

Lab assignment 3

Group A

- Title: Write an application using Hive QL for flight information system.
- Objectives:
 - a) To learn NOSQL databases such as Hive and Hbase.
 - b) To study detailed concepts of HIVE.
- Software requirements:
 - a) Ubuntu 14.04 / 14.10
 - b) GNU C compiler
 - c) Hadoop
 - d) Java and HIVE
- Problem statement: Write an application using HiveQL for flight ~~to~~ information system which will include creating, dropping and altering database tables, creating an external HIVE table to connect to the Hbase for customer information table.
Load table with data, insert new values and field in the table, join tables with HIVE, create index on flight information table. Find the average departure delay per day in 2008.
- Theory:
 - a) Introduction to Hive -
 - Hive is a data warehouse infrastructure tool to produce structured data in Hadoop.
 - It resides on top of Hadoop to summarize big data, and make querying

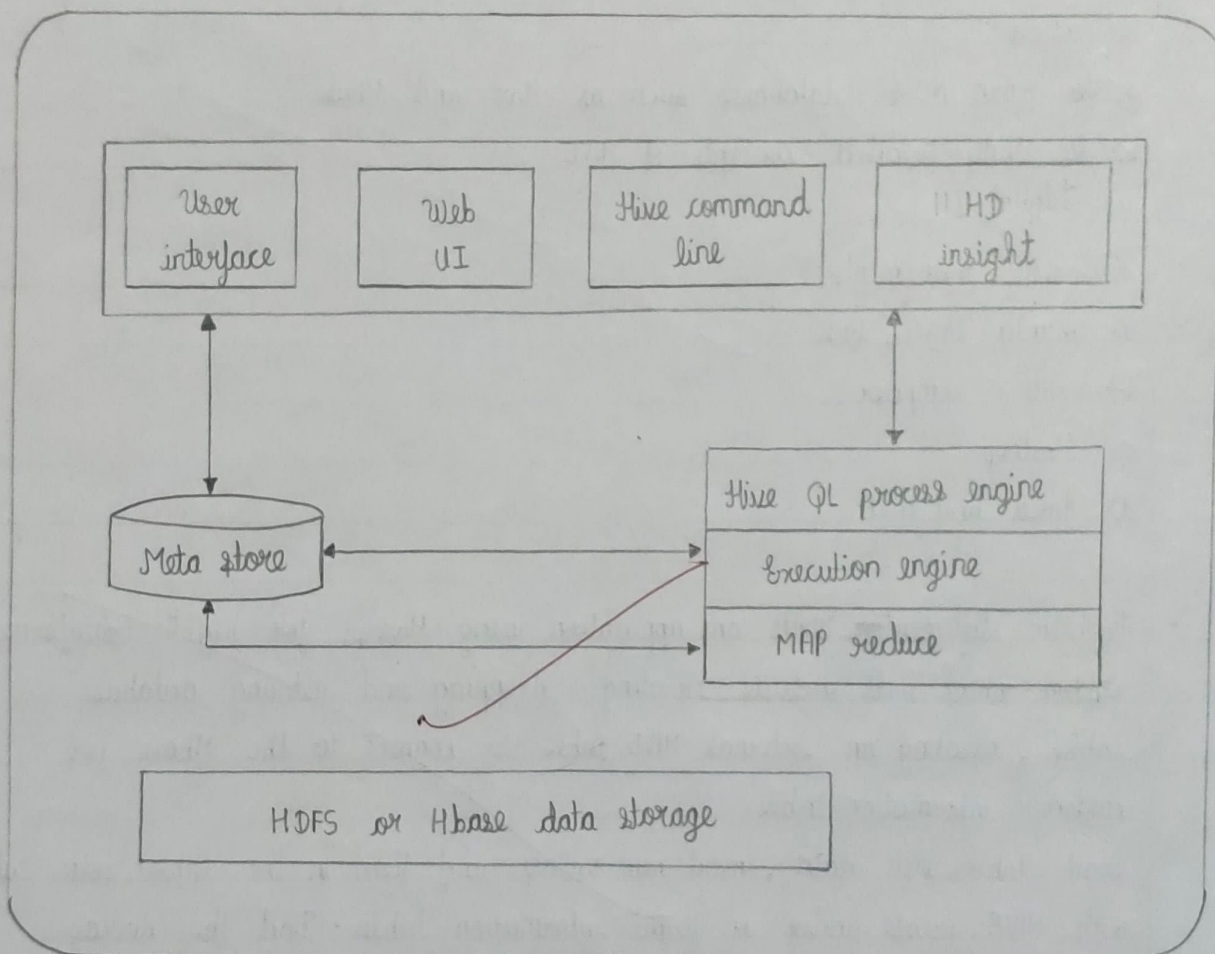


Fig. 1.: Hive architecture

and analysing easy.

- Pre-requisites: Core Java, database concepts, Hadoop file system, any of the linux operating system.
- Features: It stores schema in a database and processed data into HDFS. It is designed for OLAP. It provides SQL type language for querying called HiveQL or HQL. It is familiar, fast scalable and extensible.
- Architecture: This component diagram contains different units.
The following table describes each unit -

Unit name	Operation
a) User interface	Hive is a data warehouse infrastructure software that can create interaction between user and HDFS. The user interfaces that Hive supports are Hive web UI, Hive command line and Hive HD insight (in windows server).
b) Meta store	Hive chooses respective databases servers to store the schema/metadata of tables, database columns in a table, their data types and HDFS mapping
c) HiveQL process engine	HiveQL is similar to SQL for querying on schema or metadata of tables, databases on metastore. It is one of traditional approach replacement for map reduce program.
d) Execution engine	The conjunction part of HiveQL process engine and map Reduce is Hive execution engine. Execution engine processes the query and generates

results as same as map reduce results.

⇒ HDFS or HBase

Hadoop distributed file system or Hbase are the data storage techniques to store data into file system.

→ Conclusion: Thus, we have learnt how to design a application Hbase and HiveQL for flight information system.