# INTRODUCTION
## TO HIVE

**Data Sheet**

# Introduction to Hive

Hive is being built on top of Hadoop. It is a data warehouse system that is stored in HDFS for querying and analysis of data. Hive is an open source project that helps programmers on Hadoop to examine massive data sets.

In the business intelligence industry, the size of data sets being collected and analyzed is rising and, in a way, conventional data warehousing strategies are becoming more costly. With the MapReduce system, Hadoop is used as an alternative approach for analyzing large data sets.

Although Hadoop has proven useful for working on large data sets, its MapReduce architecture is very poor, requiring programmers to write customized programs that are difficult to maintain and reuse. Hive comes here for programmers to rescue.

Hive built on top of the Hadoop Map-Reduce architecture as a data warehousing solution.

Hive offers a declarative language similar to SQL, called HiveQL, which is used to express queries. Using Hive-QL, SQL-associated users are very quickly able to perform data analysis.

These queries are compiled by the Hive engine into Map-Reduce jobs to be performed on Hadoop. Moreover, it is also possible to plug custom Map-Reduce scripts into queries. Hive works on data stored in tables consisting of primitive data types and types of data collection, such as arrays and maps.

Hive comes with a shell interface on command line that can be used to construct tables and to execute queries.

Hive query language is equivalent to SQL in which it supports subqueries. It is possible to take a MapReduce joint through Hive tables with the Hive query language. It supports simple SQL functions such as CONCAT, SUBSTR, ROUND, and aggregation functions such as SUM, COUNT, MAX, and so on. It also supports clauses relating to Party BY and SORT BY. User-defined functions can also be written in the Hive query language.

# Difference Between Hive and Map Reduce

| Hive | Map Reduce |
| --- | --- |
| It supports SQL for interaction and data modelling, such as query language. | ✓ It compiles a language with two key tasks in it. One is the job of mapping and the other is a reducer.<br>✓ These tasks can be described using Java or Python |
| Higher abstraction level at the top of the HDFS | Lower level of abstraction |
| Comparatively less than Map ReduceC | Offering high efficiency |
| Less number of lines of code needed for execution | More number of lines of codes to be listed |
| Less development work is required | More development work is required |