**Assignment 6.3:**

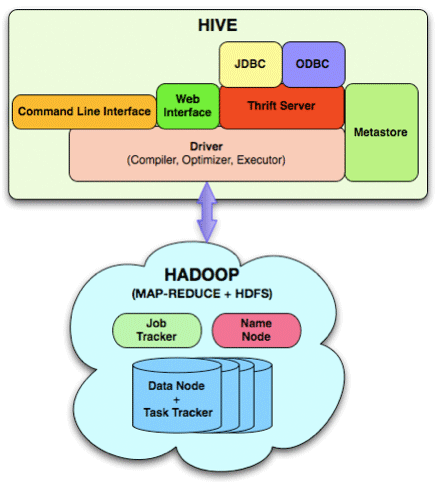
**Explain hive architecture in brief:**

**Explain hive component in brief:**

* Hive Architecture:

**Hive –** It is a data warehouse solution built on top of Hadoop which is used to facilitate the querying and managing large datasets residing in the distributed storage. It is a language identical to SQL in syntax and is known as HQL (Hive Query LANGUAGE). Hive allows traditional map reduce programs to customize mappers and reducers. It can be used to integrate with other databases like SQL Server and MYSQL.

Further It works on the batch processing methodology and one can accept latency while executing jobs. Hive is not suitable for real-time analysis. Hive is best suited for bath jobs on immutable data.



The above diagram shows a basic hive architectural diagram. The first components are CLI (Command Line Interface) which is the hive terminal, JDBC/ODBC to connect from various databases or a GUI like HUE to interact with hive. It connects using API (Thrift server) and then to Hive divers, When the user comes through an UI like Ambari/Hue its directly connects to the Hive drivers.

The hive driver receives various jobs from the user and is sent across to the hadoop architecture. The hadoop architecture uses its various components as name node, data node, YARN to receive and divide the work using bucketing and partitioning methodology.



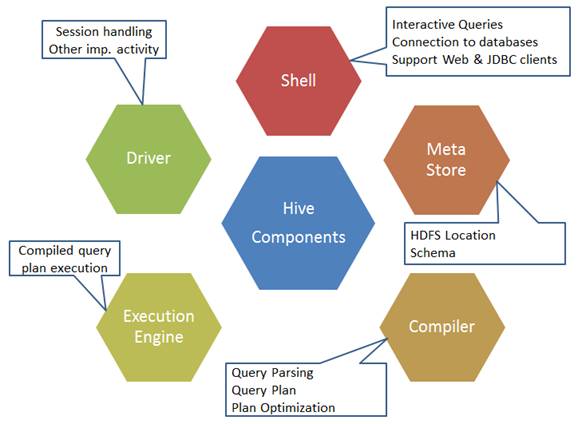
Above is the Complete hadoop architecture.

The execute query instance is initiated from the UI to the hive driver. The driver creates a session handle against each query and sends the query to compiler to generate an execution plan. The compiler needs the metadata so it sends a request to get metadata and receives the metadata from the metastore.

This metadata is used to prune partitions based on query predicates. The plan generated by the compiler is a DAG of stages with each stage being a map/reduce job. The plan contains map operator and reduce operator trees.

**Hive Components:**

The major components of Hive are:



* **UI**: UI means User interface for users to submit queries and other operations to the system. HUE and Ambari are some of the few UI provided by Hive
* **Driver**: The driver is used to receive the queries from the UI. This component implements the notion of session handles and provides execute and fetch APIs modeled on JDBC/ODBC interfaces.
* **Compiler:**The component that parses the query, does semantic analysis on the different query blocks and query expressions and eventually generates an execution plan with the help of the table and partition metadata looked up from the metastore.
* **Metastore**: This component stores all the structure information of various tables and partitions in the warehouse including column type information, the serializers and deserializers necessary to write and read data
* **Execution Engine:** The component which executes the execution plan created by the compiler. The execution engine manages the dependencies between these different stages of the plan.