

Create tables in Hive and write queries to access the data in the table

AIM:

To create tables in hive and write queries to access the data in the table using Apache Hive and Hadoop.

PROCEDURES:

1. Start Hadoop using `start-all.sh` command
2. Run the SQL Server in the Settings in Mac.
3. Open a terminal and go to the bin folder of Apache Hive and run `hive – service metastore &`
4. Run hive using the command `./hive`
5. Create a database and use the database.
6. Create a table and define its structure.
7. Import the data into the table from HDFS.
8. Query the data using SQL commands.

OUTPUT:



```
~/apache-hive-3.1.3-bin/bin ./hive
hive> CREATE DATABASE sample_database;
OK
Time taken: 0.045 seconds
hive> show databases;
OK
default
sample_database
Time taken: 0.048 seconds, Fetched: 2 row(s)
hive>
```

```
~/apache-hive-3.1.3-bin/bin
./hive

hive> CREATE TABLE employees (
    >     id INT,
    >     name STRING,
    >     age INT,
    >     department STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n';
OK
Time taken: 0.05 seconds
hive> show tables;
OK
employees
Time taken: 0.056 seconds, Fetched: 1 row(s)
hive>
```

```
~/apache-hive-3.1.3-bin/bin
./hive

hive> LOAD DATA INPATH '/user/four/employees.csv' INTO TABLE employees;
Loading data to table sample_database.employees
OK
Time taken: 0.494 seconds
hive>
```

```
~/apache-hive-3.1.3-bin/bin
./hive

hive> SELECT * FROM employees;
OK
1      John Doe      30      Engineering
2      Jane Smith    28      Marketing
3      Bob Johnson   45      Sales
4      Alice Brown   38      HR
5      Michael Davis 50      Finance
6      Emily Wilson  26      Engineering
7      Chris Martinez 32      Marketing
8      Sarah Taylor  29      Sales
9      David Anderson 44      HR
10     Jessica Thomas 31      Finance
11     Daniel Harris  36      Engineering
12     Olivia Lee     27      Marketing
13     Matthew Clark  41      Sales
14     Sophia Robinson 33      HR
15     James Walker   48      Finance
16     Ava King       25      Engineering
17     William Wright 34      Marketing
18     Isabella Scott 39      Sales
19     Alexander Green 42      HR
20     Mia Adams     30      Finance
21     Lucas Baker   28      Engineering
22     Amelia Gonzalez 35      Marketing
23     Benjamin Nelson 46      Sales
24     Ella Carter    37      HR
25     Henry Mitchell 29      Finance
26     Abigail Perez  31      Engineering
```

```
./hive
~/apache-hive-3.1.3-bin/bin
./hive

hive> SELECT name, department FROM employees;
OK
John Doe      Engineering
Jane Smith    Marketing
Bob Johnson   Sales
Alice Brown   HR
Michael Davis Finance
Emily Wilson  Engineering
Chris Martinez Marketing
Sarah Taylor   Sales
David Anderson HR
Jessica Thomas Finance
Daniel Harris  Engineering
Olivia Lee     Marketing
Matthew Clark   Sales
Sophia Robinson HR
James Walker   Finance
Ava King       Engineering
William Wright Marketing
Isabella Scott Sales
Alexander Green HR
Mia Adams     Finance
Lucas Baker   Engineering
Amelia Gonzalez Marketing
Benjamin Nelson Sales
Ella Carter   HR
Henry Mitchell  Finance
Abigail Perez  Engineering
Jackson Roberts Marketing
Lily Turner    Sales
Owen Phillips  HR
```

```
./hive
hive> SELECT * FROM EMPLOYEES WHERE age > 30
> ;
OK
3      Bob Johnson    45      Sales
4      Alice Brown     38      HR
5      Michael Davis   50      Finance
7      Chris Martinez  32      Marketing
9      David Anderson  44      HR
10     Jessica Thomas  31      Finance
11     Daniel Harris   36      Engineering
13     Matthew Clark    41      Sales
14     Sophia Robinson 33      HR
15     James Walker    48      Finance
17     William Wright  34      Marketing
18     Isabella Scott  39      Sales
19     Alexander Green 42      HR
22     Amelia Gonzalez 35      Marketing
23     Benjamin Nelson 46      Sales
24     Ella Carter     37      HR
26     Abigail Perez   31      Engineering
27     Jackson Roberts 40      Marketing
28     Lily Turner     33      Sales
29     Owen Phillips   47      HR
31     Liam Parker     38      Engineering
32     Charlotte Evans 32      Marketing
33     Mason Edwards   44      Sales
35     Ethan Stewart   50      Finance
37     Logan Morris    36      Marketing
39     Jacob Peterson  45      HR
41     Elijah Perez   41      Engineering
43     Lucas Flores   43      Sales
```

```
./hive
hive> SELECT department, COUNT(*) FROM employees GROUP BY department;
Query ID = manoj_20240822104130_1182377c-d213-4884-b3c9-12dfa40a21f7
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1724302421490_0001, Tracking URL = http://localhost:8088/proxy/application_172
4302421490_0001/
Kill Command = /Users/manoj/hadoop-3.4.0//bin/mapred job -kill job_1724302421490_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-08-22 10:41:41,517 Stage-1 map = 0%,  reduce = 0%
2024-08-22 10:41:45,714 Stage-1 map = 100%,  reduce = 0%
2024-08-22 10:41:50,879 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_1724302421490_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  HDFS Read: 14487 HDFS Write: 201 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Engineering      10
Finance          10
HR                10
Marketing         10
Sales              10
Time taken: 22.088 seconds, Fetched: 5 row(s)
hive>
```

RESULT:

Thus, to create a table and query the data using Apache Hive is completed successfully.