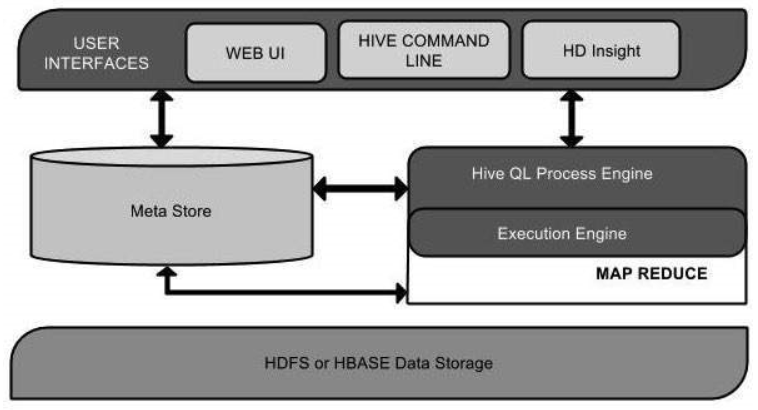
Assignment 6.3

Explain Hive Architecture in Brief

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy.

Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

The following depicts the architecture of HIVE



Hive has a user interface that helps to interact with HDFS

Metastore store schema (metadata of tables), databases, columns in tables, their data types and HDFS mapping

HiveQL Process Engine is similar to SQL. Instead of Java mapreduce program, we can write SQL like query to query mapreduce job and process it.

Execution engine processes the query and generates results as same as MapReduce results.

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

Explain Hive Components in Brief

Following are Hive components –

1. Command line interface – To access & execute HIVE query
2. Hive Server
3. Hive Web Interface – It is web based GUI
4. Execution Engine - This component executes the execution plan created by the compiler.
5. Metastore - This component stores all the structure information of the various table and partitions in the warehouse.

All components are explained as part of question#1 in Hive architecture.