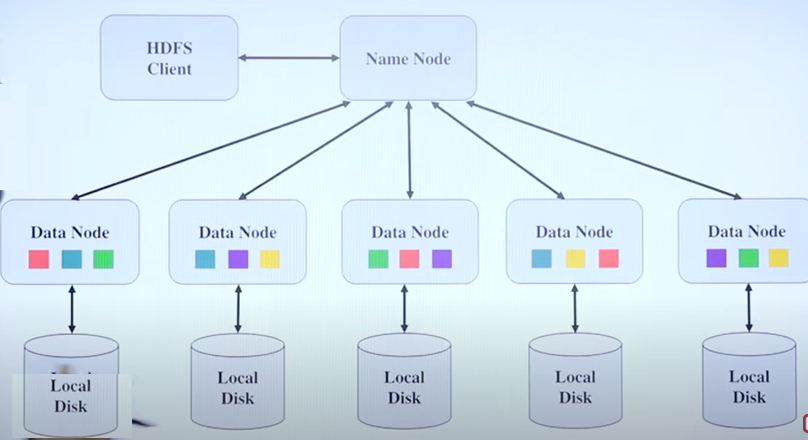
**Big Data:**

|  |  |
| --- | --- |
| Small Data(RDBMS) | Big Data(Open SQL) |
| 1. Mostly Structured | 1. Mostly Unstructured |
| 1. Storage Capacity in MB,GB,TB etc. | 1. Storage capacity in PB,EB,XB,YB etc.. |
| 1. Gradually Increase | 1. Exponentially Increase |
| 1. Local or centralised Data | 1. Global or Distributed Data |
| 1. Software : SQL Server,Oracle,MySQL. | 1. Hadoop, Spark etc. |
| 1. Single Node | 1. Multi-node Cluster |

**Hadoop:**

1. It is an opensource framework that allows us to store and process large data set in a parallel and distributed manner.
2. Two main component HDFS and MapReduce.
3. Hadoop Distributed File System(HDFS) is the primary data storage system used by Hadoop Applications.
4. MapReduce is the processing unit of Hadoop.



By default file is partition

In 128 MB

/Master Node