1. **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.

Hive Consists of Mainly 3 main parts

1. Hive Clients
2. Hive Services
3. Hive Storage and Computing

**Hive Clients:**

Hive provides different drivers for communication with a different type of applications. For Thrift based applications, it will provide Thrift client for communication.

For java related applications, it provides JDBC Drivers. Other than any type of applications provided ODBC drivers. These Clients and drivers in turn again communicate with Hive server in the Hive services.

**Hive Services:**

Client interactions with Hive can be performed through Hive Services. If the client wants to perform any query related operations in Hive, it has to communicate through Hive Services.

CLI is the command line interface acts as Hive service for DDL (Data definition Language) operations. All drivers communicate with Hive server

**Hive Storage and Computing:**

Hive services such as Meta store, File system, and Job Client in turn communicates with Hive storage and performs the following actions

* Metadata information of tables created in Hive is stored in Hive "Meta storage database".
* Query results and data loaded in the tables are going to be stored in Hadoop cluster on HDFS or HBASE

**ii. Explain Components of Hive in Brief**

**Major Components of Hive are:**

**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 queries 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.

**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.