Exericise 7 Partitioned Data

- Create a database

CREATE DATABASE IF NOTEXISTSstudent\_db;

USE student\_db;

-- Create a partitioned table

CREATETABLEstudent\_data (

student\_idINT,

student\_name STRING,

ageINT,

grade STRING

)

PARTITIONED BY (course STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY','

STORED AS TEXTFILE;

-- Load data into the partitioned table

-- You need to specify the partition value when loading data

ALTER TABLE student\_data ADD PARTITION (course='Mathematics');

LOAD DATA INPATH 'student\_data\_mathematics.csv' INTO TABLE student\_data

PARTITION (course='Mathematics');

ALTER TABLE student\_data ADD PARTITION (course='Science');

LOAD DATA INPATH '/user/hive/student\_data\_science.csv' INTO TABLE student\_data

PARTITION (course='Science');

--Display available partitions

Show partitions student\_data

-- Run queries on the partitioned table

-- Display all data from a specific partition (e.g., Mathematics)

SELECT\*FROMstudent\_dataWHERE course ='Mathematics';

-- Query across all partitions to get the count of students per course

SELECT course, COUNT(student\_id) ASstudent\_countFROMstudent\_dataGROUPBY

course;

-- Filter data for students above a certain age across all partitions

SELECT\*FROMstudent\_dataWHERE age >20;

**Start Hive CLI:**

bash

Copy code

hive

**Create a Database:**

sql

Copy code

CREATE DATABASE IF NOT EXISTS sales\_db;

**Use the Created Database:**

sql

Copy code

USE sales\_db;

**Create a Partitioned Table for Products:**

sql

Copy code

CREATE TABLE IF NOT EXISTS products (

product\_id INT,

product\_name STRING,

price FLOAT

)

PARTITIONED BY (category STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Add Two Partitions:**

Hive creates partitions automatically when loading data with a specified partition key. Below, we directly load data into two partitions.

**Load Electronics Data into Partitioned Table:**

sql

Copy code

LOAD DATA LOCAL INPATH '/path/to/electronics\_data.csv'

INTO TABLE products

PARTITION (category='Electronics');

**Load Wearable Data into Partitioned Table:**

sql

Copy code

LOAD DATA LOCAL INPATH '/path/to/wearable\_data.csv'

INTO TABLE products

PARTITION (category='Wearable');

**Query to Select All Products:**

sql

Copy code

SELECT \* FROM products;

**Query to Filter Products by Category:**

sql

Copy code

SELECT \* FROM products

WHERE category = 'Electronics';

**Query to Get Average Price by Category:**

sql

Copy code

SELECT category, AVG(price) AS avg\_price

FROM products

GROUP BY category;

**How to Prepare CSV Files:**

* **electronics\_data.csv:**

Copy code

1,Smartphone,799.99

2,Laptop,999.99

3,Tablet,450.50

* **wearable\_data.csv:**

Copy code

4,Smartwatch,199.99

5,Fitness Tracker,99.99

6,VR Headset,299.99