Ex.No: 10	Running and Analyzing HIVE Queries on a Hadoop Cluster
Date: 13.10.25	

Aim:

To create a table and run queries using Hive on a Hadoop cluster to analyze structured data stored in HDFS.

Procedure:

1. **Create an Input file:**
   A CSV or text file (e.g., students.csv) containing structured data:
   101,John,IT,85
   102,Mary,CSE,90
   103,Steve,ECE,78
2. **Start Hadoop Services:**
   start-dfs.sh
   start-yarn.sh
3. **Start Hive Command Line Interface (CLI):**
   hive
4. **Create a Hive Database:**
   CREATE DATABASE studentdb;
   USE studentdb;
5. **Create a Table in Hive:**
   CREATE TABLE students (
   id INT,
   name STRING,
   department STRING,
   marks INT
)
   ROW FORMAT DELIMITED
   FIELDS TERMINATED BY ','
   STORED AS TEXTFILE;
6. **Load Data into Hive Table:**
   First, upload the CSV file to HDFS:
   hdfs dfs -mkdir /user/hive/input
   hdfs dfs -put students.csv /user/hive/input/

   Then load data into the table:
   LOAD DATA INPATH '/user/hive/input/students.csv' INTO TABLE students;
7. **Run HiveQL Queries:**
   View all records:
   SELECT * FROM students;

   Find students with marks > 80:
   SELECT name, department FROM students WHERE marks > 80;

   Count total number of students:
   SELECT COUNT(*) FROM students;

8. **Exit Hive CLI:**
   exit;

Program:

```
-- Start Hadoop services
start-dfs.sh
start-yarn.sh

-- Start Hive CLI
hive

-- Create and use database
CREATE DATABASE studentdb;
USE studentdb;

-- Create table
CREATE TABLE students (
   id INT,
   name STRING,
   department STRING,
   marks INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;

-- Upload data to HDFS
hdfs dfs -mkdir /user/hive/input
hdfs dfs -put students.csv /user/hive/input/

-- Load data into table
LOAD DATA INPATH '/user/hive/input/students.csv' INTO TABLE students;
```

-- Run queries

```
SELECT * FROM students;

SELECT name, department FROM students WHERE marks > 80;

SELECT COUNT(*) FROM students;
```

-- Exit

```
exit;
```

Output:

```
hive> SELECT * FROM students;
        OK
        101 John  IT   85
        102 Mary  CSE  90
        103 Steve ECE  78
```

Result:

Hive queries were successfully executed on a Hadoop cluster. The student data was analyzed using HiveQL, demonstrating Hive's ability to perform SQL-like operations over data stored in HDFS.