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.

PROCEDURE:

1. Start Hadoop using start-all.sh command
2. Run the SQL Server in the Settings in Mac.
3. Run hive using the command ./hive
4. Create a database and use the database.
5. Create a table and define its structure.
6. Import the data into the table from HDFS.
7. Query the data using SQL commands.

OUTPUT:

```
[mohamedhussainshahulhameed@Mohameds-Laptop ~ % hive
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
Hive Session ID = 38686e20-9a2f-4811-8320-db0304c3cd6d

Logging initialized using configuration in jar:file:/Users/mohamedhussainshahulhameed/hive/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
Hive Session ID = bd100f86-cc24-47ec-8275-17d3f6f5817a
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

```
[hive> create database mim;
OK
Time taken: 0.618 seconds
[hive> show databases;
OK
default
mim
Time taken: 0.223 seconds, Fetched: 2 row(s)
```

```
hive> CREATE TABLE employee_info (
    >       emp_id INT,
    >       emp_name STRING,
    >       emp_age INT,
    >       emp_dept STRING,
    >       emp_salary FLOAT
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.564 seconds
```

```
hive> INSERT INTO employee_info VALUES
    > (1, 'Alice', 30, 'HR', 55000.00),
    > (2, 'Bob', 25, 'Engineering', 70000.00),
    > (3, 'Charlie', 28, 'Marketing', 60000.00),
    > (4, 'David', 35, 'Finance', 80000.00),
    > (5, 'Eva', 32, 'HR', 65000.00),
    > (6, 'Frank', 29, 'Engineering', 72000.00),
    > (7, 'Grace', 31, 'Marketing', 62000.00),
    > (8, 'Henry', 40, 'Finance', 90000.00),
    > (9, 'Ivy', 27, 'HR', 58000.00),
    > (10, 'Jack', 26, 'Engineering', 68000.00);
Query ID = mohamedhussainshahulhameed_20240904144050_a2dc20b1-759b-47e1-b607-3cc0bf1538f5
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 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_1724908930607_0003, Tracking URL = http://localhost:8088/proxy/application_1724908930607_0003/
Kill Command = /Users/mohamedhussainshahulhameed/hadoop-3.4.0//bin/mapred job  -kill job_1724908930607_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-09-04 14:41:06,606 Stage-1 map = 0%,  reduce = 0%
2024-09-04 14:41:13,210 Stage-1 map = 100%,  reduce = 0%
2024-09-04 14:41:19,703 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_1724908930607_0003
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employee_info/.hive-staging_hive_2024-09-04_14-40-50_817_4714769106197534765-1/-ext-10000
Loading data to table default.employee_info
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   HDFS Read: 22754 HDFS Write: 838 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 32.011 seconds
hive>
```

```
hive>
[    > select * from employee_info;
OK
1       Alice   30      HR      55000.0
2       Bob     25      Engineering     70000.0
3       Charlie 28      Marketing       60000.0
4       David   35      Finance 80000.0
5       Eva     32      HR      65000.0
6       Frank   29      Engineering     72000.0
7       Grace   31      Marketing       62000.0
8       Henry   40      Finance 90000.0
9       Ivy     27      HR      58000.0
10      Jack    26      Engineering     68000.0
Time taken: 0.412 seconds, Fetched: 10 row(s)
hive>
```

```
[hive> select *from employee_info where emp_salary > 70000;
OK
4       David   35      Finance 80000.0
6       Frank   29      Engineering     72000.0
8       Henry   40      Finance 90000.0
Time taken: 0.418 seconds, Fetched: 3 row(s)
hive>
```

RESULT:

     Thus , create tables in Hive and write queries to access the data in the table has been successfully executed.