Exercise 10 ---- INTERNAL AND EXTERNAL

Create a database

CREATE DATABASE IF NOTEXISTSdemo\_db;

USE demo\_db;

-- Create an internal table

CREATETABLEinternal\_table (

student\_idINT,

student\_name STRING,

course STRING,

grade STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY','

STORED AS TEXTFILE;

-- Load data into the internal table

LOAD DATA LOCAL INPATH 'students.csv'INTOTABLEinternal\_table;

-- Create an external table

CREATEEXTERNALTABLEexternal\_table (

student\_idINT,

student\_name STRING,

course STRING,

grade STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY','

LOCATION '/user/hive/external\_students';

-- Load data into the external table

LOAD DATA LOCAL INPATH 'students.csv'INTOTABLEexternal\_table;

-- Query internal table

SELECT\*FROMinternal\_table;

Class: II M.Sc DS Semester: III

P23DS3P6: Big Data Management and Analytics Lab P a g e |44

-- Query external table

SELECT\*FROMexternal\_table;

-- Drop the internal table (this will remove the data)

DROPTABLEinternal\_table;

-- Drop the external table (this will not remove the data)

DROPTABLEexternal\_table;

-- Accessing external table data directly in HDFS after dropping the table

-- Verify that the data still exists in HDFS

-- Use Hadoop commands to check the data location

dfs-ls/user/hive/external\_students;

-- Read data directly from HDFS to confirm it exists

dfs-cat /user/hive/external\_students/student\_data.csv;

-- To reuse the data, Create an external table that refers the existing data file path

CREATEEXTERNALTABLEexternal\_table (

student\_idINT,

student\_name STRING,

course STRING,

grade STRING)

ROW FORMAT DELIMITED

LOCATION '/user/hive/external\_students';

-- Query external table

SELECT\*FROMexternal\_table;

**Start the Hive CLI:**

bash

Copy code

hive

**Create a Database (if needed):**

sql

Copy code

CREATE DATABASE IF NOT EXISTS product\_db;

**Use the Created Database:**

sql

Copy code

USE product\_db;

**Create the Internal Table (pdt\_internal):**

sql

Copy code

CREATE TABLE IF NOT EXISTS pdt\_internal (

product\_id INT,

product\_name STRING,

category STRING,

price FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Load Data into the Internal Table:**

sql

Copy code

LOAD DATA LOCAL INPATH '/path/to/internal\_products.csv'

INTO TABLE pdt\_internal;

**Sample internal\_products.csv:**

Copy code

1,Laptop,Electronics,75000.00

2,Smartphone,Electronics,30000.00

3,Watch,Wearable,15000.00

4,TV,Electronics,40000.00

5,Headphones,Wearable,5000.00

**Query the Internal Table:**

sql

Copy code

SELECT \* FROM pdt\_internal;

**Create the External Table (pdt\_external):**

sql

Copy code

CREATE EXTERNAL TABLE IF NOT EXISTS pdt\_external (

product\_id INT,

product\_name STRING,

category STRING,

price FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/external\_data/products';

**Load Data into the External Table:**

1. **Move the data to HDFS location**:

bash

Copy code

hdfs dfs -mkdir -p /external\_data/products

hdfs dfs -put /path/to/external\_products.csv /external\_data/products/

1. **Query the External Table**:

sql

Copy code

SELECT \* FROM pdt\_external;

**Sample external\_products.csv:**

mathematica

Copy code

6,Tablet,Electronics,20000.00

7,Shoes,Fashion,5000.00

8,Sunglasses,Fashion,3000.00

9,Speaker,Electronics,7000.00

10,Fitness Band,Wearable,2500.00

**Drop Tables (Cleanup):**

sql

Copy code

DROP TABLE IF EXISTS pdt\_internal;

DROP TABLE IF EXISTS pdt\_external;