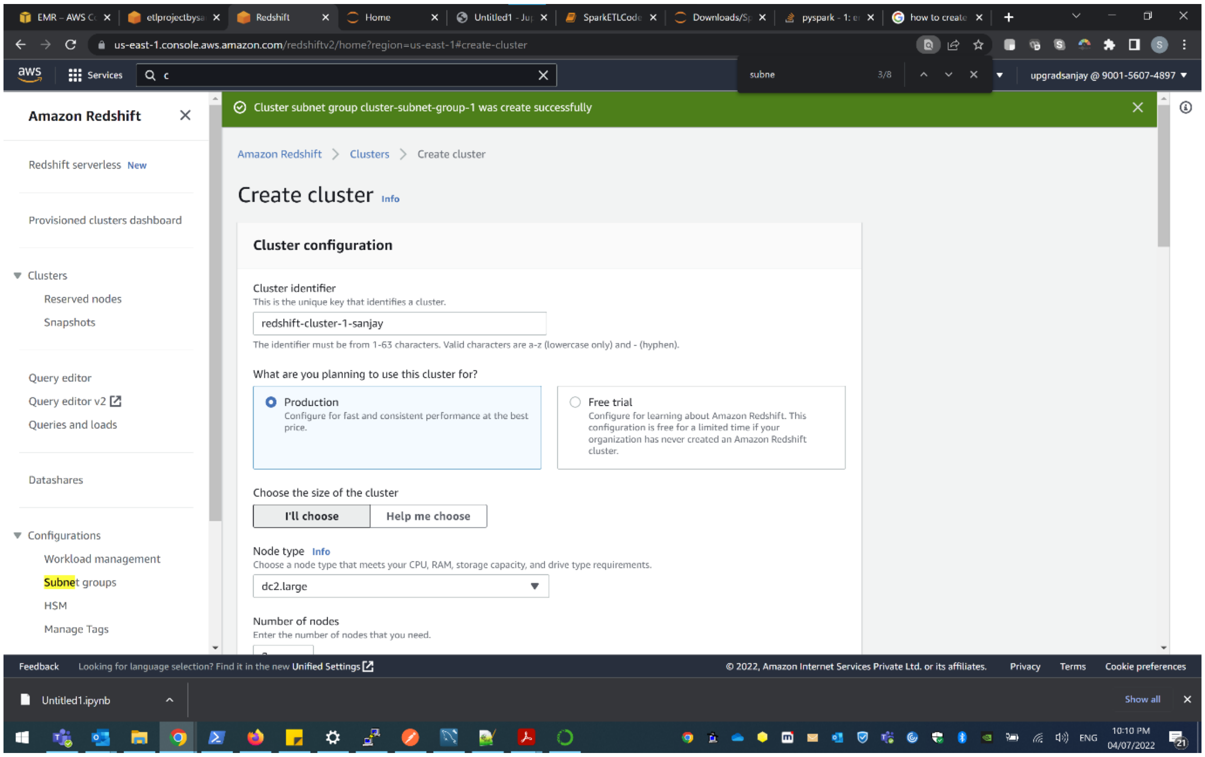
Creation of a Redshift Cluster

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



us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#create-cluster

Cluster subnet group cluster-subnet-group-1 was create successfully

Amazon Redshift > Clusters > Create cluster

Create cluster

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.
redshift-cluster-1-sanjay
The 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?

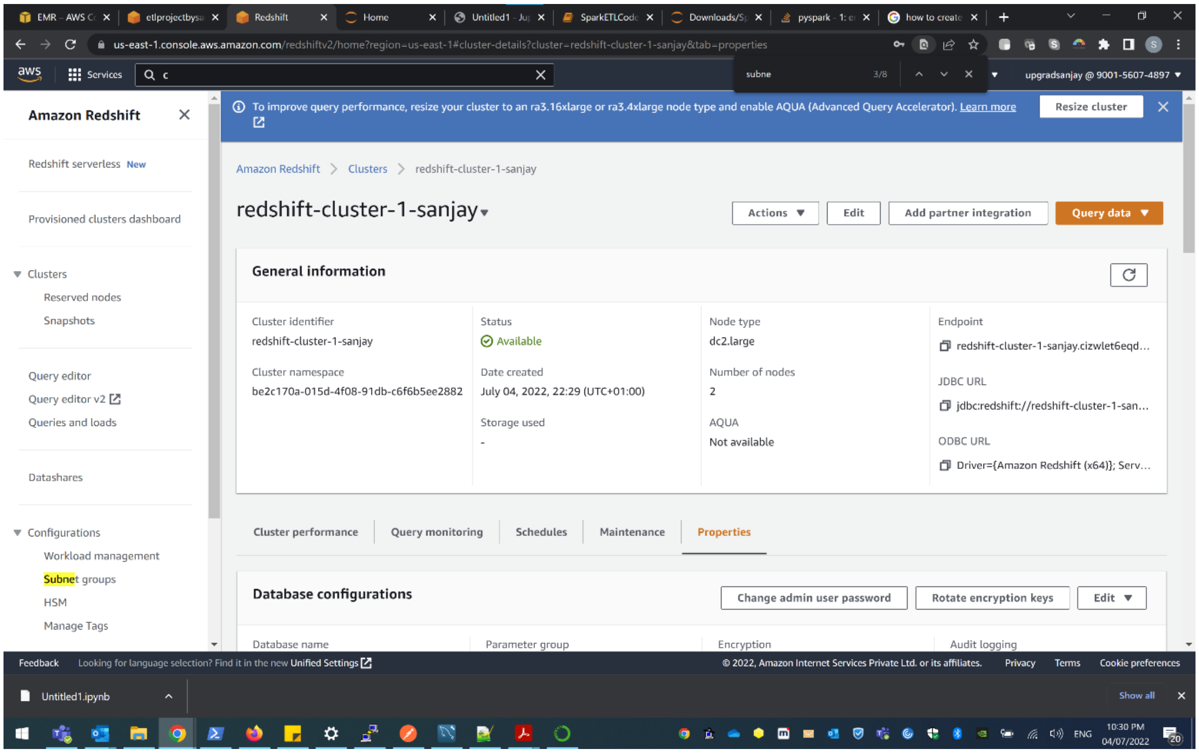
● Production
Configure for fast and consistent performance at the best price.

○ 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
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.
dc2.large

Number of nodes
Enter the number of nodes that you need.



us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#cluster-details?cluster=redshift-cluster-1-sanjay&tab=properties

To improve query performance, resize your cluster to an ra3.16xlarge or ra3.4xlarge node type and enable AQUA (Advanced Query Accelerator). Learn more

Amazon Redshift > Clusters > redshift-cluster-1-sanjay

redshift-cluster-1-sanjay

General information

Cluster identifier
redshift-cluster-1-sanjay

Cluster namespace
be2c170a-015d-4f08-91db-c6f6b5ee2882

Status
Available

Date created
July 04, 2022, 22:29 (UTC+01:00)

Storage used
-

Node type
dc2.large

Number of nodes
2

AQUA
Not available

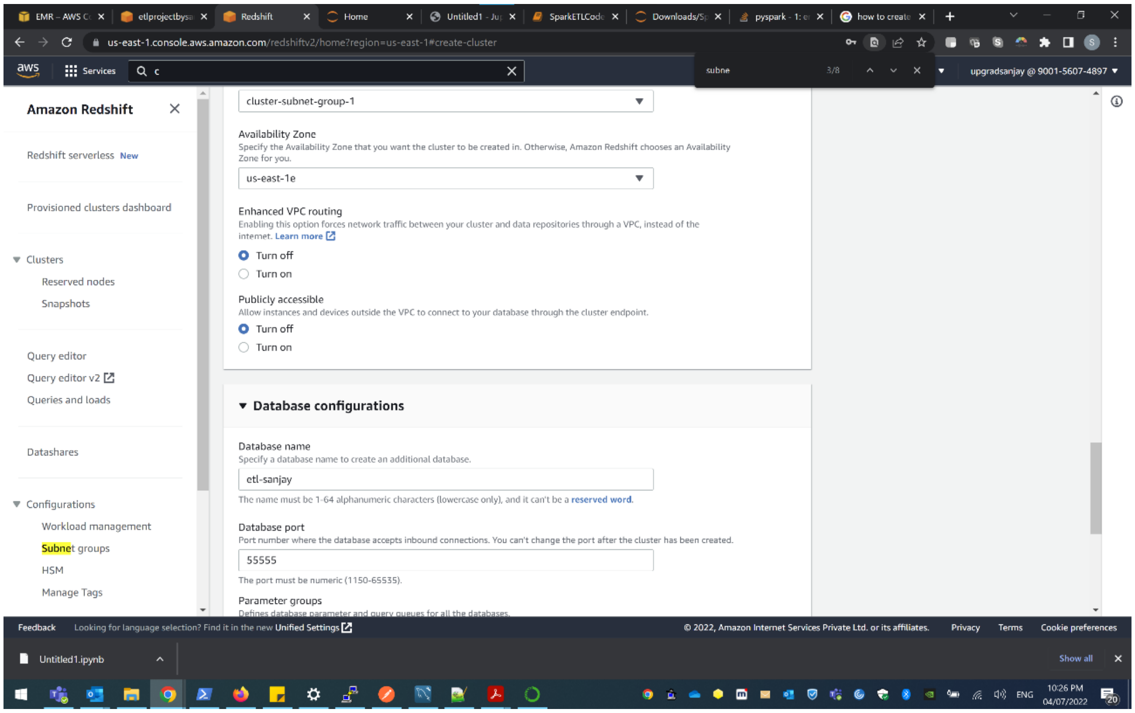
Endpoint
redshift-cluster-1-sanjay.cizwlet6eqd...

JDBC URL
jdbc:redshift://redshift-cluster-1-san...

ODBC URL
Driver={Amazon Redshift (x64)}; Serv...

Database configurations

Database name | Parameter group | Encryption | Audit logging



cluster-subnet-group-1

Availability Zone
Specify the Availability Zone that you want the cluster to be created in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

us-east-1e

Enhanced VPC routing
Enabling this option forces network traffic between your cluster and data repositories through a VPC, instead of the internet. Learn more

● Turn off
○ Turn on

Publicly accessible
Allow instances and devices outside the VPC to connect to your database through the cluster endpoint.

● Turn off
○ Turn on

Database configurations

Database name
Specify a database name to create an additional database.

etl-sanjay

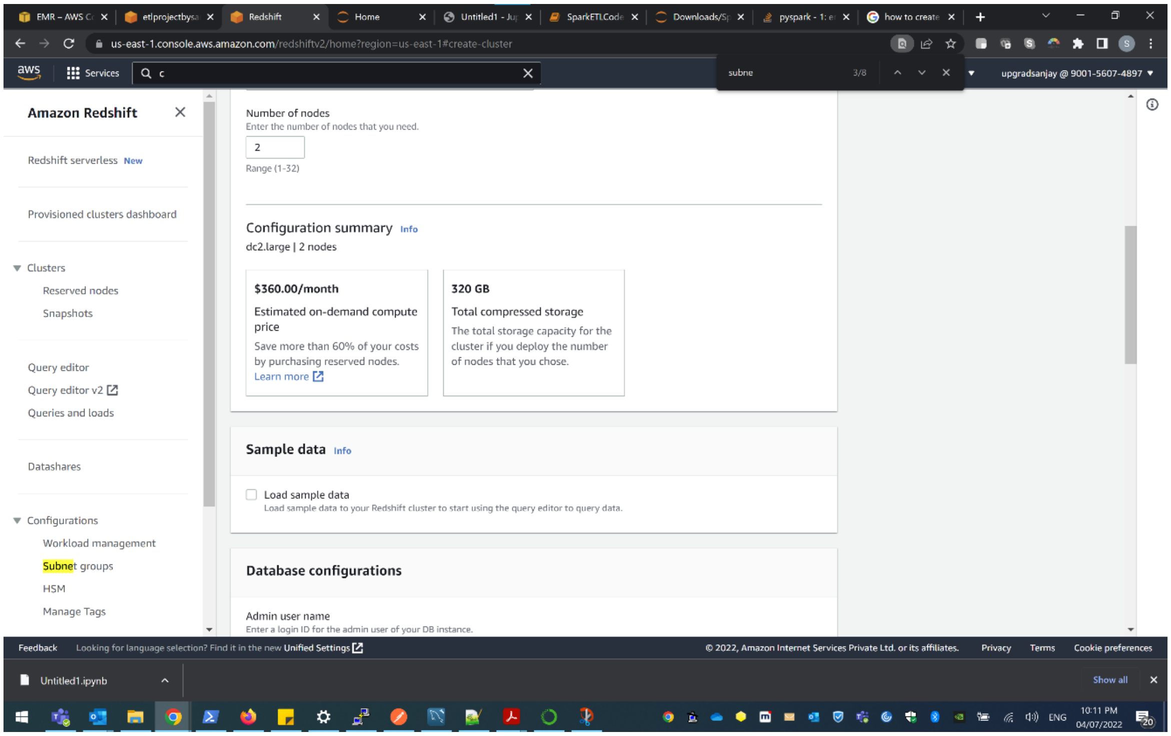
The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word.

Database port
Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.

55555

The port must be numeric (1150-65535).

Parameter groups
Defines database parameter and query queues for all the databases.



Number of nodes
Enter the number of nodes that you need.

2

Range (1-32)

Configuration summary Info
dc2.large | 2 nodes

$360.00/month
Estimated on-demand compute price
Save more than 60% of your costs by purchasing reserved nodes.
Learn more

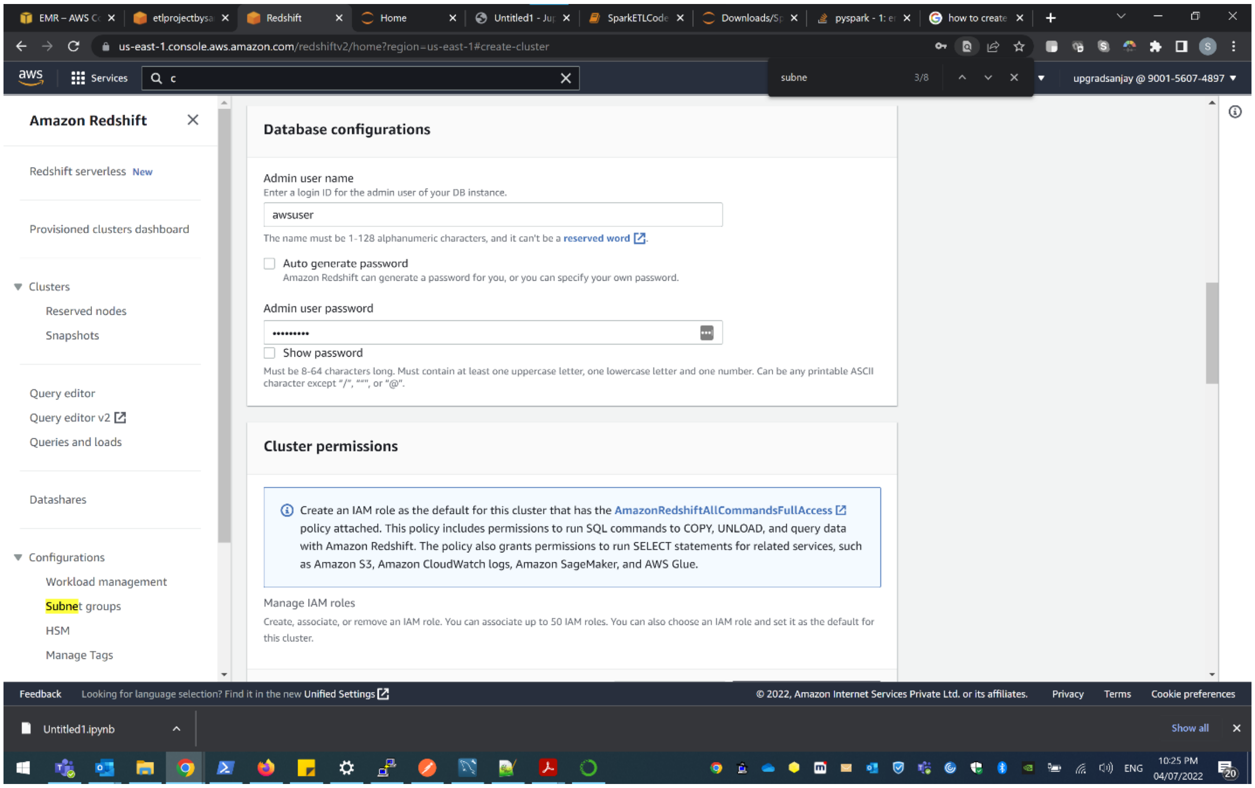
320 GB
Total compressed storage
The total storage capacity for the cluster if you deploy the number of nodes that you chose.

Sample data Info

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

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.



Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a reserved word.

□ Auto generate password
Amazon Redshift can generate a password for you, or you can specify your own password.

Admin user password

••••••••••

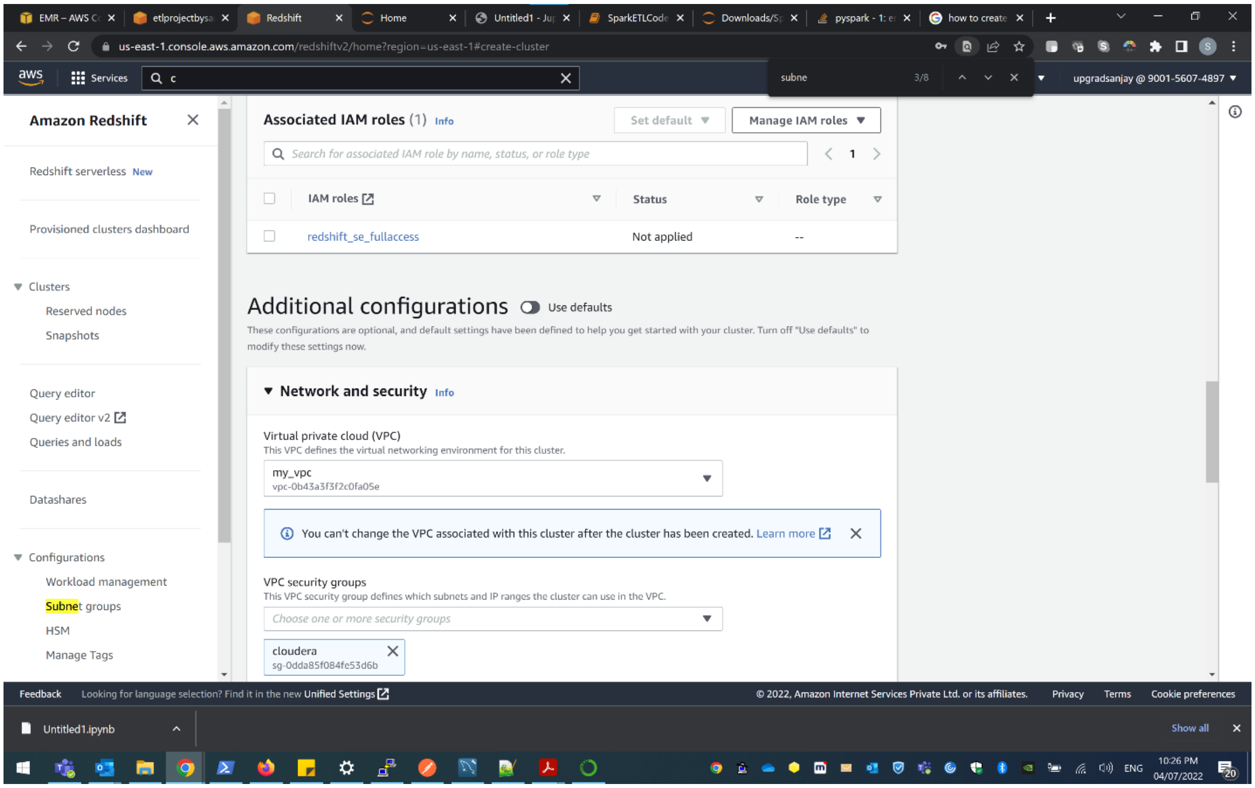
□ Show password

Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except "/", """, or "@".

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 (1) Info

□	IAM roles	Status	Role type
□	redshift_se_fullaccess	Not applied	--

Additional configurations

○ Use defaults

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 Info

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.

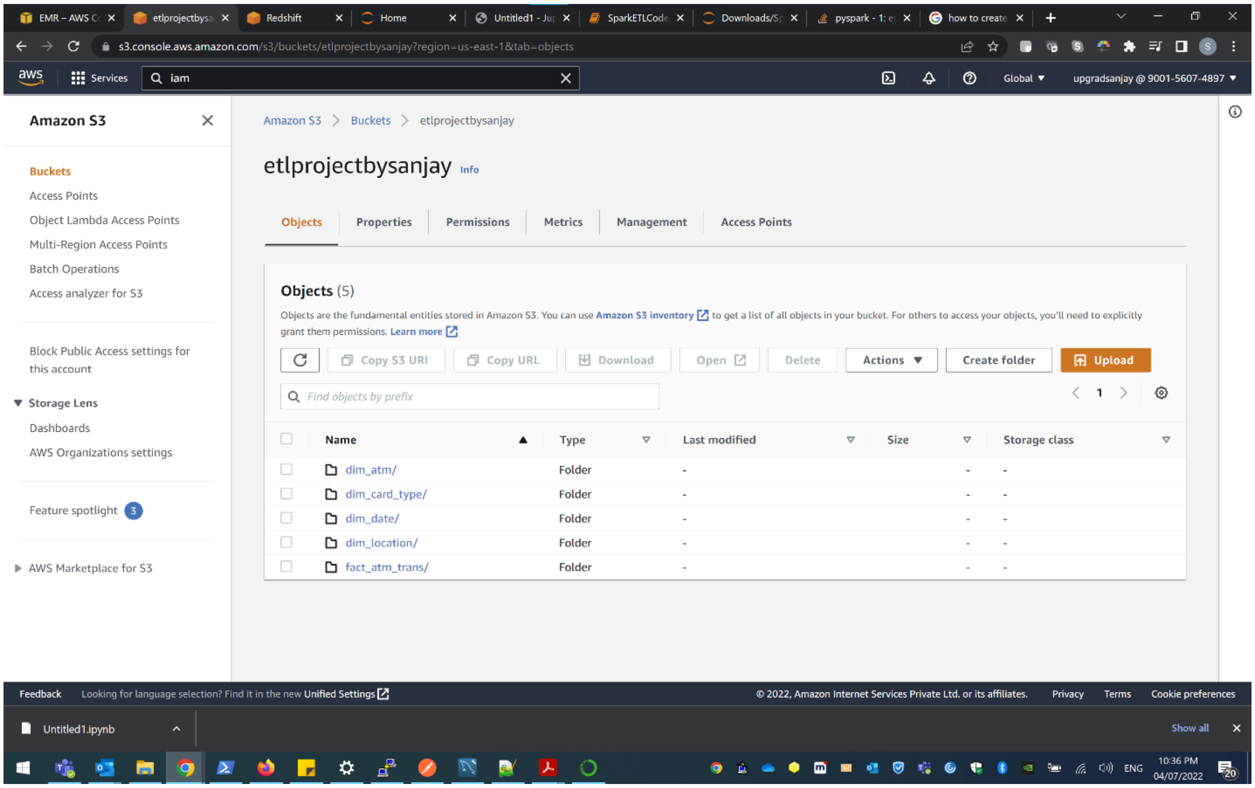
my_vpc
vpc-0b43a3f3f2c0fa05e

You can't change the VPC associated with this cluster after the cluster has been created. Learn more

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

cloudera
sg-0dda85f084fe53d6b


s3.console.aws.amazon.com/s3/buckets/etlprojectbysanjay?region=us-east-1&tab=objects
Services
iam
Global
upgradsanjay @ 9001-5607-4897
Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
AWS Marketplace for S3
Amazon S3 > Buckets > etlprojectbysanjay
etlprojectbysanjay Info
Objects
Properties
Permissions
Metrics
Management
Access Points
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences
Untitled1.ipynb
Show all
10:36 PM
04/07/2022

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

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

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:

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)
);
```

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)
);
```

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

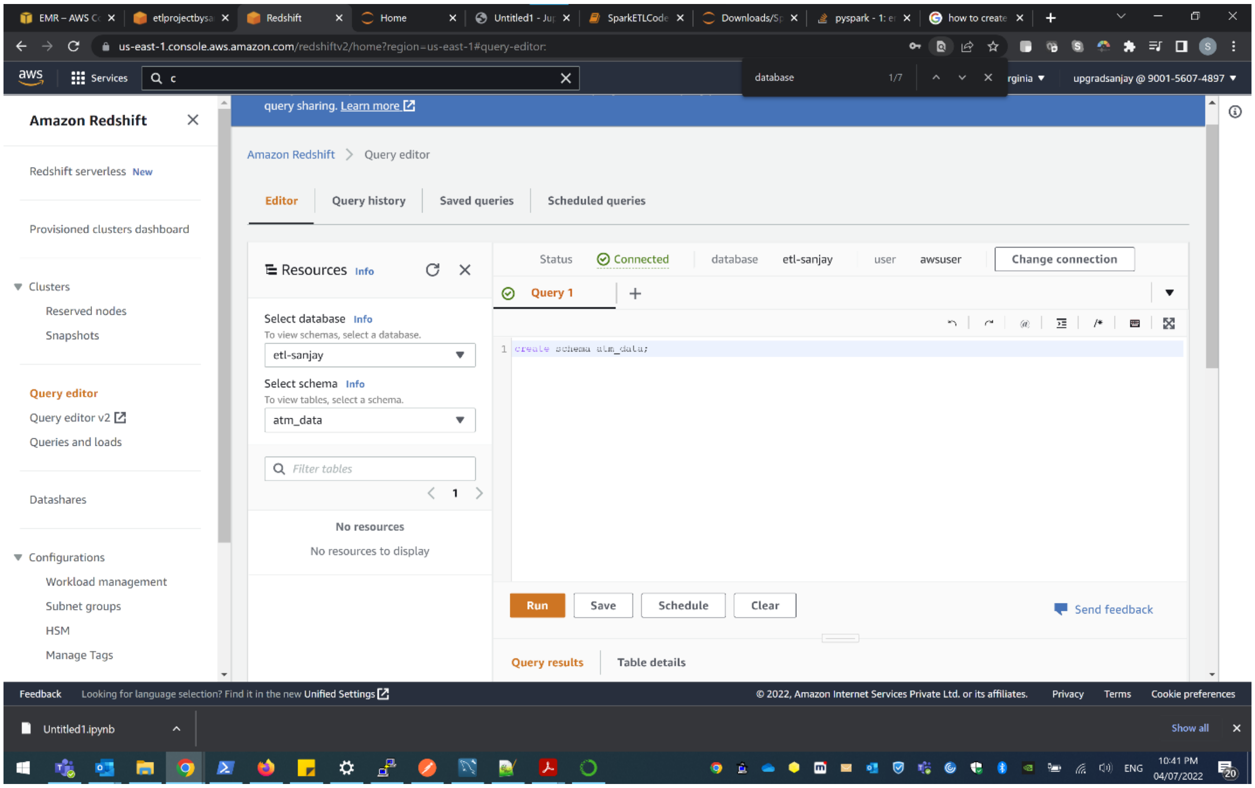
```
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30) ,
PRIMARY KEY(card_type_id)
);
```

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_DATA(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)
);
```

Screenshots:



Amazon Redshift > Query editor

Status: Connected database: etl-sanjay user: awsuser

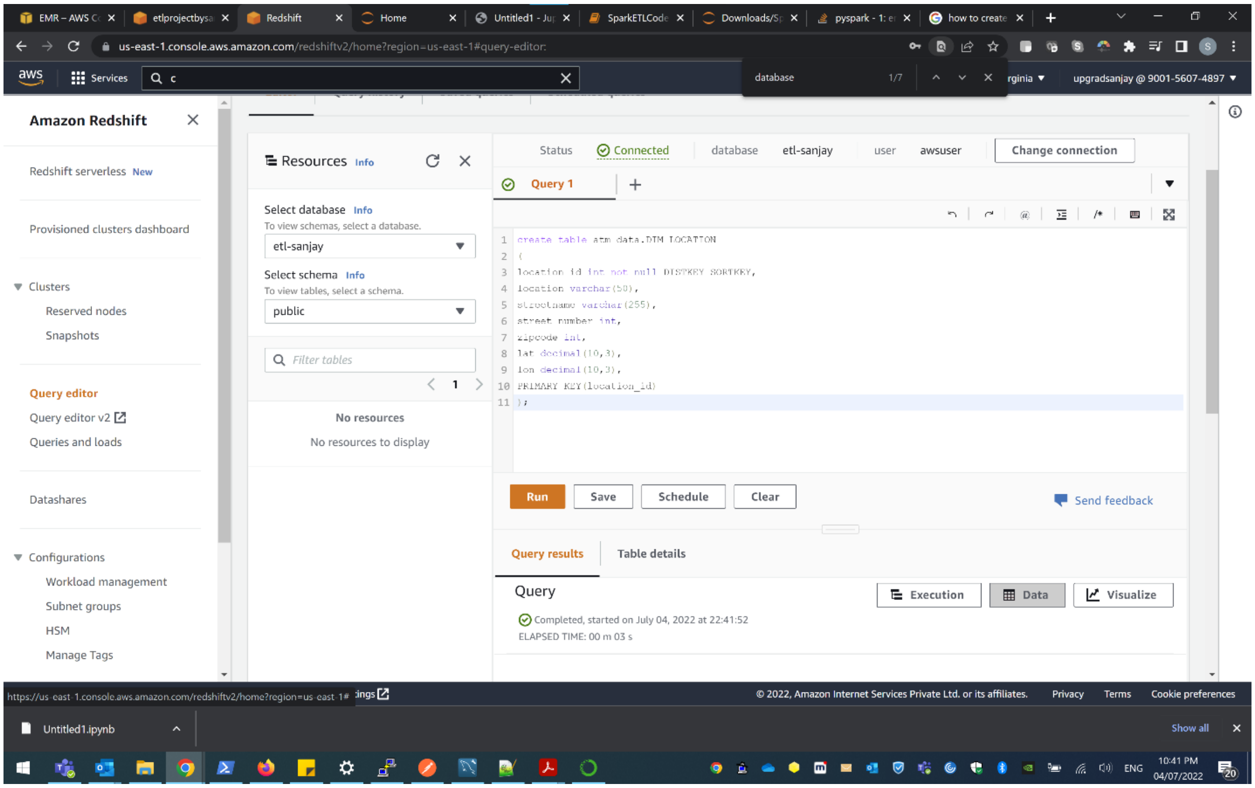
Select database: etl-sanjay

Select schema: atm_data

No resources

No resources to display

```
create schema atm_data;
```



Status: Connected database: etl-sanjay user: awsuser

Select database: etl-sanjay

Select schema: public

No resources

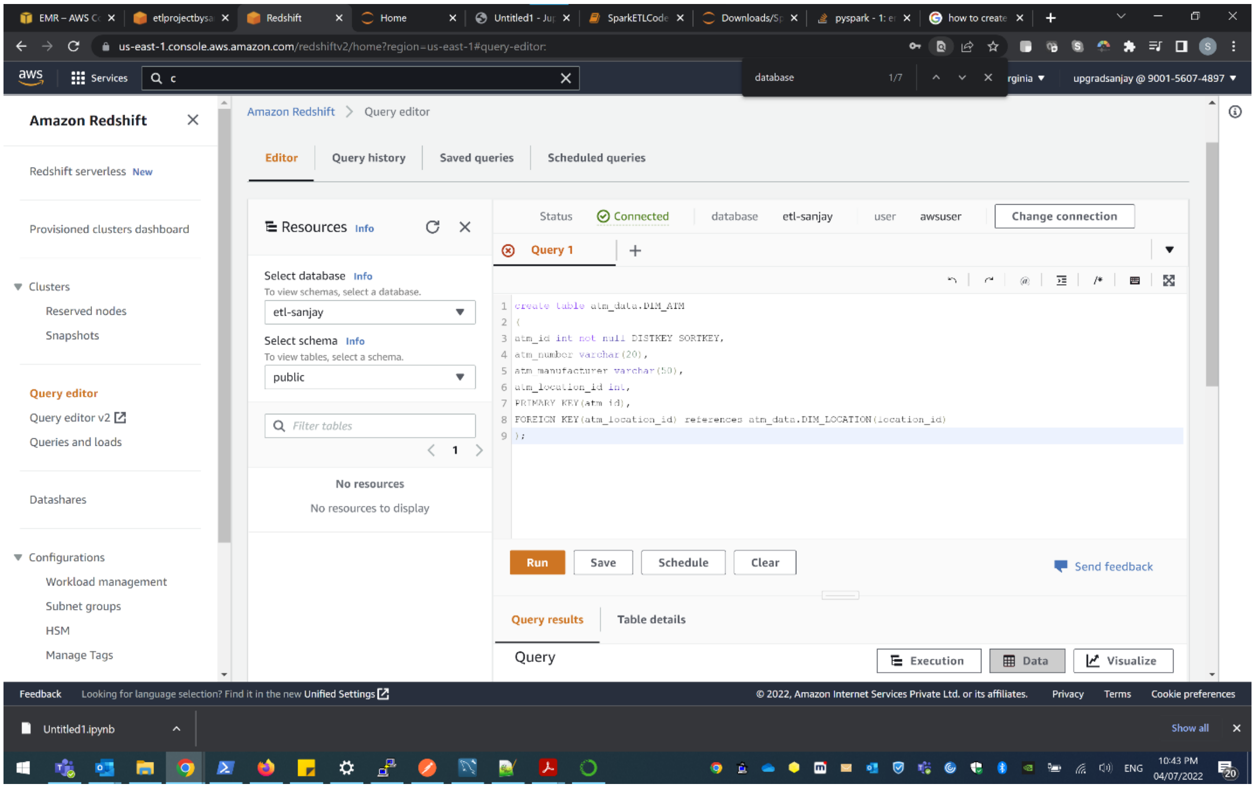
No resources to display

```
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)
);
```

Query

Completed, started on July 04, 2022 at 22:41:52

ELAPSED TIME: 00 m 03 s

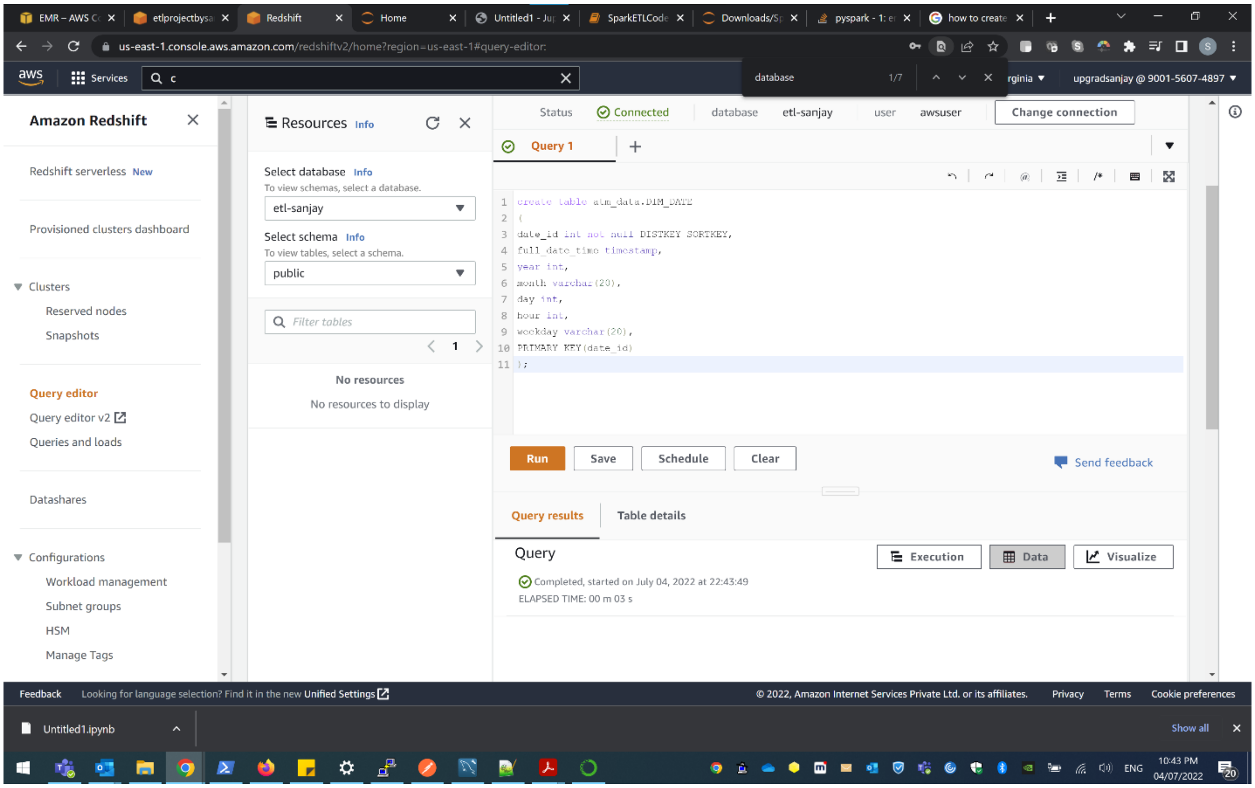


Status: Connected | database: etl-sanjay | user: awsuser | Change connection

Select database: etl-sanjay

Select schema: public

```
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)
);
```



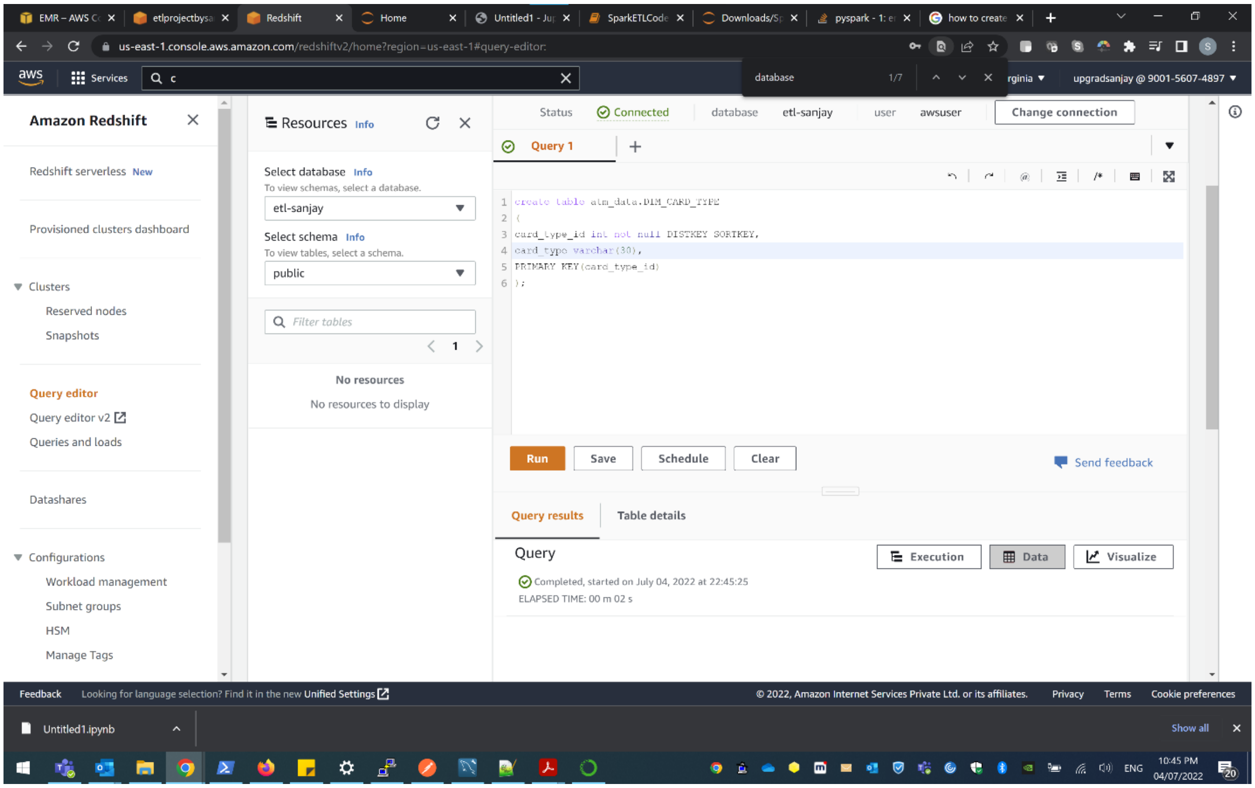
Status: Connected | database: etl-sanjay | user: awsuser | Change connection

Select database: etl-sanjay

Select schema: public

```
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)
);
```

Completed, started on July 04, 2022 at 22:43:49
ELAPSED TIME: 00 m 03 s

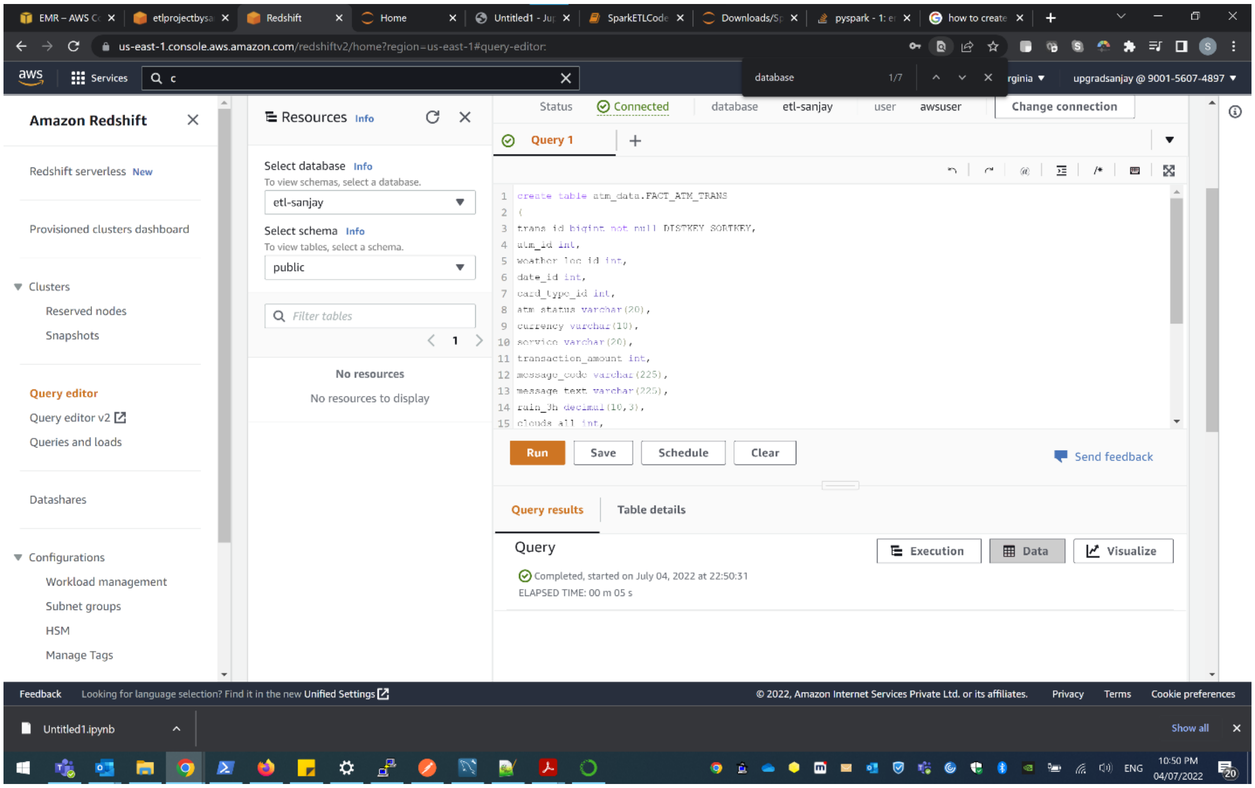


Status: Connected   database: etl-sanjay   user: awsuser

```
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);
```

Query

Completed, started on July 04, 2022 at 22:45:25
ELAPSED TIME: 00 m 02 s



Status: Connected   database: etl-sanjay   user: awsuser

```
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,
```

Query

Completed, started on July 04, 2022 at 22:50:31
ELAPSED TIME: 00 m 05 s

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

Copying the data to dim_location table

```
copy atm_data.dim_location from 's3://etlprojectbysanjay/dim_location/part-00000-4f4b02d0-919a-442e-9134-f459cbdb7909-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Copying the data to dim_atm table

```
copy atm_data.dim_atm from ' s3://etlprojectbysanjay/dim_atm/part-00000-c4425605-e626-4cd2-adb2-cef68f7cb1b9-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Copying the data to dim_date table

```
copy atm_data.dim_date from ' s3://etlprojectbysanjay/dim_date/part-00000-7a7ef505-bc12-476c-a0a6-e9e8b544fe44-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV
timeformat 'YYYY-MM-DDTHH:MI:SS';
```

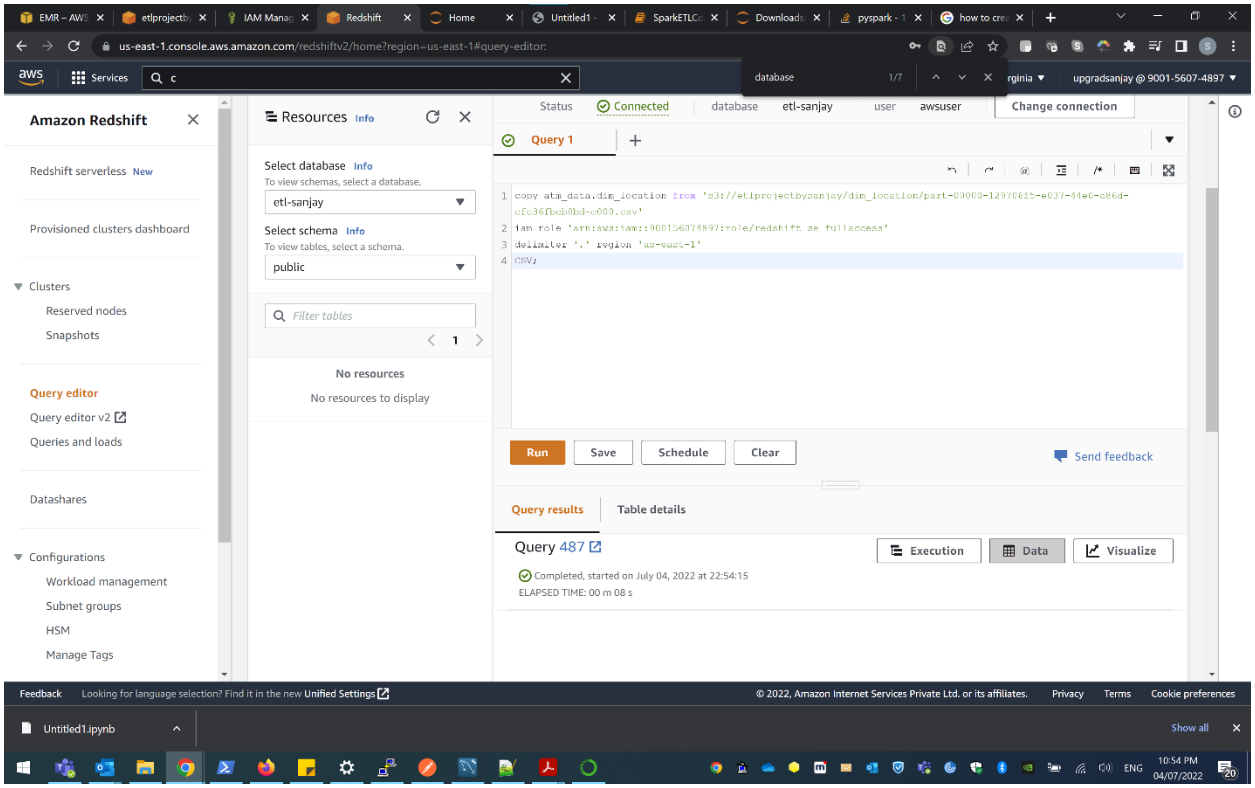
Copying the data to dim_card_type table

```
copy atm_data.dim_card_type from ' s3://etlprojectbysanjay/dim_card_type/part-00000-b9c7eb07-29c6-4445-ba0f-98de14834601-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Copying the data to fact_atm_trans table

```
copy atm_data.fact_atm_trans from ' s3://etlprojectbysanjay/fact_atm_trans/part-00000-978dd709-2ef2-4145-8ab5-9981558a8c60-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Screenshots



```
copy atm_data.dim_location from 's3://etlprojectbysanjay/dim_location/part-00000-129706f5-e037-44e0-a86d-cfc36fbcb0bd-c000.csv'
iam_role 'arn:aws:iam::900156074897:role/redshift_se_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Status: Connected | database: etl-sanjay | user: awsuser

Select database: etl-sanjay

Select schema: public

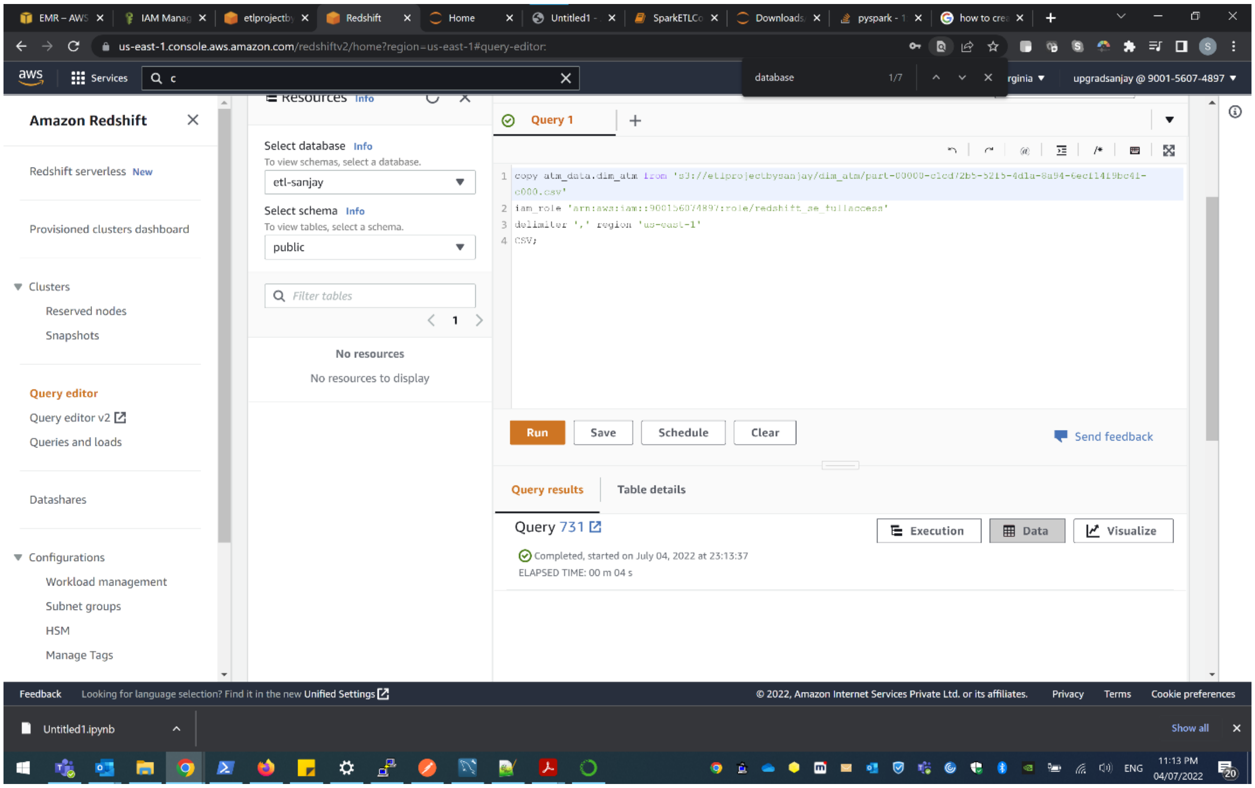
No resources

No resources to display

Query 487

Completed, started on July 04, 2022 at 22:54:15

ELAPSED TIME: 00 m 08 s



```
copy atm_data.dim_atm from 's3://etlprojectbysanjay/dim_atm/part-00000-c1cd72b5-5215-4d1a-8a94-6ec1419bc41-c000.csv'
iam_role 'arn:aws:iam::900156074897:role/redshift_se_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Select database: etl-sanjay

Select schema: public

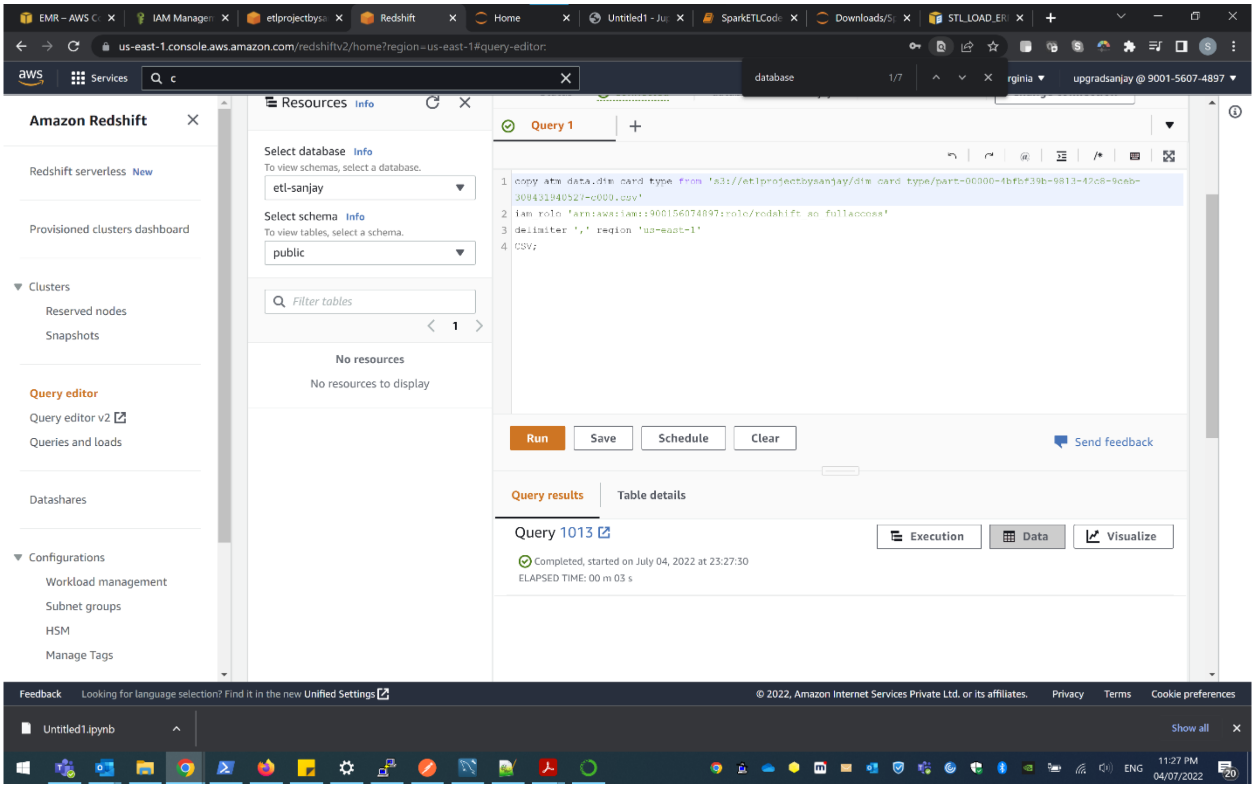
No resources

No resources to display

Query 731

Completed, started on July 04, 2022 at 23:13:37

ELAPSED TIME: 00 m 04 s

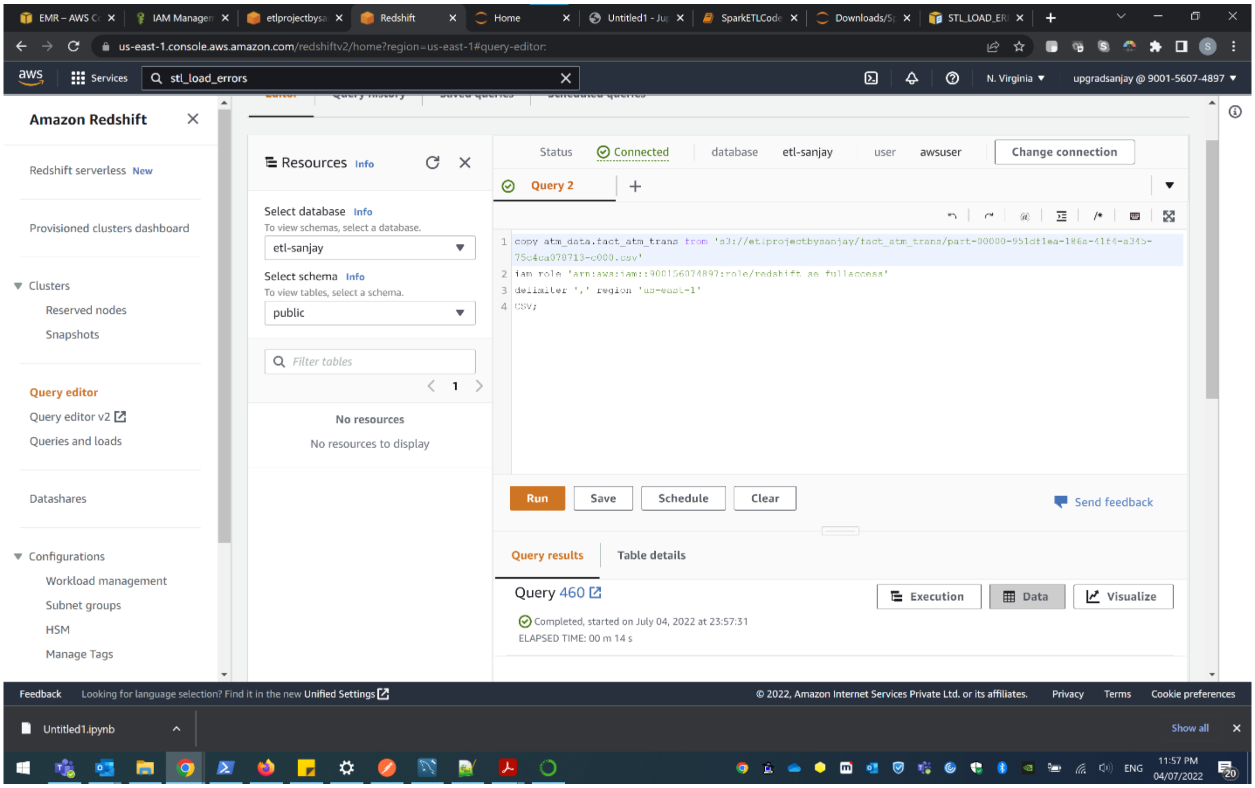


Query 1

```
copy atm_data.dim_card_type from 's3://etlprojectbysanjay/dim_card_type/part-00000-4bfbf39b-9813-42c8-9ceb-308431940527-c000.csv'
iam_role 'arn:aws:iam::900156074897:role/redshift_se_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 1013

Completed, started on July 04, 2022 at 23:27:30
ELAPSED TIME: 00 m 03 s



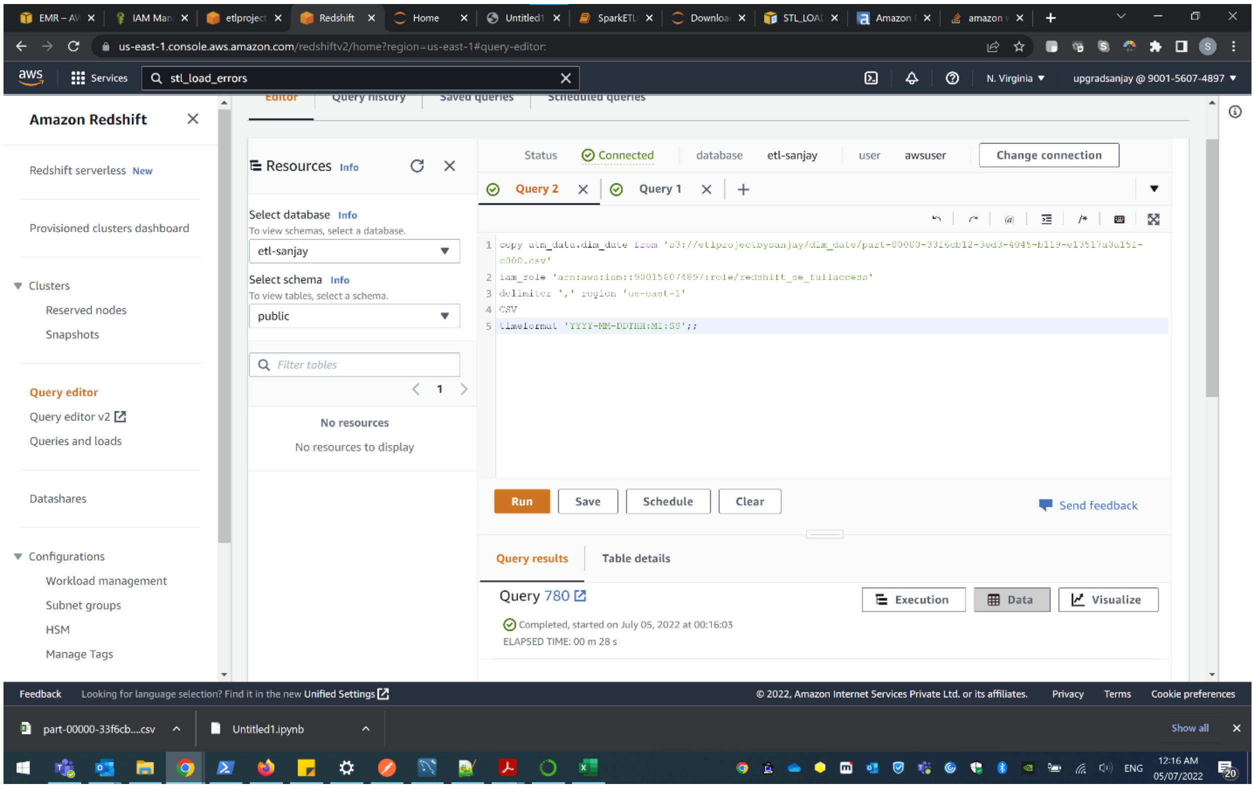
Status Connected database etl-sanjay user awsuser

Query 2

```
copy atm_data.fact_atm_trans from 's3://etlprojectbysanjay/fact_atm_trans/part-00000-951df1ea-186a-41f4-a345-75c4ca070713-c000.csv'
iam_role 'arn:aws:iam::900156074897:role/redshift_se_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 460

Completed, started on July 04, 2022 at 23:57:31
ELAPSED TIME: 00 m 14 s



```
copy atm_data.dim_date from 's3://etlprojectbysanjay/dim_date/part-00000-33f6cb12-3ed3-4045-b119-e13517a0a15f-c000.csv'
iam_role 'arn:aws:iam::900156074897:role/redshift_se_fullaccess'
delimiter ',' region 'us-east-1'
CSV
timeformat 'YYYY-MM-DDTHH:MI:SS';;
```

Query 780

Completed, started on July 05, 2022 at 00:16:03

ELAPSED TIME: 00 m 28 s