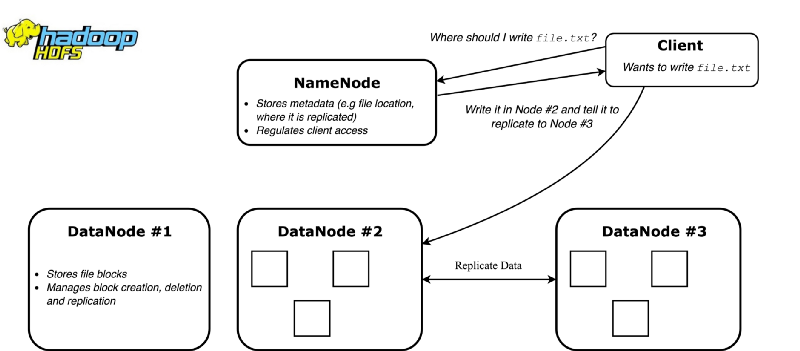
**HDFS**

Hadoop Distributed File System (HDFS) is the distributed file system used for distributed computing via the Hadoop framework. Boasting widespread adoption, it is used to store and replicate large files (GB or TB in size) across many machines.

Its architecture consists mainly of NameNodes and DataNodes. NameNodes are responsible for keeping metadata about the cluster, like which node contains which file blocks. They act as coordinators for the network by figuring out where best to store and replicate files, tracking the system’s health. DataNodes simply store files and execute commands like replicating a file, writing a new one and others.



Details: <https://www.youtube.com/watch?v=ECxlFPy7s-k&t=119s>