

## Creation of a Redshift Cluster

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

### Cluster Configuration

Amazon Redshift > Clusters > upgrad-etlproject

upgrad-etlproject ▾

Actions ▾ Edit Add partner integration Query data ▾

#### General information

Cluster identifier upgrad-etlproject	Status Available	Node type dc2.large	Endpoint upgrad-etlproject.ce0nzschaagg.us-east-1.redshift.a...
Cluster namespace 0e082ede-bd5f-44fe-a1cb-866f7da1a26e	Date created January 30, 2022, 15:47 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://upgrad-etlproject.ce0nzschaagg.us-e...
	Storage used 0.02% (0.06 of 320 GB used)	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=upgrad-etl...

Cluster performance Query monitoring Schedules Maintenance Properties

► Recommendations (0)  
To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

► Alarms (0)  
CloudWatch alarms are triggered when a metric threshold is met.

► Events (1)  
Amazon Redshift tracks events that occur on your cluster.

Cluster performance Query monitoring Schedules Maintenance Properties

#### Database configurations

Change admin user password Rotate encryption keys Edit ▾

Database name upgrad	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5440	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...	AWS KMS key ID -	
Admin user name awsuser			

#### Network and security settings

Edit

Virtual private cloud (VPC) vpc-0e26715bac6684eff	Availability Zone us-east-1b	VPC security group Specify which instances and devices can connect to the cluster. sg-00c00e16523040ee7 sg-007aef85271829933	Publicly accessible Allow instances and devices outside the VPC to connect to your database through the cluster endpoint. Disabled
Subnet upgrad-cluster-subnet-group-1	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.

Manage IAM roles

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.

Associated IAM roles (2) [Info](#)

Set default

Manage IAM roles

Q

Search for associated IAM role by name, status, or role type

<

1

>

<input type="checkbox"/>	<div><div>IAM roles</div><div></div></div>	<div><div>Status</div><div></div></div>	<div><div>Role type</div><div></div></div>
<input type="checkbox"/>	<a href="#">AWSServiceRoleForRedshift</a>	<div><div></div><div>in-sync</div></div>	--
<input type="checkbox"/>	<a href="#">upgrad_Redshift_s3</a>	<div><div></div><div>in-sync</div></div>	--

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

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:**

#### Location Table

```
create table DIM_LOCATION(  
location_id integer not null primary key,  
location varchar(50),  
streetname varchar(255),  
street_number integer,  
zipcode integer,  
lat decimal(10,3),  
lon decimal(10,3));
```

#### ATM Table

```
create table DIM_ATM(  
atm_id integer not null primary key,  
atm_number varchar(20),  
atm_manufacturer varchar(50),  
atm_location_id integer not null,  
foreign key(atm_location_id) references DIM_ATM(atm_id));
```

#### Date Table

```
create table DIM_DATE(  
date_id integer not null primary key,  
full_date_time timestamp,  
year integer,  
month varchar(20),  
day integer,  
hour integer,  
weekday varchar(20));
```

#### Card type table

```
create table DIM_CARD_TYPE(  
card_type_id integer not null primary key,  
card_type varchar(23));
```

#### Transaction Fact Table

```
create table FACT_ATM_TRANS(  
trans_id BIGINT not null primary key,  
atm_id integer,  
weather_loc_id integer,  
date_id VARCHAR(10),  
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 DECIMAL(10,3),
clouds_all DECIMAL(10,3),
weather_id integer,
weather_main VARCHAR(50),
weather_description VARCHAR(255),
foreign key(atm_id) references DIM_ATM(atm_id),
foreign key(date_id) references DIM_DATE(date_id),
foreign key(card_type_id) references DIM_CARD_TYPE(card_type_id),
foreign key(weather_loc_id) references DIM_LOCATION(location_id));
```

Amazon Redshift > Query editor

Editor | **Query history** | Saved queries | Scheduled queries

Queries and loads (7) [Info](#)

[Open in editor](#)

< 1 > ⚙

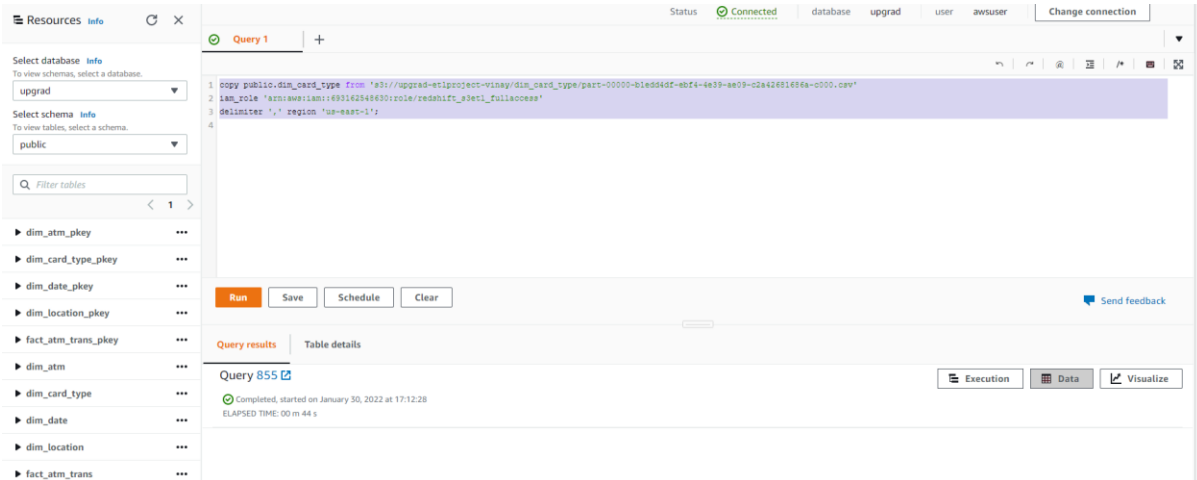
<input type="checkbox"/>	Start time	Status	Duration	SQL
<input type="checkbox"/>	Jan 30th, 2022 04:12:38 PM 6 minutes ago	Finished	555 ms	create table FACT_ATM_TRANS( trans_id BIGINT not null primary key, atm_id integer not null, weather_
<input type="checkbox"/>	Jan 30th, 2022 04:12:21 PM 7 minutes ago	Finished	560 ms	create table DIM_CARD_TYPE( card_type_id integer not null primary key, card_type varchar(23));
<input type="checkbox"/>	Jan 30th, 2022 04:12:06 PM 7 minutes ago	Finished	562 ms	create table DIM_DATE( date_id integer not null primary key, full_date_time timestamp, year integer,
<input type="checkbox"/>	Jan 30th, 2022 04:11:32 PM 8 minutes ago	Finished	2 sec	create table DIM_ATM( atm_id integer not null primary key, atm_number varchar(20), atm_manufacturer
<input type="checkbox"/>	Jan 30th, 2022 04:10:07 PM 9 minutes ago	Failed	486 ms	create table DIM_LOCATION( location_id integer not null primary key, location varchar(50), streetnam
<input type="checkbox"/>	Jan 30th, 2022 04:07:59 PM 11 minutes ago	Finished	557 ms	create table DIM_LOCATION( location_id integer not null primary key, location varchar(50), streetnam
<input type="checkbox"/>	Jan 30th, 2022 04:06:35 PM 13 minutes ago	Finished	577 ms	create schema redshift_etl;

## 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

#### Card\_type

```
copy public.dim_card_type from 's3://upgrad-etlproject-vinay/dim_card_type/part-00000-b1edd4df-ebf4-4e39-ae09-
c2a42681686a-c000.csv'
iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
delimiter ',' region 'us-east-1';
```



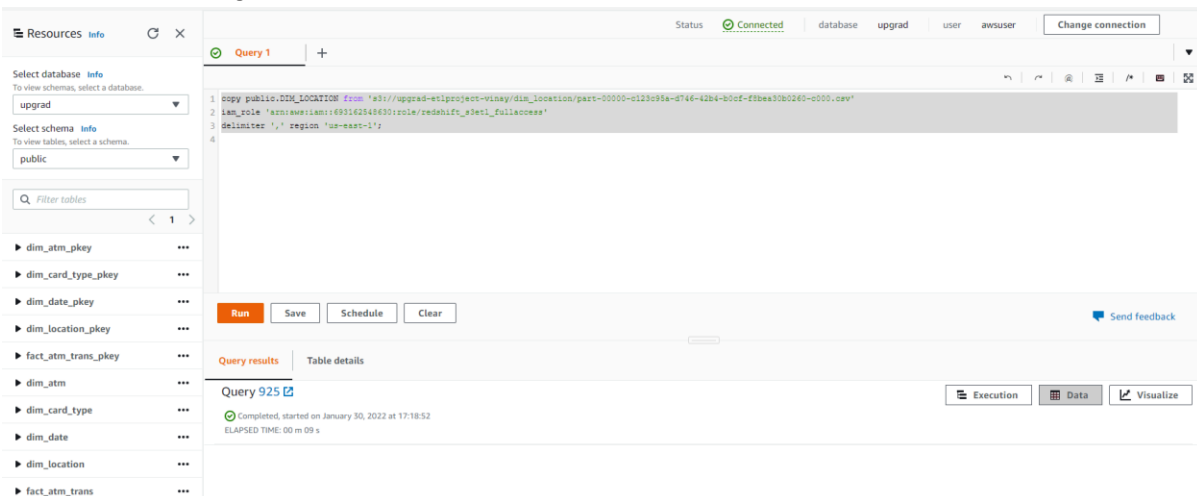
The screenshot shows the upGrad Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there's a 'Select database' dropdown set to 'upgrad' and a 'Select schema' dropdown set to 'public'. A list of tables is shown, including 'dim\_atm\_pkey', 'dim\_card\_type\_pkey', '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 area displays a SQL query for 'Query 1':

```
1 copy public.dim_card_type from 's3://upgrad-etlproject-vinay/dim_card_type/part-00000-b1e994df-ebf4-ee39-ae09-c2a426f1e96a-c000.csv'
2 iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
3 delimiter ',' region 'us-east-1';
4
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there's a 'Query results' tab showing 'Query 855' with a status of 'Completed, started on January 30, 2022 at 17:12:28' and 'ELAPSED TIME: 00 m 44 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## Location

```
copy public.DIM_LOCATION from 's3://upgrad-etlproject-vinay/dim_location/part-00000-c123c95a-d746-42b4-b0cf-f8bea30b0260-c000.csv'
iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
delimiter ',' region 'us-east-1';
```



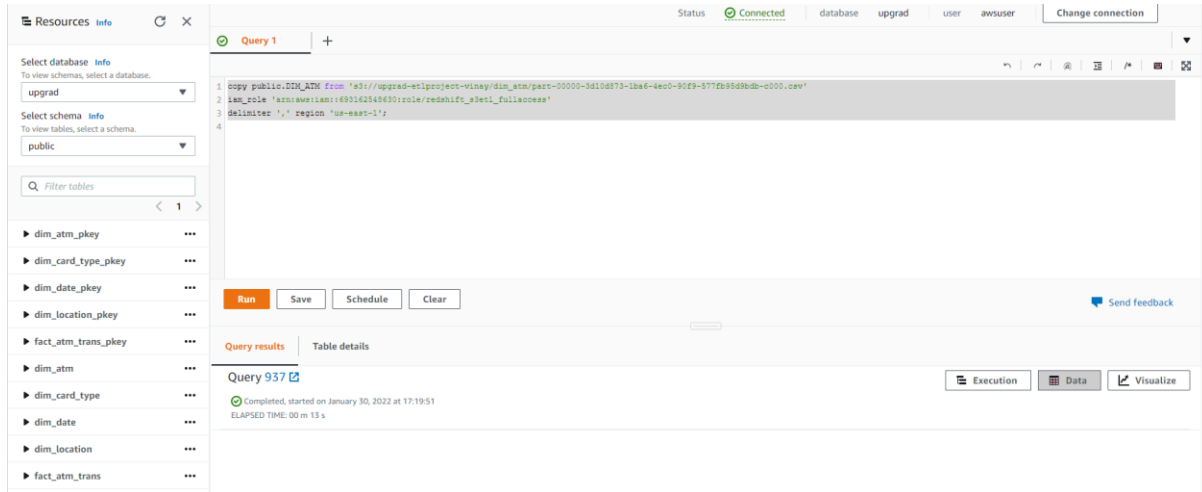
The screenshot shows the upGrad Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there's a 'Select database' dropdown set to 'upgrad' and a 'Select schema' dropdown set to 'public'. A list of tables is shown, including 'dim\_atm\_pkey', 'dim\_card\_type\_pkey', '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 area displays a SQL query for 'Query 1':

```
1 copy public.DIM_LOCATION from 's3://upgrad-etlproject-vinay/dim_location/part-00000-c123c95a-d746-42b4-b0cf-f8bea30b0260-c000.csv'
2 iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
3 delimiter ',' region 'us-east-1';
4
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there's a 'Query results' tab showing 'Query 925' with a status of 'Completed, started on January 30, 2022 at 17:18:52' and 'ELAPSED TIME: 00 m 09 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## ATM

```
copy public.DIM_ATM from 's3://upgrad-etlproject-vinay/dim_atm/part-00000-3d10d873-1ba6-4ec0-90f9-577fb95d9bdb-c000.csv'
iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
delimiter ',' region 'us-east-1';
```



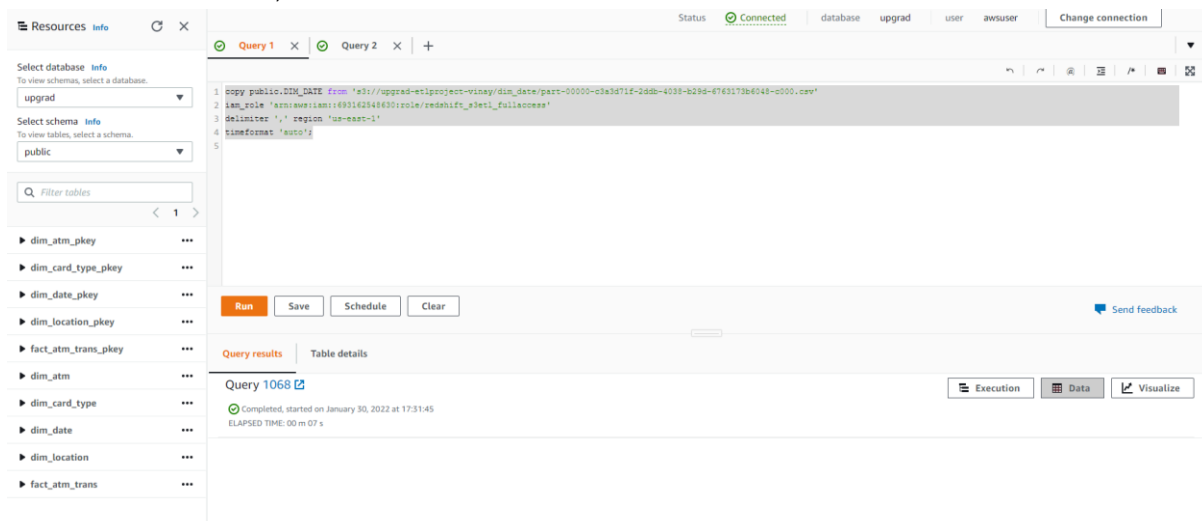
The screenshot shows the upGrad Redshift console interface. On the left, there's a 'Resources' sidebar with a tree view of databases and schemas. The main area displays 'Query 1' with the following SQL code:

```
1 copy public.DIM_ATM from 's3://upgrad-etlproject-vinay/dim_atm/part-00000-3d10d873-1ba6-4ec0-90f9-577f96d9b0b-c000.csv'
2 iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
3 delimiter ',' region 'us-east-1';
4
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 937' with a status of 'Completed, started on January 30, 2022 at 17:19:51' and an 'ELAPSED TIME: 00 m 13 s'.

## Date

```
copy public.DIM_DATE from 's3://upgrad-etlproject-vinay/dim_date/part-00000-c3a3d71f-2ddb-4038-b29d-
6763173b6048-c000.csv'
iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
delimiter ',' region 'us-east-1'
timeformat 'auto';
```



The screenshot shows the upGrad Redshift console interface. On the left, there's a 'Resources' sidebar with a tree view of databases and schemas. The main area displays 'Query 2' with the following SQL code:

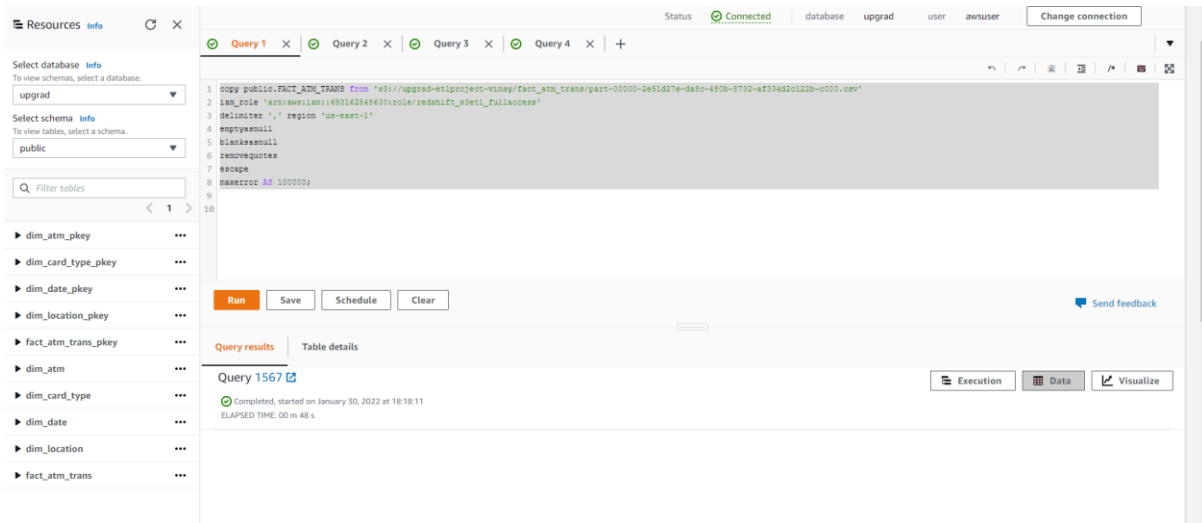
```
1 copy public.DIM_DATE from 's3://upgrad-etlproject-vinay/dim_date/part-00000-c3a3d71f-2ddb-4038-b29d-6763173b6048-c000.csv'
2 iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
3 delimiter ',' region 'us-east-1'
4 timeformat 'auto';
5
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 1068' with a status of 'Completed, started on January 30, 2022 at 17:31:45' and an 'ELAPSED TIME: 00 m 07 s'.

## ATM\_TRANS

```
copy public.FACT_ATM_TRANS from 's3://upgrad-etlproject-vinay/fact_atm_trans/part-00000-2e51d27e-da8c-480b-
8732-af334d2c122b-c000.csv'
iam_role 'arn:aws:iam::693162548630:role/redshift_s3etl_fullaccess'
delimiter ',' region 'us-east-1'
```

```
emptyasnull
blanksasnull
removequotes
escape
maxerror AS 100000;
```



The screenshot shows the upGrad SQL editor interface. On the left, there's a sidebar with 'Resources' and 'Select database' (upgrad) and 'Select schema' (public). The main area displays a SQL query with a syntax error. The query is:

```
1 copy public.FACT_ATM_TRANS from 's3://upgrad-es1project-vinay/fact_atm_trans/part-00000-2e51d27e-da50-480b-b702-af394d2c122b-c000.csv'
2 into_role 'ath:aws:iam:693142b49630:role/redshift_ath1_fullaccess'
3 delimited ',' region 'us-east-1'
4 emptyasnull
5 blanksasnull
6 removequotes
7 escape
8 maxerror AS 100000;
```

The query execution result shows 'Query 1567' with a status of 'Completed, started on January 30, 2022 at 18:18:11' and 'ELAPSED TIME: 00 m 48 s'. The error message 'maxerror AS 100000;' is highlighted in red.