Assignment No. 4

Aim -

Write an application using HBase & HiveQL for Flight information system.

Problem Statement -

Write application using HBase & HiveQL for flight information system which will include,

a) Create, drop & altering tables
b) Creating external hive table to connect to HBase
c) Load table with data, insert new values & fields in table.

Theory :-

HBase via Hive :

This is the second of two posts ~~examing~~ examining the use of Hive for interaction with HBase tables.

Hive :

Hive is datawarehouse infrastructure tool to process structured data in hadoop.

PAGE NO.....
DATE ...../...../.....

Hive is not :

a) A Relational database

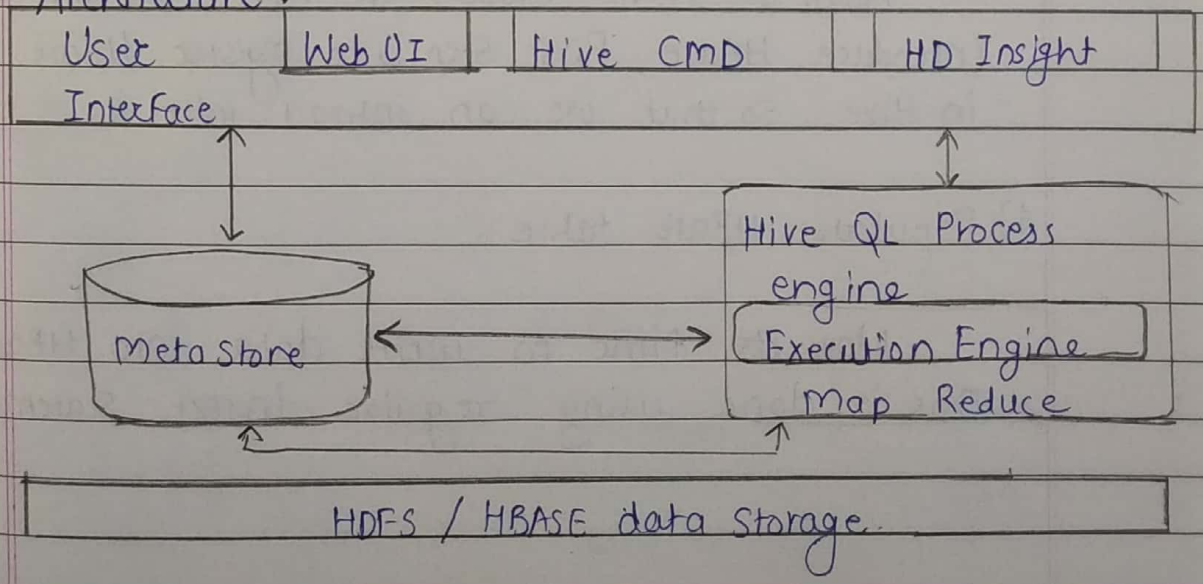
b) Design for OLTP

c) A language for realtime queries & row level updates.

Feature of Hive:

- Stores schema in database & processed data into HDFS.
- Designed for OLAP
- Provides SQL type language for querying called HiveQL.
- It is Familiar, Fast, scalable & extensible.

Architecture :



PAGE NO.
DATE ../../..

Hive interact with HBase

1) Grab some data & register it in Hive.

```
$ cat 00-pagecounts.ddl
Create table IF not exists pagecounts (
projectcode, String, pageviews String)
Row Format
    delimited Fields terminated by ' '
    lines terminated by '\n'
    Stored as text file.
```

Run script to register dataset with Hive.

2) Transfer schema for HBase

Next step is to transform row data in a schema that makes sense for HBase

3) Register HBase table

Now, we have dataset in Hive, its time to introduce HBase. First step is to register HBase table in Hive so that we can interact with it.

4) Populate HBase table

Now its time to write data into HBase. This is done using regular insert statement.

PAGE NO.....

DATE ...../...../.....

5) Query data from HBase - Land

You now have data in HBase

6) Verify data from Hive

7) Continue using Hive for Analysis.

Conclusion :

Understand to use various aggregate functions using Map-Reduce.