* **Hive**

The Apache Hive is a data warehouse software built on the top of Hadoop that facilitates reading, writing, and managing large datasets stored in HDFS. The data stored by Hive may residing on distributed storage which can be queried using SQL. It is mainly used for storing the data in a data warehouse and performs different SQL operations over that. The hive provides a simple SQL-like query language called Hive Query Language (HQL) for querying and managing the large datasets. Hive engine compiles these queries into Map-Reduce jobs to be executed on Hadoop.

Using the Hive query language (HiveQL), which is very similar to SQL, queries are converted into a series of jobs that execute on a Hadoop cluster through MapReduce or Apache Spark. The Users can run batch processing workloads with Hive as well as can analyze the same data simultaneously using interactive SQL or using machine-learning workloads over the tools like Apache Impala or Apache Spark within a single platform. As Hive is a petabyte-scale data warehouse system built on the top of Hadoop platform, it allows programmers to write the custom Map-Reduce framework to perform more sophisticated analysis and data processing. Therefore, Hive on MapReduce or Spark is best-suited for batch data preparation or ETL.

The various SQL commands used in hive are given in Table IV. It has support for different aggregation functions of SQL like SUM, COUNT, MAX, MIN etc. and other functions like CONCAT, SUBSTR, ROUND etc. It also supports GROUP BY and SORT BY clauses along with joins. It comes with a command-line shell interface which can be used to create tables and execute queries. In addition, custom Map-Reduce scripts can also be plugged into queries.

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Function** | **Command** |
| 1 | Create database | hive> create database dbname; |
| 2 | View databases | hive> show databases; |
| 3 | Select database | hive> use dbname; |
| 4 | Copy input data to HDFS from local | hive> hadoop dfs –copyFromLocal <local path to file> hdfs:/ |
| 5 | Create table | hive> CREATE TABLE <tablename> (attribute-1 datatype, attribute-2 datatype, -attribute-n datatype); |
| 6 | Copy local data in Hive table | hive> LOAD DATA INPATH <filename> OVERWRITE INTO TABLE <tablename> |
| 7 | Insert data | hive> INSERT INTO [TABLE] <tablename> [(column\_list)]  [PARTITION (partition\_clause)] |
| 8 | Retrieving all the values | hive> SELECT \* FROM <tablename>; |
| 9 | Update data | hive> UPDATE <tablename> SET [ column = value ...] [WHERE expression] |
| 10 | Delete Touple | hive> DELETE FROM <tablename> [WHERE expression] |