**Apache Hive** is a Data warehousing tool built on top of Hadoop. It is primarily used for analysis of huge datasets.

# Hive Architecture

The below diagram describes the Architecture of Hive and Hive components. It also describes the flow in which a query is submitted into Hive and finally processed using the MapReduce framework:



Above diagram shows the major components of Apache Hive-

**Hive Clients** – Apache Hive supports all application written in languages like C++, Java, Python etc. using JDBC, Thrift and ODBC drivers. Thus, one can easily write Hive client application written in a language of their choice.

**Hive Services** – Hive provides various services like web Interface, CLI etc. to perform queries.

1. **CLI (Command Line Interface)** – This is the default shell that Hive provides, in which one can execute their Hive queries and command directly.
2. **Web Interface** – Hive also provides web based GUI for executing Hive queries and commands.
3. **Hive Server**- It is built on Apache Thrift and thus is also called as Thrift server. It allows different clients to submit requests to Hive and retrieve the final result.
4. **Hive Driver** – Driver is responsible for receiving the queries submitted Thrift, JDBC, ODBC, CLI, Web UL interface by a Hive client.
   1. **Complier** –After that hive driver passes the query to the compiler where parsing, type checking, and semantic analysis takes place with the help of schema present in the metastore.
   2. **Optimizer** – It generates the optimized logical plan in the form of a DAG (Directed Acyclic Graph) of MapReduce and HDFS tasks.
   3. **Executor** – Once compilation and optimization complete, execution engine executes these tasks in the order of their dependencies using Hadoop.
5. **Metastore** – Metastore is the central repository of Apache Hive metadata in the Hive Architecture. It stores metadata for Hive tables (like their schema and location) and partitions in a relational database. It provides client access to this information by using metastore service API. Hive metastore consists of two fundamental units:
   1. A service that provides metastore access to other Apache Hive services.
   2. Disk storage for the Hive metadata which is separate from HDFS storage.

**Processing framework and Resource Management** – Hive internally uses Hadoop MapReduce framework to execute the queries.

**Distributed Storage** – Hive is built on the top of Hadoop, so it uses the underlying HDFS for the distributed storage.