

Redshift Setup

Redshift cluster creation

The screenshot displays the AWS Management Console interface for an Amazon Redshift cluster. The top navigation bar includes the AWS logo, 'Services', a search bar, and the user's profile information. The breadcrumb trail indicates the path: Amazon Redshift > Clusters > etl-redshift-cluster. The cluster name 'etl-redshift-cluster' is prominently displayed at the top of the main content area, accompanied by buttons for 'Actions', 'Edit', 'Add partner integration', and 'Query data'. Below this, the 'General information' tab is selected, showing a table of cluster details. The table is organized into four columns: Cluster identifier, Status, Node type, and Endpoint. The 'Status' column shows the cluster is 'Available'. The 'Node type' is 'dc2.large'. The 'Endpoint' column provides the JDBC and ODBC URLs. The bottom of the console features a series of tabs for 'Cluster performance', 'Query monitoring', 'Databases', 'Datashares', 'Zero-ETL integrations', 'Resource Policy', 'Schedules', and 'Maintenance'. The footer contains the 'CloudShell' icon, a 'Feedback' link, and the copyright notice for Amazon Web Services, Inc.

aws Services Search [Alt+S] N. Virginia voclabs/user3148501=bharatpanera69@gmail.com @ 6546-5431-0080

Amazon Redshift > Clusters > etl-redshift-cluster

etl-redshift-cluster

Actions Edit Add partner integration Query data

General information Info

Cluster identifier etl-redshift-cluster	Status Available	Node type dc2.large	Endpoint etl-redshift-cluster.cky9xmqudfknn.us-east-1.redshift.amazonaws.com:5440/upgrad
Custom domain name -	Date created June 10, 2024, 20:20 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://etl-redshift-cluster.cky9xmqudfknn.us-east-1.redshift.amazonaws.com:5440/upgrad
Cluster namespace ARN arn:aws:redshift:us-east-1:654654310080:namespace:63c5da94-0db3-4ac6-aba8-e95ddf1d6787	Storage used 0.02% (0.07 of 320 GB used)	Patch version Patch 181	ODBC URL Driver={Amazon Redshift (x64)}; Server=etl-redshift-cluster.cky9xmqudfknn.us-east-1.redshift.amazonaws.com; Database=upgrad
Cluster configuration Production	Multi-AZ No		

Cluster performance Query monitoring Databases Datashares Zero-ETL integrations Resource Policy Schedules Maintenance

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Create schema

The screenshot displays the AWS CloudShell interface. At the top, the AWS logo and 'Services' menu are visible. The search bar contains the text 'Search'. The status bar shows 'Status' as 'Connected' with a green checkmark, and the connection details include 'database', 'upgrad', 'user', and 'awsuser'. A 'Change connection' button is present. The main area shows a SQL query editor with the text '1 CREATE SCHEMA atm_data;'. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also visible. The 'Query results' tab is selected, showing a 'Query' section with a green checkmark indicating the query was 'Completed, started on June 10, 2024 at 20:57:23' and 'ELAPSED TIME: 00 m 54 s'. To the right of the query results are buttons for 'Execution', 'Data', and 'Visualize'. The bottom of the interface features a 'CloudShell' logo, a 'Feedback' link, and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates.' along with links for 'Privacy', 'Terms', and 'Cookie preferences'.

aws Services Search [Alt+S] N. Virginia voclabs/user3148501=bharatpanera69@gmail.com @ 6546-5431-0080

Status **Connected** database upgrad user awsuser Change connection

Query 1 +

```
1 CREATE SCHEMA atm_data;
```

Run Save Schedule Clear Send feedback

Query results Table details

Query

Completed, started on June 10, 2024 at 20:57:23
ELAPSED TIME: 00 m 54 s

Execution Data Visualize

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Create dimension table - DIM_LOCATION

```
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)  
);
```

Screenshot: Create dimension table - DIM_LOCATION

The screenshot displays the AWS CloudShell interface. At the top, the AWS logo and 'Services' are visible. The status bar shows 'N. Virginia' and the user 'voclabs/user3148501=bharatpanera69@gmail.com @ 6546-5431-0080'. The main interface is divided into three sections:

- Resources:** On the left, under 'Resources Info', the 'upgrad' database is selected. Under 'Select schema Info', the 'atm_data' schema is selected. Below this, a list of tables is shown, including 'dim_location_pkey' and 'dim_location'.
- Query Editor:** The central area shows a SQL query for 'Query 1'. The query is:

```
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)  
);
```
- Execution and Results:** At the bottom, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. Below these, the 'Query results' tab is active, showing a status of 'Completed, started on June 10, 2024 at 21:00:30' and 'ELAPSED TIME: 00 m 22 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

The bottom of the interface shows the 'CloudShell' logo, a 'Feedback' link, and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

Create dimension table - DIM_ATM

```
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)
);
```

Screenshot: Create dimension table - DIM_ATM

The screenshot displays the AWS Glue console interface. On the left sidebar, under 'Resources', the 'upgrad' database and 'atm_data' schema are selected. The main panel shows a SQL query editor with the following code:

```
16  
17 CREATE TABLE atm_data.DIM_ATM  
18 (  
19   atm_id INT NOT NULL DISTKEY SORTKEY,  
20   atm_number VARCHAR(20),  
21   atm_manufacturer VARCHAR(50),  
22   atm_location_id INT,  
23   PRIMARY KEY(atm_id),  
24   FOREIGN KEY(atm_location_id) REFERENCES atm_data.DIM_LOCATION(location_id)  
25 );  
26  
27
```

Below the query editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link. Below the query editor, the 'Query results' tab is active, showing a status of 'Completed, started on June 10, 2024 at 21:02:14' and 'ELAPSED TIME: 00 m 07 s'. The 'Table details' tab is also visible. At the bottom of the console, there are links for 'CloudShell', 'Feedback', and '© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

Create dimension table - DIM_DATE

```
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)  
);
```


Screenshot: Create dimension table - DIM_DATE

The screenshot displays the AWS Glue console interface. On the left sidebar, under 'Resources', the 'upgrad' database and 'atm_data' schema are selected. The main panel shows a SQL query editor with the following code:

```
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)
);
```

Below the query editor, the 'Run' button is highlighted in orange. The 'Query results' tab is active, showing a status of 'Completed, started on June 10, 2024 at 21:03:29' and 'ELAPSED TIME: 00 m 06 s'. The bottom of the console features a footer with 'CloudShell', 'Feedback', and copyright information for Amazon Web Services.

Create dimension table - DIM_CARD_TYPE

```
CREATE TABLE atm_data.DIM_CARD_TYPE
```

```
(
```

```
card_type_id INT NOT NULL DISTKEY SORTKEY,
```

```
card_type VARCHAR(30),
```

```
PRIMARY KEY(card_type_id)
```

```
);
```

Screenshot: Create dimension table - DIM_CARD_TYPE

The screenshot displays the AWS Glue console interface. On the left, the 'Resources' sidebar shows the 'upgrad' database and 'atm_data' schema selected. The main panel shows a SQL query editor with the following code:

```
38  
39 CREATE TABLE atm_data.DIM_CARD_TYPE  
40 (  
41   card_type_id INT NOT NULL DISTKEY SORTKEY,  
42   card_type VARCHAR(30),  
43   PRIMARY KEY(card_type_id)  
44 );  
45  
46  
47  
48  
49  
50
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is active, showing the status 'Completed, started on June 10, 2024 at 21:04:42' and 'ELAPSED TIME: 00 m 05 s'. The bottom of the console shows the 'CloudShell' tab and the footer with copyright information for Amazon Web Services, Inc. or its affiliates.

Create Fact table - FACT_ATM_TRANS

```
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)
);
```

Create Fact table - FACT_ATM_TRANS

The screenshot displays the AWS Glue console interface. On the left, the 'Resources' sidebar shows a tree view of databases and schemas. The 'upgrad' database is selected, and the 'atm_data' schema is chosen. Below this, a list of tables is shown, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', and several dimension tables like 'dim_atm', 'dim_card_type', 'dim_date', and 'dim_location'. The 'fact_atm_trans' table is expanded, showing its columns: trans_id, atm_id, weather_loc_id, date_id, card_type_id, atm_status, currency, and service.

The main area displays 'Query 1' with the following SQL code:

```
46 CREATE TABLE atm_data.FACT_ATM_TRANS
47 (
48   trans_id BIGINT NOT NULL DISTKEY SORTKEY,
49   atm_id INT,
50   weather_loc_id INT,
51   date_id INT,
52   card_type_id INT,
53   atm_status VARCHAR(20),
54   currency VARCHAR(10),
55   service VARCHAR(20),
56   transaction_amount INT,
57   message_code VARCHAR(225),
58   message_text VARCHAR(225),
59   rain_3h DECIMAL(10,3),
60   clouds_all INT,
61   weather_id INT,
62   weather_main VARCHAR(50),
63   weather_description VARCHAR(255),
64   PRIMARY KEY(trans_id),
65   FOREIGN KEY(weather_loc_id) REFERENCES atm_data.DIM_LOCATION(location_id),
66   FOREIGN KEY(atm_id) REFERENCES atm_data.DIM_ATM(atm_id),
67   FOREIGN KEY(date_id) REFERENCES atm_data.DIM_DATE(date_id),
68   FOREIGN KEY(card_type_id) REFERENCES atm_data.DIM_CARD_TYPE(card_type_id)
69 );
```

Below the SQL editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The bottom section shows 'Query results' and 'Table details' tabs. The 'Query' tab is active, displaying a status of 'Completed, started on June 10, 2024 at 21:07:57' and 'ELAPSED TIME: 00 m 09 s'. At the bottom right, there are buttons for 'Execution', 'Data', and 'Visualize'.

At the very bottom of the console, there is a footer with the text '© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

Load data into a Redshift cluster from AWS S3 bucket

```
COPY atm_data.dim_location
FROM 's3://etl-atm-data/dim_location/part-00000-fdd8607e-c7f3-4bf1-85d6-034256dcf3ae-c000.csv'
IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;
```

Copy the data to table - dim_location

The screenshot displays the AWS Redshift console interface. At the top, the navigation bar includes the AWS logo, 'Services', a search bar, and account information for 'N. Virginia' and 'voclabs/user3148501=bharatpanera69@gmail.com @ 6546-5431-0080'.

The main workspace is divided into three sections:

- Left Panel (Resources Info):** Contains dropdowns for 'Select database' (set to 'upgrad') and 'Select schema' (set to 'atm_data'). Below these is a 'Filter tables' search bar and a list of tables including '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'.
- Center Panel (Query Editor):** Shows a SQL query for 'Query 1'. The query is:

```
77  
78  
79  
80  
81 COPY atm_data.dim_location  
82 FROM 's3://etl-atm-data/dim_location/part-00000-fdd8607e-c7f3-4bf1-85d6-034256dcf3ae-c000.csv'  
83 IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'  
84 DELIMITER ','  
85 REGION 'us-east-1'  
86 CSV;
```
- Right Panel (Query Results):** Displays the execution status of 'Query 2400'. It shows a green checkmark indicating 'Completed, started on June 10, 2024 at 21:21:12' with an 'ELAPSED TIME: 00 m 06 s'. Buttons for 'Execution', 'Data', and 'Visualize' are present.

At the bottom of the console, there is a 'CloudShell' button, a 'Feedback' link, and a footer with copyright information for Amazon Web Services, Inc. and links to 'Privacy', 'Terms', and 'Cookie preferences'.

Copy the data to table - dim_atm

```
COPY atm_data.dim_atm
```

```
FROM 's3://etl-atm-data/dim_atm/part-00000-13ec11a9-5a99-4e67-bf3d-77d36be09256-c000.csv'
```

```
IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
```

```
DELIMITER ','
```

```
REGION 'us-east-1'
```

```
CSV;
```


Copy the data to table - dim_atm

The screenshot displays the AWS Redshift console interface. On the left, the 'Resources' sidebar shows the database 'upgrad' and schema 'atm_data'. The main panel shows a SQL query for 'Query 1' with the following text:

```
89  
90  
91  
92  
93 COPY atm_data.dim_atm  
94 FROM 's3://etl-atm-data/dim_atm/part-00000-13ec11a9-5a99-4e67-bf3d-77d36be09256-c000.csv'  
95 IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'  
96 DELIMITER ','  
97 REGION 'us-east-1'  
98 CSV;  
99  
100  
101
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link. Below the buttons, the 'Query results' tab is active, showing 'Query 2441' with a status of 'Completed, started on June 10, 2024 at 21:23:13' and an 'ELAPSED TIME: 00 m 06 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'. The bottom of the console shows the 'CloudShell' tab and a footer with copyright information and links for 'Privacy', 'Terms', and 'Cookie preferences'.

Copy the data to table - dim_card_type

```
COPY atm_data.dim_card_type
```

```
FROM 's3://etl-atm-data/dim_card_type/part-00000-28105b18-516b-4f16-a32c-27e204b9d04c-c000.csv'
```

```
IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
```

```
DELIMITER ','
```

```
REGION 'us-east-1'
```

```
CSV;
```

Copy the data to table - dim_card_type

The screenshot displays the AWS CloudShell interface. The top navigation bar includes the AWS logo, 'Services', a search bar, and system information like 'N. Virginia' and a user email. The main workspace is divided into three sections:

- Resources Info:** On the left, it shows 'Select database' with 'upgrad' selected and 'Select schema' with 'public' selected. Below this is a 'Filter tables' search bar and a message stating 'No resources'.
- Query Editor:** The central area shows a SQL query for 'Query 1'. The query is:

```
110
111 COPY atm_data.dim_card_type
112 FROM 's3://etl-atm-data/dim_card_type/part-00000-28105b18-516b-4f16-a32c-27e204b9d04c-c000.csv'
113 IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
114 DELIMITER ','
115 REGION 'us-east-1'
116 CSV;
```
- Query Results:** At the bottom, it shows 'Query 2572' with a status of 'Completed, started on June 10, 2024 at 21:29:58' and 'ELAPSED TIME: 00 m 26 s'. It includes tabs for 'Execution', 'Data', and 'Visualize'.

The bottom of the interface features a 'CloudShell' label, a 'Feedback' link, and a copyright notice for Amazon Web Services, Inc. along with links for 'Privacy', 'Terms', and 'Cookie preferences'.

Copy the data to table - dim_date

```
COPY atm_data.dim_date
```

```
FROM 's3://etl-atm-data/dim_date/part-00000-89fb4560-dd8d-4f8a-a610-dd9bf1b054b0-c000.csv'
```

```
IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
```

```
DELIMITER ','
```

```
REGION 'us-east-1'
```

```
CSV
```

```
timeformat 'auto';
```

Copy the data to table - dim_date

The screenshot displays the AWS Redshift console interface. On the left sidebar, under 'Resources', the 'Select database' dropdown is set to 'upgrad' and the 'Select schema' dropdown is set to 'atm_data'. Below these, a list of tables is shown, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main panel shows a SQL query editor with the following code:

```
108
109
110
111
112
113 COPY atm_data.dim_date
114 FROM 's3://etl-atm-data/dim_date/part-00000-89fb4560-dd8d-4f8a-a610-dd9bf1b054b0-c000.csv'
115 IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
116 DELIMITER ','
117 REGION 'us-east-1'
118 CSV
119 timeformat 'auto';
120
```

Below the query editor, 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 query editor, the 'Query results' tab is active, showing 'Query 3193' with a status of 'Completed, started on June 10, 2024 at 22:14:22' and 'ELAPSED TIME: 00 m 11 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

At the bottom of the console, there is a 'CloudShell' button and a 'Feedback' link. The footer contains the copyright notice '© 2024, Amazon Web Services, Inc. or its affiliates.' and links for 'Privacy', 'Terms', and 'Cookie preferences'.

Copy the data to fact table - fact_atm_trans

```
COPY atm_data.fact_atm_trans
```

```
FROM 's3://etl-atm-data/fact_atm_trans/part-00000-2f09b28c-0ba5-493c-9c7b-e778ea20bae6-c000.csv'
```

```
IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
```

```
DELIMITER ','
```

```
REGION 'us-east-1'
```

```
CSV;
```

Copy the data to fact table - fact_atm_trans

The screenshot displays the Amazon Redshift Query Editor interface. The top navigation bar includes the AWS logo, a search bar, and the user's location (N. Virginia) and account ID. The main interface is divided into several sections:

- Resources Info:** A sidebar on the left showing the selected database (`upgrad`) and schema (`public`). It also includes a search bar for tables and a message stating "No resources to display".
- Query Editor:** The central area where the SQL query is written. The query is as follows:

```
121
122
123
124
125 COPY atm_data.fact_atm_trans
126 FROM 's3://etl-atm-data/fact_atm_trans/part-00000-2f09b28c-0ba5-493c-9c7b-e778ea20bae6-c000.csv'
127 IAM_ROLE 'arn:aws:iam::654654310080:role/myRedshiftRole'
128 DELIMITER ','
129 REGION 'us-east-1'
130 CSV;
```
- Query Results:** A section at the bottom showing the execution status of the query. It indicates that the query (ID 3248) is completed, started on June 10, 2024 at 22:16:29, and has an elapsed time of 00 m 12 s. There are buttons for "Execution", "Data", and "Visualize".

The interface also includes a "Run" button, a "Save" button, a "Schedule" button, and a "Clear" button. A "Send feedback" link is also present.