

Apache Hadoop vs other Database technologies

Most database management systems are not up to scratch for operating at such lofty levels of Big Data exigencies either due to the sheer technical inefficiency or the insurmountable financial challenges posed. When the type of data is totally unstructured, the volume of data is humongous, the results needed are at breakneck speeds, then the only platform that can effectively stand up to the challenge is Apache Hadoop.

Hadoop owes its runaway success to a certain processing framework called as MapReduce that is central to its existence. The MapReduce technology lets ordinary programmers contribute their part where large data sets are divided and are independently processed in parallel. These coders need not know the nuances of high-performance computing and can work efficiently without having to worry about intra-cluster complexities, monitoring of tasks, node failure management, and so on.

Another aspect of Hadoop is the Hadoop Distributed File System (**HDFS**). The biggest strength of HDFS is its ability to rapidly scale and work without a hitch irrespective of any fault with the nodes. HDFS in essence divides large file into smaller blocks or chunks usually ranging from 64 to 128MB which are then copied onto a couple of nodes of the cluster. This way HDFS ensures no work would stop even in the case of some nodes going out of service. HDFS also has APIs to ensure The MapReduce program can go about reading and writing data simultaneously at high speeds. When there is a need to speed up performance or when there is extra incoming data that needs to be processed then all one has to do is add extra nodes in parallel to the cluster and the increased demand can be immediately met.