

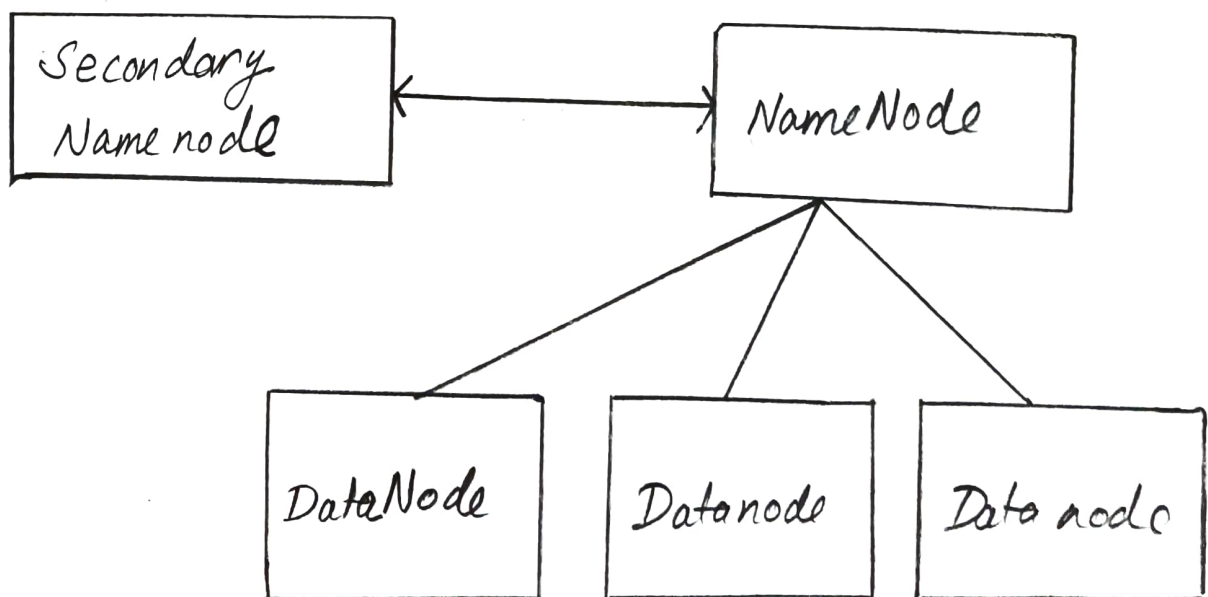
Hadoop distributed file system:- (HDFS)

- It is the storage component of Hadoop that stores data in the form of files.
- Each file is divided into blocks of 128 MB (configurable) and stores them on different machines in the cluster.

Components of HDFS:-

It has a master-slave architecture with two main components:

1. Name node
2. Data node



Name node is the master node and there is only one per cluster. Its task is to know where each block belonging to a file is lying in the cluster.

- Master daemon
- Maintains and manages Data nodes
- Records Metadata eg. location of block stored, the size of files, permissions, hierarchy etc.
- Receives heartbeat and block report from all the Data nodes.

Data node is the slave node that stores the blocks of data and there are more than one per cluster. Its task is to retrieve the data as and when required. It keeps in constant touch with the Namenode through heartbeats.

- slave daemons
- stores actual data
- serves read and write requests.

Rasool