

# Assignment 6.3

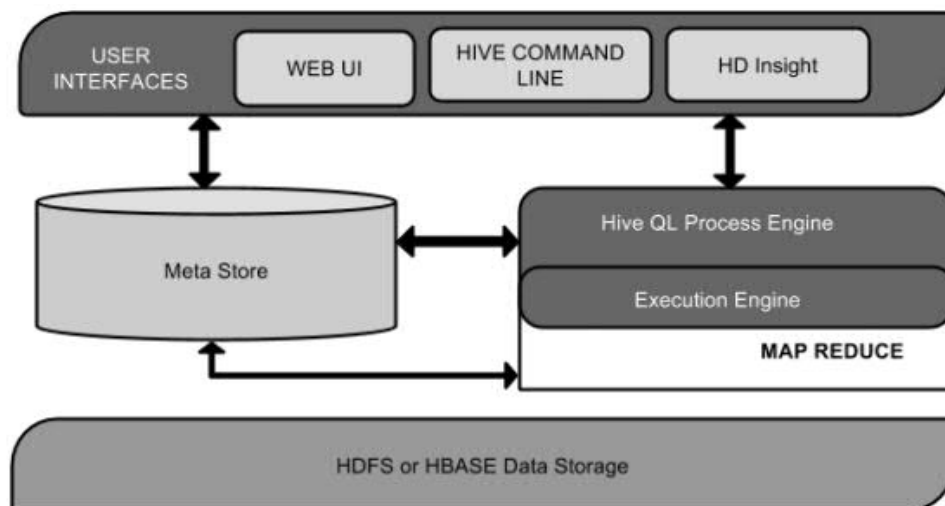
## Problem Statement 1:-

Explain Hive Architecture in Brief.

Solution:-

### Architecture of Hive

The following component diagram depicts the architecture of Hive:



This component diagram contains different units. The following table describes each unit:

### 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).

### **Meta Store**

Hive chooses respective database servers to store the schema or Metadata of tables, databases, columns in a table, their data types, and HDFS mapping.

### **HiveQL Process Engine**

HiveQL is similar to SQL for querying on schema info on the Metastore. It is one of the replacements of traditional approach for MapReduce program. Instead of writing MapReduce program in Java, we can write a query for MapReduce job and process it.

### **Execution Engine**

The conjunction part of HiveQL process Engine and MapReduce is Hive Execution Engine. Execution engine processes the query and generates results as same as MapReduce results. It uses the flavor of MapReduce.

### **HDFS or HBASE**

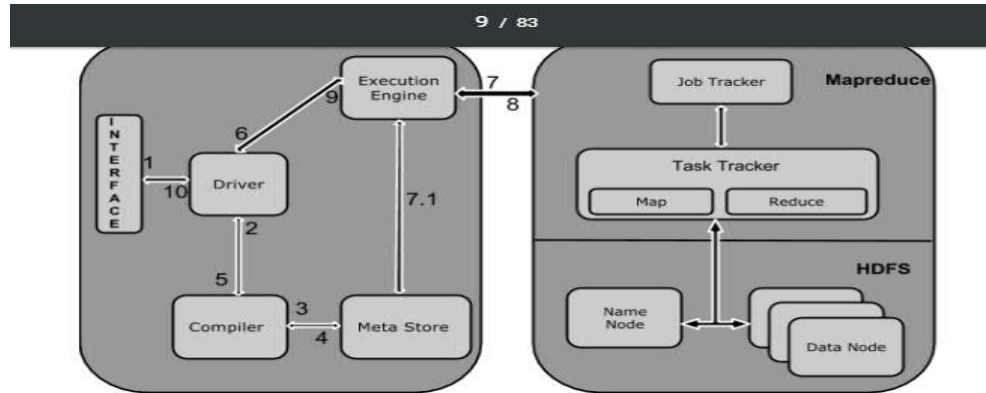
Hadoop distributed file system or HBASE are the data storage techniques to store data into file system.

## Problem Statement 2:-

Explain Hive Components in Brief.

Solution:-

The following diagram will show the components of hive :



### 1 Execute Query

The Hive interface such as Command Line or Web UI sends query to Driver (any database driver such as JDBC, ODBC, etc.) to execute.

### 2 Get Plan

The driver takes the help of query compiler that parses the query to check the syntax and query plan or the requirement of query.

### 3 Get Metadata

The compiler sends metadata request to Metastore (any database).

### 4 Send Metadata

Metastore sends metadata as a response to the compiler.

## **5 Send Plan**

The compiler checks the requirement and resends the plan to the driver. Up to here, the parsing and compiling of a query is complete.

## **6 Execute Plan**

The driver sends the execute plan to the execution engine.

## **7 Execute Job**

Internally, the process of execution job is a MapReduce job. The execution engine sends the job to JobTracker, which is in Name node and it assigns this job to TaskTracker, which is in Data node. Here, the query executes MapReduce job. Metadata Ops Meanwhile in execution, the execution engine can execute metadata operations with Metastore.

## **8 Fetch Result**

The execution engine receives the results from Data nodes. Send Results the execution engine sends those resultant values to the driver. Send Results the driver sends the results to Hive Interfaces