC1yc2EAAQAB...		

Network and security settings:

Virtual private cloud (VPC): vpc-07890c5e6e87a2ff6	Availability Zone: us-east-1a	VPC security group: sg-015a20323efac6217	Publicly accessible: Disabled
Subnet: dj-subnet	Enhanced VPC routing: Disabled		
Endpoint URL: -			

Cluster permissions:

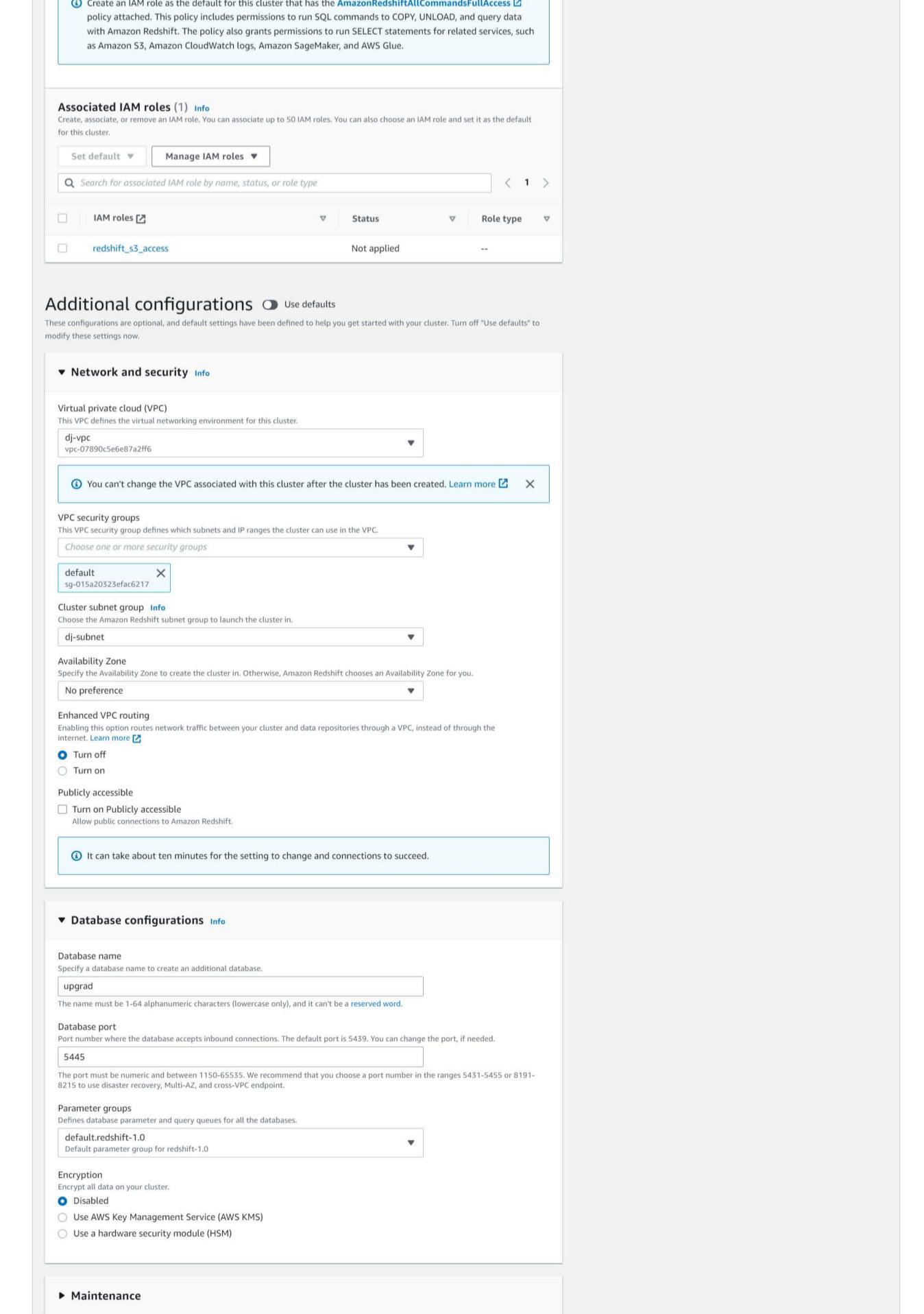
Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1) [Info](#)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

IAM roles	Status	Role type
redshift_s3_access	in-sync	--

Screenshots of the various configurations associated with cluster creation:



The screenshot shows the AWS Create cluster configuration page.

Cluster configuration:

Cluster identifier: redshift-dj-etl
Cluster identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

Production: Configure for fast and consistent performance at the best price.
 Free trial: Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster:
 I'll choose
 Help me choose

Node type: dc2.large

Number of nodes: 2

Configuration summary:

\$360.00/month	Estimated on-demand compute price
Estimated on-demand compute price	Save more than 60% of your costs by purchasing reserved nodes.
dc2.large 2 nodes	Total compressed storage
	The total storage capacity for the cluster if you deploy the number of nodes that you chose.

Sample data:

Load sample data: Load sample data to your Redshift cluster to start using the query editor to query data.

Database configurations:

Admin user name: awssuser
Admin user password: (password field)

Cluster permissions:

Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1) [Info](#)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

IAM roles	Status	Role type
redshift_s3_access	Not applied	--

Additional configurations:

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

Network and security:

Virtual private cloud (VPC): dj-vpc
You can't change the VPC associated with this cluster after the cluster has been created.

VPC security groups:

This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.
Choose one or more security groups: sg-015a20323efac6217

Cluster subnet group:

Choose the Amazon Redshift subnet group to launch the cluster in.
dj-subnet

Availability Zone:

Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.
No preference

Enhanced VPC routing:

Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the internet.

Turn off
 Turn on

Publicly accessible:

Turn on Publicly accessible: Allow public connections to Amazon Redshift.

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

Database configurations:

Database name: upgrad
Database port: 5445
Parameter groups: default.redshift-1.0

Maintenance:

Monitoring:

Backup:

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

Viewing all the data in Amazon S3 bucket:

The screenshot shows the AWS S3 console interface. The left sidebar is collapsed. The main area shows the 'dj-etlproject' bucket. The 'Objects' tab is selected. There are 5 objects listed:

Name	Type	Last modified	Size	Storage class
dim_atm/	Folder	-	-	-
dim_card_type/	Folder	-	-	-
dim_date/	Folder	-	-	-
dim_location/	Folder	-	-	-
fact_atm_trans/	Folder	-	-	-

Viewing all the data inside dim_atm:

The screenshot shows the AWS S3 console interface. The left sidebar is collapsed. The main area shows the 'dim_atm/' folder within the 'dj-etlproject' bucket. The 'Objects' tab is selected. There are 4 objects listed:

Name	Type	Last modified	Size	Storage class
_SUCCESS	-	December 15, 2022, 12:00:40 (UTC+05:30)	0 B	Standard
_SUCCESS.crc	crc	December 15, 2022, 12:00:39 (UTC+05:30)	8.0 B	Standard
.part-00000-a030b129-6452-4a54-88d6-3c58c3523aec-c000.csv.crc	crc	December 15, 2022, 12:00:38 (UTC+05:30)	32.0 B	Standard
.part-00000-a030b129-6452-4a54-88d6-3c58c3523aec-c000.csv	csv	December 15, 2022, 12:00:40 (UTC+05:30)	2.6 KB	Standard

Viewing all the data inside dim_card_type:

The screenshot shows the AWS S3 console interface. The left sidebar is collapsed. The main area shows the 'dim_card_type/' folder within the 'dj-etlproject' bucket. The 'Objects' tab is selected. There are 4 objects listed:

Name	Type	Last modified	Size	Storage class
_SUCCESS	-	December 15, 2022, 12:01:08 (UTC+05:30)	0 B	Standard
_SUCCESS.crc	crc	December 15, 2022, 12:01:07 (UTC+05:30)	8.0 B	Standard
.part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv.crc	crc	December 15, 2022, 12:01:06 (UTC+05:30)	12.0 B	Standard
.part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv	csv	December 15, 2022, 12:01:08 (UTC+05:30)	195.0 B	Standard

Viewing all the data inside dim_date:

The screenshot shows the AWS S3 console interface. The left sidebar is collapsed. The main area shows the 'dim_date/' folder within the 'dj-etlproject' bucket. The 'Objects' tab is selected. There are 4 objects listed:

Name	Type	Last modified	Size	Storage class
_SUCCESS	-	December 15, 2022, 12:01:36 (UTC+05:30)	0 B	Standard
_SUCCESS.crc	crc	December 15, 2022, 12:01:31 (UTC+05:30)	8.0 B	Standard
.part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv.crc	crc	December 15, 2022, 12:01:06 (UTC+05:30)	4.1 KB	Standard
.part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv	csv	December 15, 2022, 12:01:35 (UTC+05:30)	521.2 KB	Standard

Viewing all the data inside fact_atm_trans:

Query to create a schema for the dimension and fact tables:

```
create schema atm_data;
```

The screenshot shows the Amazon Redshift Query Editor interface. On the left, there's a sidebar with 'Resources' and 'Info' sections. The main area has tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. A status bar at the top indicates 'Connected' to 'upgrad' database as user 'awsuser'. The query editor window contains a 'Query 1' tab with the SQL command 'create schema atm_data;'. Below the query are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab shows a completed execution with a duration of 00 m 16 s.

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

❖ Creating location dimension table

```
CREATE TABLE atm_data.dim_location
(
    location_id INT NOT NULL distkey sortkey,
    location VARCHAR(50),
    streetname VARCHAR(255),
    street_number INT,
    zipcode INT,
    lat DECIMAL(10, 3),
    lon DECIMAL(10, 3),
    PRIMARY KEY(location_id)
);
```

The screenshot shows the Amazon Redshift Query Editor interface. The sidebar shows 'dim_location_pkey' and 'dim_location' under the 'atm_data' schema. The main area has tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. A status bar at the top indicates 'Connected' to 'upgrad' database as user 'awsuser'. The query editor window contains a 'Query 1' tab with the SQL command 'create table atm_data.DIM_LOCATION ...'. Below the query are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab shows a completed execution with a duration of 00 m 20 s.

❖ Creating atm dimension table

```
CREATE TABLE atm_data.dim_atm
(
    atm_id INT NOT NULL distkey sortkey,
    atm_number VARCHAR(20),
    atm_manufacturer VARCHAR(50),
    atm_location_id INT,
    PRIMARY KEY(atm_id),
    FOREIGN KEY(atm_location_id) REFERENCES atm_data.dim_location(location_id)
);
```

The screenshot shows the Amazon Redshift Query Editor interface. The sidebar shows 'dim_atm_pkey', 'dim_location_pkey', 'dim_atm', and 'dim_location' under the 'atm_data' schema. The main area has tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. A status bar at the top indicates 'Connected' to 'upgrad' database as user 'awsuser'. The query editor window contains a 'Query 1' tab with the SQL command 'create table atm_data.DIM_ATM ...'. Below the query are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab shows a completed execution with a duration of 00 m 11 s.

❖ Creating date dimension table

```
CREATE TABLE atm_data.dim_date
(
    date_id INT NOT NULL distkey sortkey,
    full_date_time timestamp,
    year INT,
    month VARCHAR(20),
    day INT,
    hour INT,
    weekday VARCHAR(20),
    PRIMARY KEY(date_id)
);
```

❖ Creating card type dimension table

```
CREATE TABLE atm_data.dim_card_type
(
    card_type_id INT NOT NULL distkey sortkey,
    card_type     VARCHAR(30) PRIMARY KEY(card_type_id)
);
```

The screenshot shows the Amazon Redshift Query Editor interface. On the left, there's a sidebar with 'Resources' and 'Info' sections. The 'Info' section shows the current database is 'upgrad' and the schema is 'atm_data'. Below it is a 'Filter tables' search bar. The main area has a status bar at the top indicating 'Connected' to 'N. Virginia' with user 'awsuser'. A query editor window titled 'Query 1' is open, containing the SQL code for creating the 'dim_card_type' table. Below the editor are tabs for 'Query results', 'Table details', 'Execution', 'Data', and 'Visualize'. The status bar at the bottom indicates the query completed successfully on December 15, 2022, at 14:40:50.

❖ Creating atm transactions fact table

```
CREATE TABLE atm_data.fact_atm_trans
(
    trans_id          BIGINT NOT NULL distkey sortkey,
    atm_id            INT,
    weather_loc_id   INT,
    date_id           INT,
    card_type_id     INT,
    atm_status        VARCHAR(20),
    currency          VARCHAR(10),
    service           VARCHAR(20),
    transaction_amount INT,
    message_code      VARCHAR(225),
    message_text      VARCHAR(225),
    rain_3h           DECIMAL(10, 3),
    clouds_all        INT,
    weather_id        INT,
    weather_main       VARCHAR(50),
    weather_description VARCHAR(255),
    PRIMARY KEY(trans_id),
    FOREIGN KEY(weather_loc_id) REFERENCES atm_data.dim_location(location_id),
    FOREIGN KEY(atm_id) REFERENCES atm_data.dim_atm(atm_id),
    FOREIGN KEY(date_id) REFERENCES atm_data.dim_date(date_id),
    FOREIGN KEY(card_type_id) REFERENCES atm_data.dim_card_type(card_type_id)
);
```

The screenshot shows the Amazon Redshift Query Editor interface. The sidebar and status bar are identical to the previous screenshot. The 'Query 1' editor contains the SQL code for creating the 'fact_atm_trans' table. The status bar at the bottom indicates the query completed successfully on December 15, 2022, at 14:48:51.

❖ Loading data into a RedShift cluster from Amazon S3 bucket

Queries to copy the data from S3 bucket to the RedShift cluster in the appropriate tables:

❖ Copying the data to dim_location table

```
copy atm_data.dim_location FROM 's3://dj-etlproject/dim_location/part-00000-f18338f1-3918-4b2c-8849-37e5cc2fcc5f-c000.csv'
iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'
delimiter ',' region 'us-east-1'
csv;
```

The screenshot shows the Amazon Redshift Query Editor interface. The sidebar and status bar are identical to the previous screenshots. The 'Query 1' editor contains the COPY command for the 'dim_location' table. The status bar at the bottom indicates the query completed successfully on December 15, 2022, at 14:53:20.

❖ Copying the data to dim_atm table

```
copy atm_data.dim_atm FROM 's3://dj-etlproject/dim_atm/part-00000-a030b129-6452-4a54-88d6-3c58c3523aec-c000.csv'
iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'
delimiter ',' region 'us-east-1'
csv;
```

The screenshot shows the Amazon Redshift Query Editor interface. The sidebar and status bar are identical to the previous screenshots. The 'Query 1' editor contains the COPY command for the 'dim_atm' table. The status bar at the bottom indicates the query completed successfully on December 15, 2022, at 14:56:18.

❖ Copying the data to dim_date table

```
copy atm_data.dim_date FROM 's3://dj-etlproject/dim_date/part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv'  
iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
delimiter ',' region 'us-east-1'  
timeformat 'YYYY-MM-DDTHH:MI:SS'  
csv;
```

The screenshot shows the AWS Management Console interface for Amazon Redshift. The top navigation bar includes the AWS logo, 'Services' dropdown, search bar, and account information ('N. Virginia' and 'upgraddeapepanshjha @ 5518-1487-1935'). Below the navigation is the 'Amazon Redshift > Query editor' section. The left sidebar contains 'Resources' and 'Info' sections for database and schema selection. The main area is titled 'Query 1' and contains the following SQL code:

```
1 copy atm_data.dim_date from 's3://dj-etlproject/dim_date/part-00000-d4f50d1e-d29e-4e99-90c1-df835133903a-c000.csv'  
2 iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
3 delimiter ',' region 'us-east-1'  
4 timeformat 'YYYY-MM-DDTHH:MI:SS'  
5 CSV;
```

Below the code are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a completed query (Query 200676) with a green status icon, indicating it started on December 15, 2022, at 15:00:34 and took 00 m 21 s. Other tabs include 'Table details', 'Execution', 'Data', and 'Visualize'. At the bottom of the page are 'Feedback', 'Unified Settings' (link), copyright notice ('© 2022, Amazon Internet Services Private Ltd. or its affiliates.'), and links for 'Privacy', 'Terms', and 'Cookie preferences'.

❖ Copying the data to dim_card_type table

```
copy atm_data.dim_card_type FROM 's3://dj-etlproject/dim_card_type/part-00000-2ad35cbe-49dc-486e-b473-461a3e8d6998-c000.csv'  
iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
delimiter ',' region 'us-east-1'  
csv;
```

The screenshot shows the AWS Management Console interface for Amazon Redshift. The top navigation bar includes the AWS logo, 'Services' dropdown, search bar, and account information ('N. Virginia' and 'upgraddeapepanshjha @ 5518-1487-1935'). Below the navigation is the 'Amazon Redshift > Query editor' section. The left sidebar contains 'Resources' and 'Info' sections for database and schema selection. The main area is titled 'Query 1' and contains the following SQL code:

```
1 copy atm_data.dim_card_type from 's3://dj-etlproject/dim_card_type/part-00000-2ad35cbe-49dc-486e-b473-461a3e8d6998-c000.csv'  
2 iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

Below the code are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a completed query (Query 200727) with a green status icon, indicating it started on December 15, 2022, at 15:03:08 and took 00 m 10 s. Other tabs include 'Table details', 'Execution', 'Data', and 'Visualize'. At the bottom of the page are 'Feedback', 'Unified Settings' (link), copyright notice ('© 2022, Amazon Internet Services Private Ltd. or its affiliates.'), and links for 'Privacy', 'Terms', and 'Cookie preferences'.

❖ Copying the data to fact_atm_trans table

```
copy atm_data.fact_atm_trans FROM 's3://dj-etlproject/fact_atm_trans/part-00000-8f9a5f90-fdf3-472c-bf66-3dc131a877dd-c000.csv'  
iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
delimiter ',' region 'us-east-1'  
csv;
```

The screenshot shows the AWS Management Console interface for Amazon Redshift. The top navigation bar includes the AWS logo, 'Services' dropdown, search bar, and account information ('N. Virginia' and 'upgraddeapepanshjha @ 5518-1487-1935'). Below the navigation is the 'Amazon Redshift > Query editor' section. The left sidebar contains 'Resources' and 'Info' sections for database and schema selection. The main area is titled 'Query 1' and contains the following SQL code:

```
1 copy atm_data.fact_atm_trans from 's3://dj-etlproject/fact_atm_trans/part-00000-8f9a5f90-fdf3-472c-bf66-3dc131a877dd-c000.csv'  
2 iam_role 'arn:aws:iam::551814871935:role/redshift_s3_access'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

Below the code are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a completed query (Query 200769) with a green status icon, indicating it started on December 15, 2022, at 15:05:39 and took 00 m 25 s. Other tabs include 'Table details', 'Execution', 'Data', and 'Visualize'. At the bottom of the page are 'Feedback', 'Unified Settings' (link), copyright notice ('© 2022, Amazon Internet Services Private Ltd. or its affiliates.'), and links for 'Privacy', 'Terms', and 'Cookie preferences'.