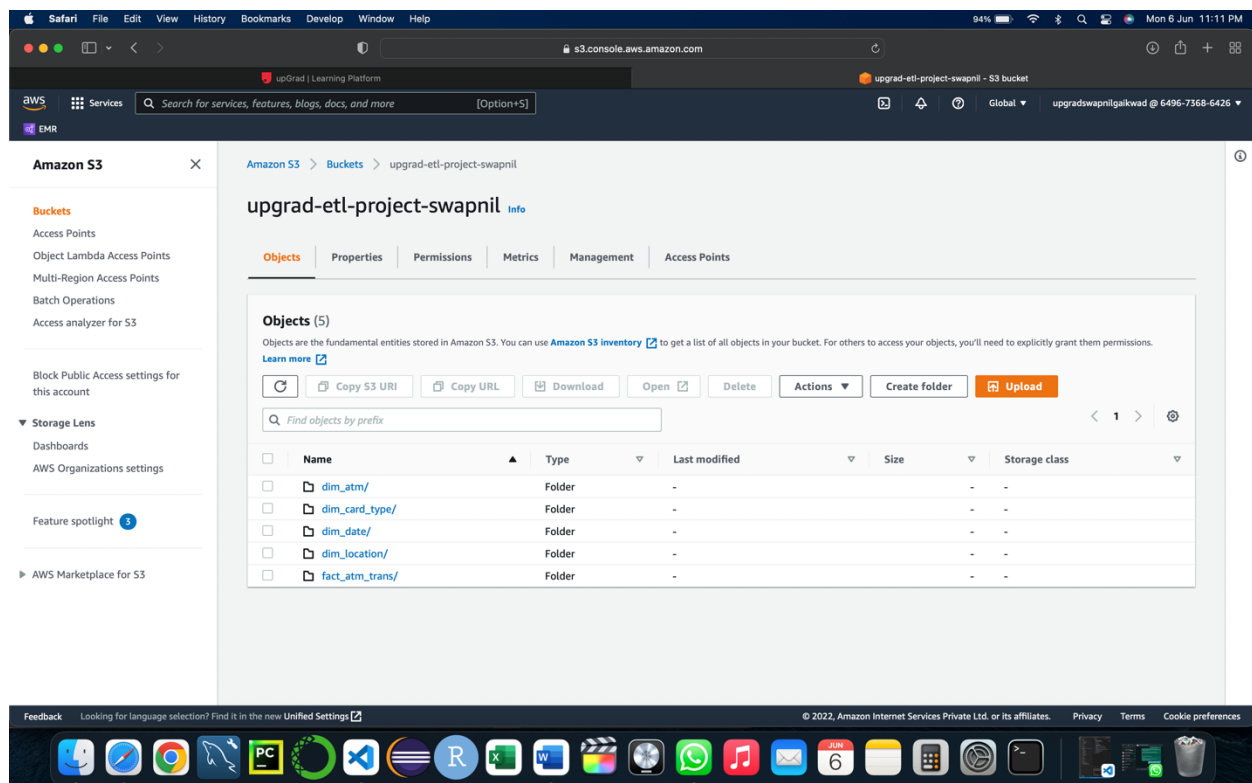
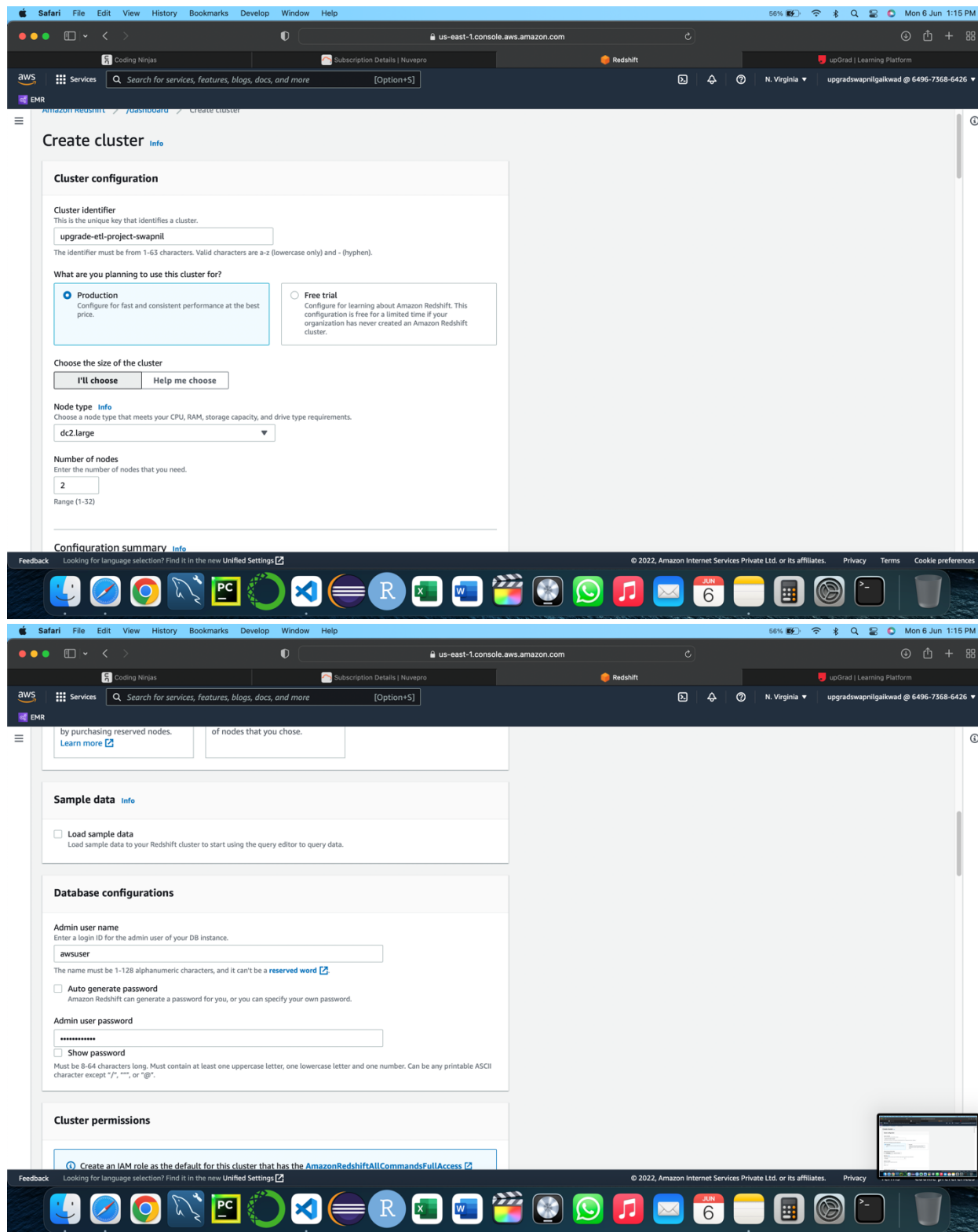


Creation of a Redshift Cluster

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





Create cluster [Info](#)

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

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

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

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

Range (1-32)

Configuration summary [Info](#)

Looking for language selection? Find it in the new Unified Settings [\[Option+5\]](#)

by purchasing reserved nodes. [Learn more](#)

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.

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

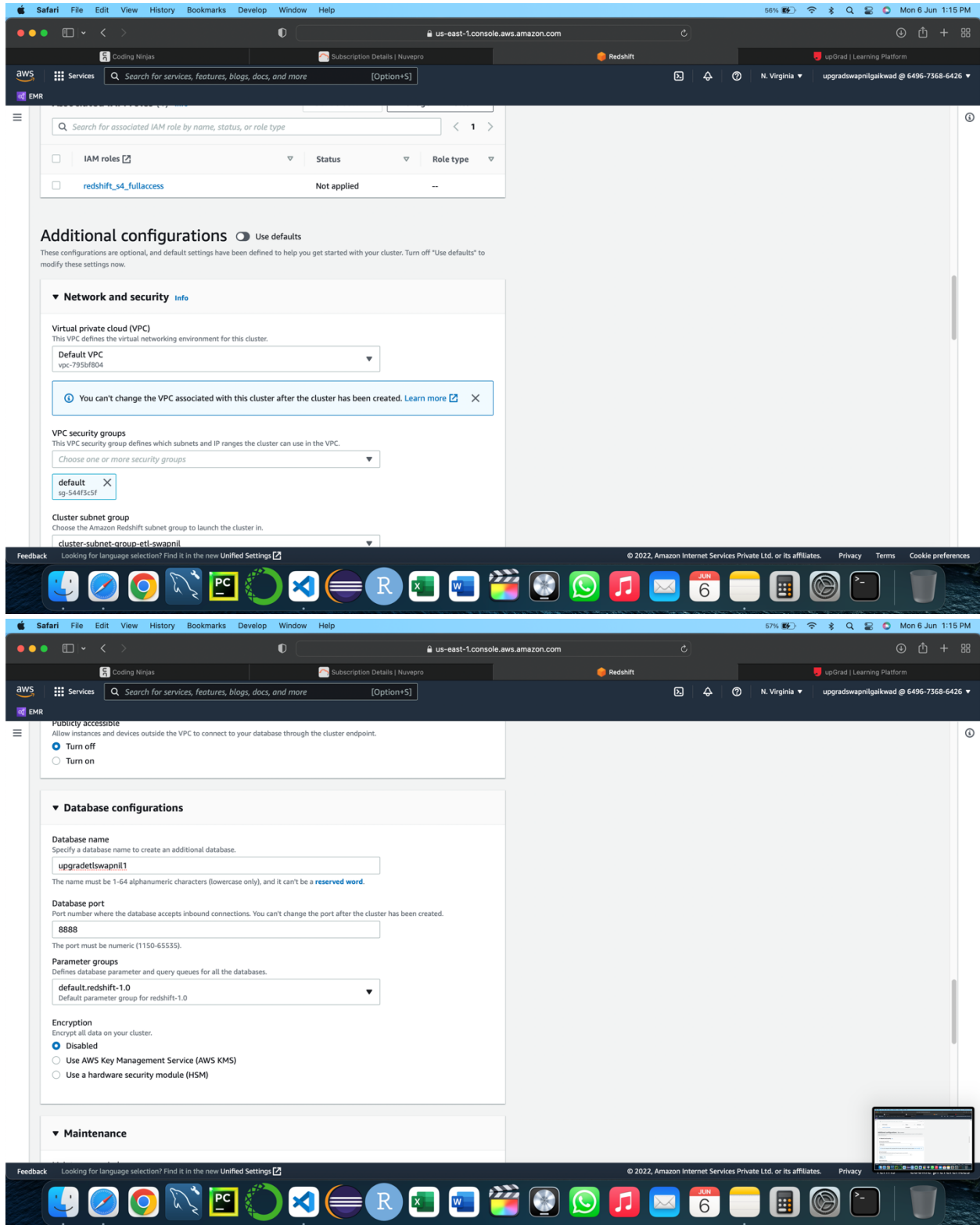
☐ **Auto generate password**
Amazon Redshift can generate a password for you, or you can specify your own password.

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](#)



The image shows a sequence of two screenshots from the AWS Redshift console, illustrating the configuration of a new Redshift cluster. The browser is Safari on a Mac, and the user is logged in as 'upgradwapnilgaikwad' in the 'us-east-1' region.

Top Screenshot: Additional configurations

- Network and security:**
 - Virtual private cloud (VPC):** Set to 'Default VPC' (vpc-795bf804). A warning states: "You can't change the VPC associated with this cluster after the cluster has been created."
 - VPC security groups:** Set to 'default' (sg-544f3c3f).
 - Cluster subnet group:** Set to 'cluster-subnet-group-etl-swapnil'.

Bottom Screenshot: Database configurations

- Publicly accessible:** Set to 'Turn off'.
- Database configurations:**
 - Database name:** 'upgradetlswapnil1'. Note: "The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word."
 - Database port:** '8888'. Note: "The port must be numeric (1150-65535)."
 - Parameter groups:** 'default.redshift-1.0'.
 - Encryption:** Set to 'Disabled'.
- Maintenance:** (Section header visible, details not fully shown).

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:

1. Query to create a schema for the deimension an fact tables :

```
create schema atm_data ;
```

2. Query to create the 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)
);
```

3. 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)
);
```

- 4.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)  
);
```

5.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)  
);
```

6.Creating atm Transaction 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),  
cloud_all int ,  
weather_id int ,  
weather_main varchar(50),  
weather_description varchar(225),  
PRIMARY KEYS(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)  
);
```

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

Following are the queries for loading the data from S3 to REDSHIFT :

1. copying data from s3 to dim_location table

```
copy atm_data.dim_location from s3://upgrad-etl-project-swapnil/dim_location/part-00000-e87d47c0-c8e1-4f8e-b880-54db72073c5c-c000.csv'  
iam_role'arn:aws:iam::649673686426:role/redshift_s4_fullaccess'  
delimiter','region'us-east-1'  
CSV;
```

3. copying data from s3 to dim_atm table

```
copy atm_data.dim_atm from 's3://upgrad-etl-project-swapnil/dim_atm/part-00000-9857a98e-1f27-4945-b1a0-01a135d4a2fa-c000.csv'  
iam_role'arn:aws:iam::649673686426:role/redshift_s4_fullaccess'  
delimiter','region'us-east-1'  
CSV;
```

4. copying data from s3 to dim_date table

```
copy atm_data.dim_location from 's3://upgrad-etl-project-swapnil/dim_date/part-00000-b9044941-f5e4-45cd-a9a3-18d4407645b5-c000.csv'  
iam_role'arn:aws:iam::649673686426:role/redshift_s4_fullaccess'  
delimiter','region'us-east-1'  
CSV;
```

5. Copying data from s3 to dim_card type

```
copy atm_data.dim_card from s3://upgrad-etl-project-swapnil/dim_card_type/part-00000-0870eee5-794c-4421-bc76-7580d5a1aaec-c000.csv'  
iam_role'arn:aws:iam::649673686426:role/redshift_s4_fullaccess'  
delimiter','region'us-east-1'  
CSV;
```

6. Copying data from s3 to fact_atm_trans table

```
copy atm_data.dim_location from 's3://upgrad-etl-project-swapnil/dim_location/part-00000-747fb822-b657-4bd5-8cfb-3ec4717a1070-c000.csv'  
iam_role'arn:aws:iam::649673686426:role/redshift_s4_fullaccess'  
delimiter','region'us-east-1'  
CSV;
```