Exp. No : 5a

Designing and testing various schema models to optimize data storage and retrieval using Hive.

1. Start hive

```
harish@fedora:~/hive/lib$ hive
which: no hbase in (/home/harish/.local/bin:/home/harish/bin:/
rish/pig/bin:/usr/lib/jvm/jdk1.8.0_202/bin:/home/harish/hive/b
me/harish/hive/bin)
Hive Session ID = 72f6d993-0190-4d45-bd88-71acc8c52db4

Logging initialized using configuration in jar:file:/home/hari
Hive Session ID = 6d0c44f6-0f4e-4f77-8b7b-783351811c7d
Hive-on-MR is deprecated in Hive 2 and may not be available in
ases.
hive>
```

2. CREATE DATABASE financials in hive

```
hive> CREATE DATABASE financials;
OK
Time taken: 8.319 seconds
```

3. Use financials database in hive

```
hive> use financials;
OK
Time taken: 0.591 seconds
```

4. Create Finance_table table in hive

```
hive> CREATE TABLE finance_table( id INT, name STRING );
OK
Time taken: 7.98 seconds
```

5. Insert records in finance_table table

```
hive> INSERT INTO finance_table VALUES (1, 'Alice'), (2, 'Bob'), (3, 'Charlie');
Query ID = hayagreevan_20240920110953_28c8669c-6564-4aa2-9392-f8adbe779f31
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_1726810143118_0001, Tracking URL = http://fedora:8088/proxy/application_1726810143118_0001/
Kill Command = /home/hayagreevan/hadoop/bin/mapred job  -kill job_1726810143118_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-09-20 11:11:23,784 Stage-1 map = 0%,  reduce = 0%
2024-09-20 11:11:57,686 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.06 sec
2024-09-20 11:12:17,572 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 14.5 sec
MapReduce Total cumulative CPU time: 14 seconds 500 msec
Ended Job = job_1726810143118_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_hive_2024-09-20_11-09-53_315_6641591007717348106-
Loading data to table financials.finance_table
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 14.5 sec   HDFS Read: 15715 HDFS Write: 291 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 500 msec
OK
Time taken: 157.02 seconds
```

6. Creating new VIEW named myview for finance_table

```
hive> CREATE VIEW myview AS SELECT name, id FROM finance_table;
OK
Time taken: 1.849 seconds
```

7. Display myview.

```
hive> SELECT * FROM myview;
OK
Alice   1
Bob     2
Charlie 3
Time taken: 4.11 seconds, Fetched: 3 row(s)
```

8. Describing finance_table structure.

```
hive> DESCRIBE finance_table;
OK
id                      int
name                    string
Time taken: 0.447 seconds, Fetched: 2 row(s)
```

9. Add new age column to Finance_table

```
hive> ALTER TABLE finance_table ADD COLUMNS (age INT);
OK
Time taken: 1.46 seconds
```