* Explain Hive Architecture in Brief.
  + HIVE is data warehouse tool used for querying big data. And was initially developed by facebook later on taken up by Apache and developed further and made available for free.



Above shows the schema of hive architecture

There are different blocks in the hive

**UI –** User interface is the one which interacts with the user and helps the end user to write the query.

There are two types of hive UI

Command line and web interface but the command line is the most used ui.

Hive is written in HQL(Hive Query Language) which is very similar to that of the SQL query

**Driver-** Driver receives the code from the UI and then optimizes the query if required and implements the notion of session handles and provides execute and fetch APIs modelled on JDBC/ODBC interfaces.

**Compiler-** After the Driver the control is transferred to the compiler and it performs the sematic analysis on the query blocks and query expressions and then generates the execution plan with the help of the table and partition metadata looked up from the metastore.

**Metastore-** It is 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 –** Execution engine takes plan from the compiler and then executes the stages one by one in order and also it maintains the interdependences of the different stages with each other. It is the one which deals with HDFS for the data to be processed and also stores data in the HDFS

* Explain Hive Components in Brief

**MetaStore -** Stores the system catalog and metadata about tables, columns, partitions etc.

**Driver-** Manages lifecycle of a HiveQL statement as it moves through Hive. It also maintains a session handle and session statistics.

**Compiler -** Compiles HiveQL into a directed acryclic graph of MapReduce tasks.

**Execution Engine -** Executes the tasks produced by the compiler in proper dependency order.

**Hive Server –** It allows the user to submit the query to hive using different programming languages and retrieve the results. It is optional service and it is also called as the Thrift Server.

**Extensibility Interfaces-** Is like the Command Line Interface (CLI), the web UI and JDBC/ODBC driver.It includes SerDe as well as the UDF(User Defined Function) and UDAF(User Defined Aggregate Function) interfaces that enable users to define their own custom functions.