Exp.No.: 5a

Design and test various schema models to optimize data storage and retrieval Using Hive

Aim:

To Design and test various schema models to optimize data storage and retrieval Using Hbase.

Procedure:

Step 1: Start Hive

Open a terminal and start Hive by running:

$hive

Step 2: Create a Database

Create a new database in Hive: hive>CREATE DATABASE financials;

```
hive> CREATE DATABASE financials;
OK
Time taken: 0.063 seconds
```

***Step 3: Use the Database**:*

Switch to the newly created database: hive>use financials;

```
hive> show databases;
OK
default
financials
Time taken: 0.358 seconds, Fetched: 2 row(s)
hive> use financials;
OK
Time taken: 0.044 seconds
```

Step 4: Create a Table:

Create a simple table in your database:

hive>CREATE TABLE finance_table(id INT, name STRING);

```
hive> CREATE TABLE finance_table( id INT, name STRING );
OK
Time taken: 2.013 seconds
```

Step 5: Load Sample Data:

You can insert sample data into the table: hive>INSERT INTO finance_tableVALUES (1, 'Alice');

```
hive> INSERT INTO finance_table VALUES
    > (1,'Alice');
Query ID = vaisharli_20240920110300_1dfe3b3e-9ec8-4fde-ae4d-2bd4bc5aeab4
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>
Job running in-process (local Hadoop)
2024-09-20 11:03:02,558 Stage-1 map = 0%,  reduce = 0%
2024-09-20 11:03:03,604 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local527335742_0001
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://localhost:9000/user/hive/warehouse/financials.db/finance_table/.hive-staging_h
ive_2024-09-20_11-03-00_141_7332587824649571237-1/-ext-10000
Loading data to table financials.finance_table
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 0 HDFS Write: 176 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 5.348 seconds
hive> select * from finance_table;
OK
1       Alice
Time taken: 0.138 seconds, Fetched: 1 row(s)
```

Step 6: Query Your Data

Use SQL-like queries to retrieve data from your table:
hive>CREATE VIEW myview AS SELECT name, id FROM finance_table;

```
hive> CREATE VIEW myview AS SELECT name, id FROM finance_table;
OK
Time taken: 0.244 seconds
```

Step 7: View the data:

*To see the data in the view, you would need to query the view hive>SELECT*FROM myview;*

```
hive> select * from myview;
OK
Alice   1
Time taken: 0.131 seconds, Fetched: 1 row(s)
```

Step 8: Describe a Table:

```
hive> DESCRIBE finance_table;
OK
id                      int
name                    string
age                     int
Time taken: 0.729 seconds, Fetched: 3 row(s)
```

You can describe the structure of a table using the DESCRIBE command: hive>DESCRIBE finance_table;

Step 9: Alter a Table:

You can alter the table structure by adding a new column: hive>ALTER TABLE finance_table ADD COLUMNS (age INT);

```
hive> alter table finance_table add columns (age INT);
OK
Time taken: 0.501 seconds
```

Step 10: Quit Hive:

To exit the Hive CLI, simply type: hive>quit;

```
hive> quit;
vaisharli@vaisharli:~$
```

Result:

Thus, the usage of various commands in Hive has been successfully completed.