

HIVE

Apache Hive is an open source data warehouse system built on top of Hadoop used for querying and analyzing large datasets stored in Hadoop files.

Instead have to write complex Map-Reduce jobs, with the help of the Hive, one just need to submit merely SQL queries. Hive is mainly targeted towards users who are comfortable with SQL. Hive use language called HiveQL (HQL), which is like SQL. HiveQL automatically translates SQL-like queries into MapReduce jobs.

Hive abstracts the complexity of Hadoop. The main thing to notice is that there is no need to learn java for Hive.

The Hive generally runs on your workstation and converts your SQL query into a series of jobs for execution on a Hadoop cluster. Apache Hive organizes data into tables. This provides a means for attaching the structure to data stored in HDFS.

