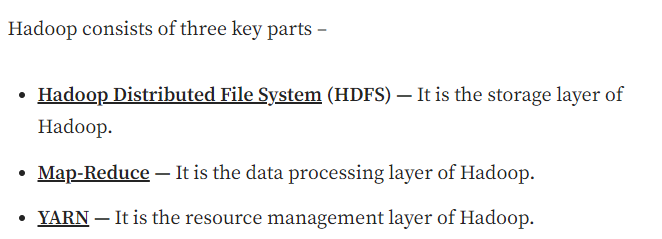


Big Data and Hadoop efficiently processes large volumes of data on a cluster of commodity hardware. Hadoop is for processing huge volume of data. Commodity hardware is the low-end hardware, they are cheap devices which are very economical. Hence, Hadoop is very economic.



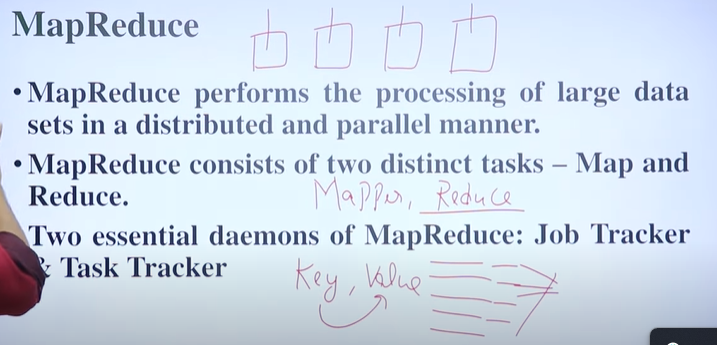
Useful Link: <https://medium.com/@patelharshali136/hadoop-tutorial-for-beginners-learn-hadoop-from-a-to-z-e4f849ee83eb>

HDFS is developed to handle huge volumes of data. The file size expected is in the range of GBs to TBs. A file is split up into blocks (default 128 MB) and stored distributedly across multiple machines. These blocks replicate as per the replication factor.(default 3)

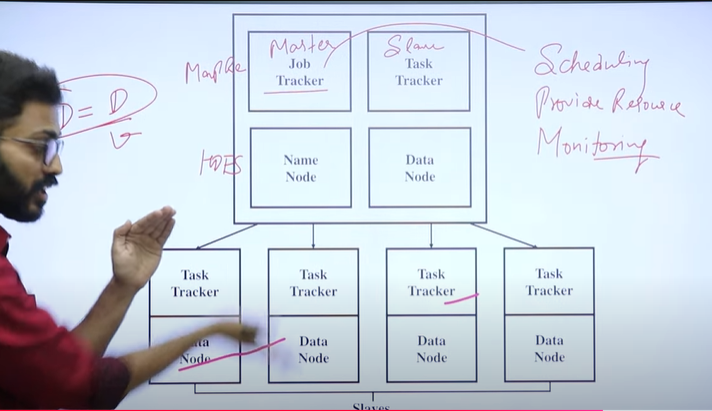
**Map Reduce:**

Works on Divide and conquer

Daemons: Background processes.



Processing delay is very less in the Hadoop because it processes the data where it’s stored.



Here there are different nodes:

Job Tracker:

It’s the master node for processing it keeps track of:-

* Monitoring
* Resource Management
* Scheduling

Task Tracker:

It’s the slave node which actually processes the data provided by the Job Tracker.