**E-Commerce Analytics**

DESCRIPTION

You work for an e-commerce company as a big data consultant. Your job entails analyzing sales data. The company operates at several locations around the world. They want you to analyze the data from their sales transactions on a daily and weekly basis and want you to derive significant insights to understand their sales in various cities and states. You've also been asked to include a variety of other details (that are provided below) about the product evaluation.

**Description:**

 Use Spark features for data analysis to derive valuable insights

**Domain:**E-commerce

**Analysis to be done:**Exploratory analysis to determine actionable insights

**nsights on Historical Data**

**1. Daily insights**

**a. Sales**

* Total sales
* Total sales in each city
* Total sales in each state

1. **Orders**

* Total number of orders
* City-wise order distribution
* State-wise order distribution
* Average review score per order
* Average freight charges per order
* Average time taken to approve the orders (order approved –

order purchased)

* Average order delivery time

**2.       Weekly insights**

**a. Sales**

* Total sales
* Total sales in each city
* Total sales in each state

**b. Orders**

* Total number of orders
* City-wise order distribution
* State-wise order distribution
* Average review score per order
* Average freight charges per order
* Average time taken to approve the orders (order approved – order purchased)
* Average order delivery time

**c . Total freight charges**

**d. Distribution of freight charges in each city**

**Tasks to perform:**

**Week 1: Overview and basic configurations**

**Step 1:**Choose a suitable cloud provider and set up a Spark shell environment

**I choose AWS cloud services to develop and run my project .**

**Step 2:** Configure the necessary dependencies

**Create Key Pair to login to ec2 using Putty**

Graphical user interface, text, application, email

Description automatically generated

Then create EMR Cluster with needed configurations

Graphical user interface

Description automatically generated

Select configuration of cluster as image and press next

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Now use our PK KEY and click create cluster

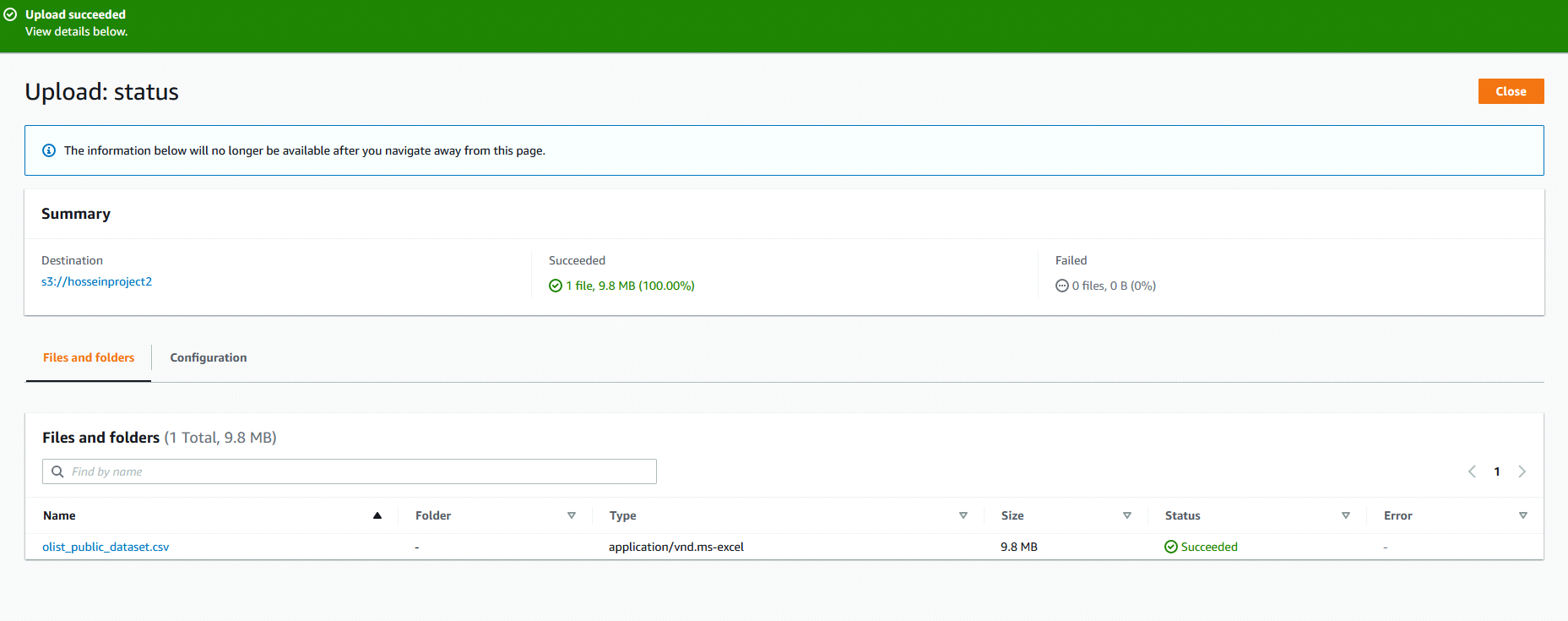
Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Then create s3 bucket and Upload data file



After that we will open Hue to connect Hive to create needed tables for insights

Graphical user interface, text, application

Description automatically generated

Load data from the csv file s3 to Hive Table

use Ecommercy\_capstone ;

load data INPATH 's3://ecomerrcy/' into table ecommercy\_olist;

Graphical user interface, text, application, email

Description automatically generated

Convert “order\_purchase\_timestamp” to week and day date

Graphical user interface, text

Description automatically generated

Then to make our insights we need to create table for each query

Sales by cities

Graphical user interface, application

Description automatically generated

Sales table per state

Graphical user interface, application

Description automatically generated

**Weekly Sales and Orders**

We’re going to do the calculation on another time frame which is a week.

* 1. Weekly orders by city

Graphical user interface

Description automatically generated with medium confidence

* 1. Weekly orders by state

Graphical user interface, table

Description automatically generated with medium confidence

1. Average review score, average freight value, average order approval, and delivery time

Graphical user interface, text, application

Description automatically generated

1. The freight charges per city and total freight charges

Graphical user interface, text

Description automatically generated with medium confidence

**Week 4: Data analysis and visualization**

* 1. Write data to HDFS with format parquet

insert overwrite directory 'hdfs://adhamfcihgmail@ip-10-0-41-79/user/hadoop/' STORED AS PARQUET select \* from city\_avg\_frreight;

insert overwrite directory 'hdfs://adhamfcihgmail@ip-10-0-41-79/user/hadoop/' STORED AS PARQUET select \* from weeklySalesOrdersByState;

insert overwrite directory 'hdfs://adhamfcihgmail@ip-10-0-41-79/user/hadoop/' STORED AS PARQUET select \* from weeklySalesOrdersByCity;

insert overwrite directory 'hdfs://adhamfcihgmail@ip-10-0-41-79/user/hadoop/' STORED AS PARQUET select \* from DailySalesOrdersByState;

insert overwrite directory 'hdfs://adhamfcihgmail@ip-10-0-41-79/user/hadoop/' STORED AS PARQUET select \* from DailySalesOrdersByCity;

then create S3 Buckets with the name of tables and load data with above query.

**Step 2:** Save the final dataset into object storage service per the cloud platform

another way to load data directly load data from hive to AWS S3 Buckets

INSERT OVERWRITE DIRECTORY 's3n://bucket/directory/' select \* from city\_avg\_frreight;

INSERT OVERWRITE DIRECTORY 's3n://bucket/directory/' select \* from weeklySalesOrdersByState;

INSERT OVERWRITE DIRECTORY 's3n://bucket/directory/' select \* from weeklySalesOrdersByCity;

INSERT OVERWRITE DIRECTORY 's3n://bucket/directory/' select \* from DailySalesOrdersByState;

INSERT OVERWRITE DIRECTORY 's3n://bucket/directory/' select \* from DailySalesOrdersByCity;

**Step 3:** Create a DB cluster that is also a NoSQL using the relevant service on the

             cloud platform

Create a DynamoDB tables with the same name of Buckets name or hive tables,

In the search box, write DynamoDB, and press the DynamoDB service Select Tables from the left side menu, and press Create table

Graphical user interface, application

Description automatically generated

Create tables with below details .

Graphical user interface, application

Description automatically generated

**Create IAM role** In the search box, write IAM, and select IAM service From the left side menu, select Roles, and press Create role

Step 1: Select Lambda from Use case and press Next

Step 2: Select these permissions (AmazonS3ReadOnlyAccess, AmazonDynamoDBFullAccess, CloudWatchLogsFullAccess), and press Next

Step 3: Enter a Role name “import\_data\_to\_dynamodb” and press Create role

**Create Lambda Function to Import data into DynamoDB**

In the search box, write Lambda, and select Lambda service From the left side menu, select Functions, and press Create function

1. Enter a name to the Lambda function “Import\_data\_to\_dynamodb”

2. Select the function language “Python 3.9”

3. Select Architecture “x86\_64” 4. Select Use an existing role

5. In Existing role, Select the IAM role “Import\_data\_to\_dynamodb”

6. Press Create function