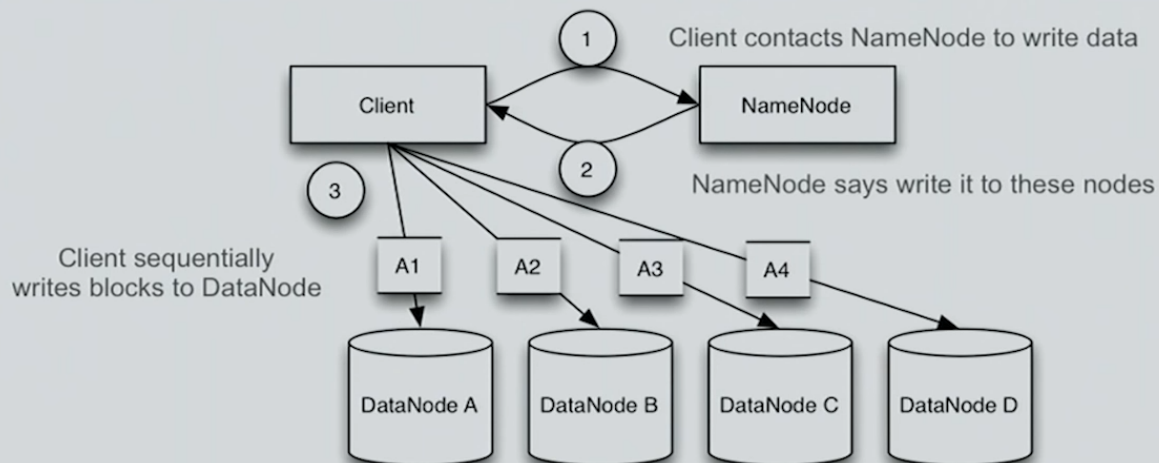


Big Data

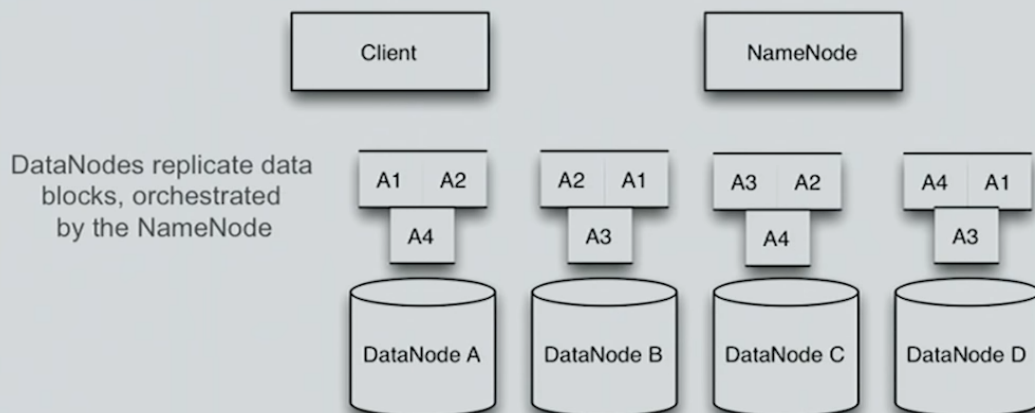
Components of Hadoop

- **Storage Unit - Hadoop Distributed FileSystem (HDFS)** : Data is distributed amongst many computers and stored in blocs . Each file is split into multiple blocks of data that are then stored onto several nodes in a cluster (128 Mb is default size of each block).

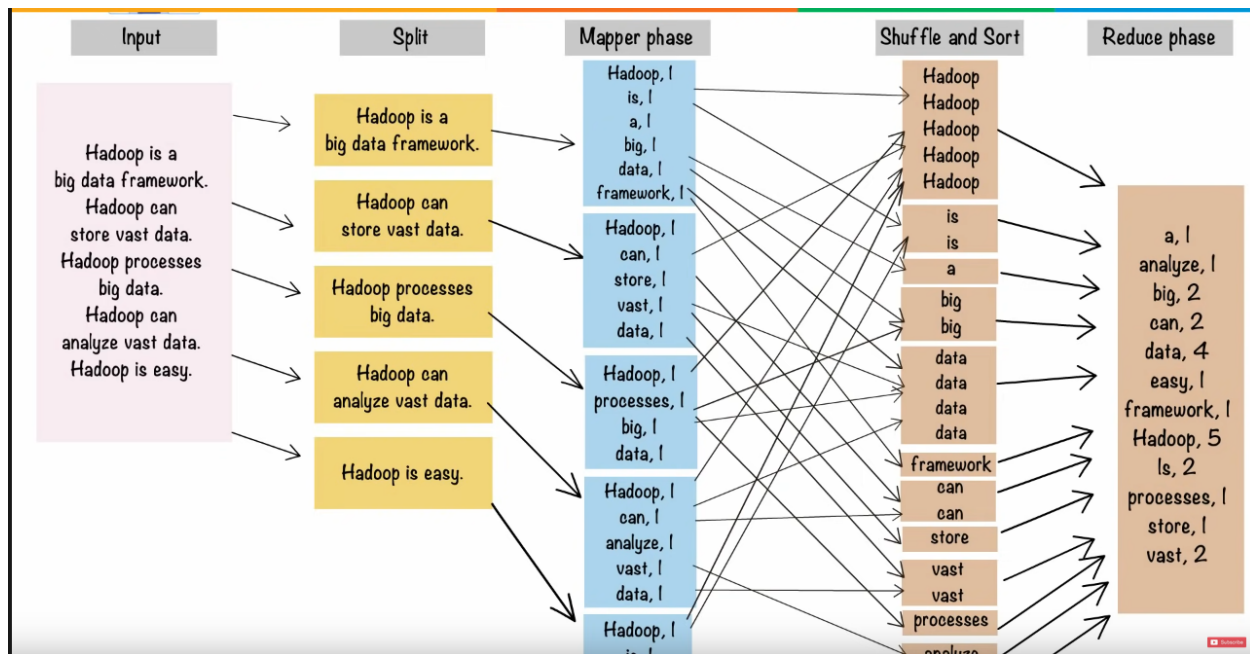
How HDFS Works - Writes



HDFS also makes multiple copies of the data and stores them across multiple systems. When a block is created, it is replicated with a replication factor of 3 (Replication Method)



- **MapReduce (Processing Unit)** : It splits Data into parts and processes them separately on different nodes, the individual results are then aggregated to give the final result.



- **YARN - Yet Another Resource Negotiator** : It consists of
 - Resource Manager : Assigns resources
 - Node Manager : Handle the nodes and monitor resource usage in the node
 - Application Master
 - Containers : Hold a collection of physical resources

How it works :

- The application Master requests the container from the node manager
- Node Manager sends the resources to Resource manager