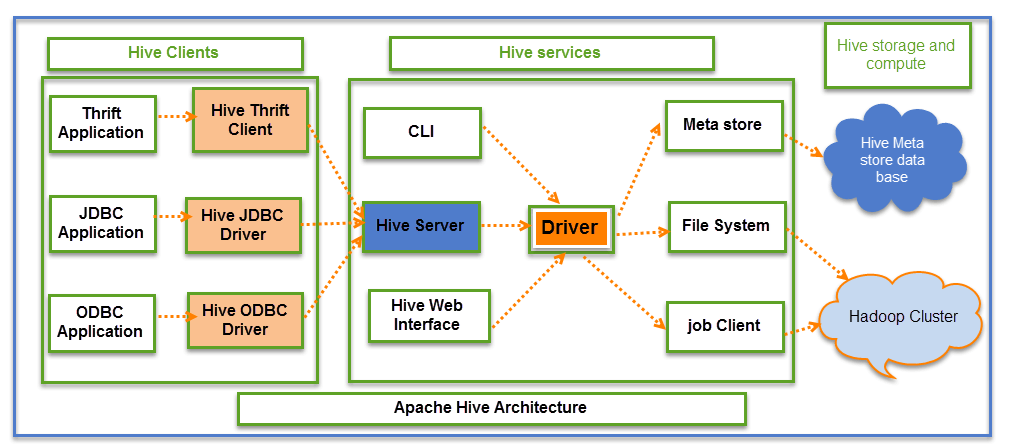
Hive Architecture:



* Hive is a ETL and data warehousing component of hadoop which is built on top of HDFS. It works on row-columnar approach.
* Hive neither stored data nor it process the data, what it does is it converts SQl kind of queries written on it to MapReduce programs and pushes this M/R jobs into Resource manager, which in turn is used for processing.
* To store the metadata info like table schema, data types of each columns we can use default derby DB, Thrift server or MySQL.
* This submission of HiveQL queriesis done by :

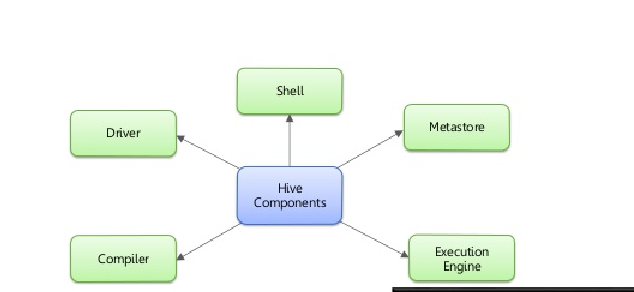
CLI – Command Line Interface (Hive shell)

HWI – Hive Web Interface – HUE, Ambari

Thrift server – JDBC/ODBC.

* So hive can be considered as a grid between HDFS and user.
* This converted M/R jobs will be internally contact Hive metastore for meta data information for processing.

Hive components:



**UI :** UI means User Interface, The user interface for users to submit queries and other operations to the system.

**Driver :** The Driver is used for receives the quires from 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:** The component that stores all the structure information of the various tables and partitions in the warehouse including column and column type information, the serializers and deserializers necessary to read and write data and the corresponding HDFS files where the data is stored.

**Execution Engine:**The component which executes the execution plan created by the compiler. The plan is a DAG of stages. The execution engine manages the dependencies between these different stages of the plan and executes these stages on the appropriate system components.