EXERCISE 8 Indexing

-- Step 1: Create a database

CREATE DATABASE IF NOT EXISTS demo\_db;

USE demo\_db;

-- Step 2: Create a table

CREATE TABLE student\_data (

student\_id INT, student\_name STRING, course STRING, grade STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

-- Step 3: Load data into the table

LOAD DATA LOCAL INPATH 'student\_data.csv' INTO TABLE student\_data;

-- Step 4: Create an index on the 'course' column

CREATE INDEX idx\_course ON TABLE student\_data (course) AS 'COMPACT';

Class: II M.Sc DS Semester: III

P23DS3P6: Big Data Management and Analytics Lab P a g e |33

-- Step 5: Rebuild the index

ALTER INDEX idx\_course ON student\_data REBUILD;

-- Step 6: Query to select data from the indexed table

SELECT \* FROM student\_data WHERE course = 'Mathematics';

-- Step 7: Display query execution plan using EXPLAIN

EXPLAIN SELECT \* FROM student\_data WHERE course = 'Mathematics';

-- Step 8: Drop the index and the table (cleanup)

DROP INDEX IF EXISTS idx\_course ON student\_data;

DROP TABLE IF EXISTS student\_data;

**Start Hive CLI:**

bash

Copy code

hive

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

sql

Copy code

CREATE DATABASE IF NOT EXISTS company\_db;

**Use the Created Database:**

sql

Copy code

USE company\_db;

**Create the Employees Table:**

sql

Copy code

CREATE TABLE IF NOT EXISTS employees (

employee\_id INT,

employee\_name STRING,

department STRING,

salary FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Load Data into the Employees Table:**

sql

Copy code

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

INTO TABLE employees;

**Sample employees\_data.csv (Make sure this is available at the given path):**

mathematica

Copy code

1,John Doe,Sales,55000

2,Jane Smith,IT,80000

3,Robert Brown,HR,60000

4,Emily Davis,Sales,58000

5,Michael Johnson,IT,90000

**Create an Index on the Employees Table:**

sql

Copy code

CREATE INDEX emp\_dept\_idx

ON TABLE employees (department)

AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'

WITH DEFERRED REBUILD;

**Rebuild the Index:**

sql

Copy code

ALTER INDEX emp\_dept\_idx ON employees REBUILD;

**Query to Select All Employees:**

sql

Copy code

SELECT \* FROM employees;

**Query to Filter Employees by Department:**

sql

Copy code

SELECT \* FROM employees

WHERE department = 'IT';