**Hive:**

Apache Hive is a data warehouse system developed by Facebook to process a huge amount of structure data in Hadoop. We know that to process the data using Hadoop, we need to right complex map-reduce functions which is not an easy task for most of the developers. Hive makes this work very easy for us.

It uses a scripting language called HiveQL which is almost similar to the SQL. So now, we just have to write SQL-like commands and at the backend of Hive will automatically convert them into the map-reduce jobs.

**Hive architecture:**



Hive is a data warehouse infrastructure software that can create interaction between user and HDFS. Hive chooses respective database servers to store the schema or Metadata of tables, databases, columns in a table, their data types, and HDFS mapping. HiveQL is similar to SQL for querying on schema info on the Metastore.  Execution engine processes the query and generates results as same as MapReduce results. It uses the flavor of MapReduce. Hadoop distributed file system or HBASE are the data storage techniques to store data into file system.

**Working of Hive:**



**Data Types in Apache Hive**

Hive data types are divided into the following 5 different categories:

1. Numeric Type: TINYINT, SMALLINT, INT, BIGINT
2. Date/Time Types: TIMESTAMP, DATE, INTERVAL
3. String Types: STRING, VARCHAR, CHAR
4. Complex Types: STRUCT, MAP, UNION, ARRAY
5. Misc Types: BOOLEAN, BINARY