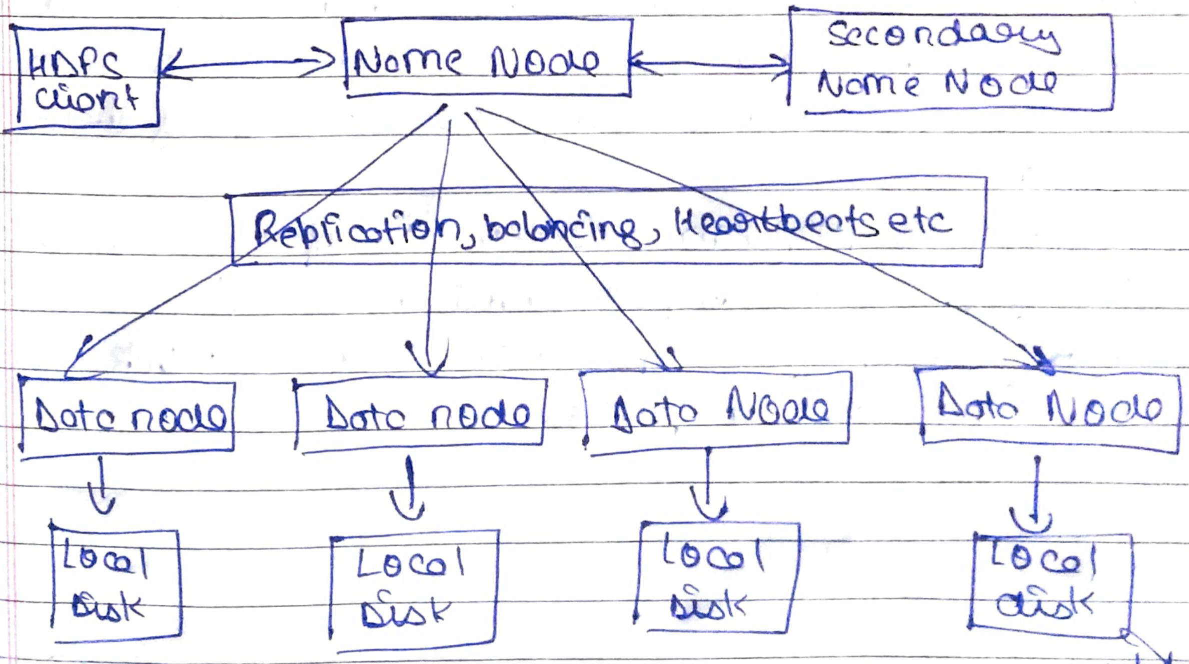
Amit Gusain
1918203
sec-K
Data Sci

Hadoop Fundamental

ans2 b-> HDFS is a distributed file system that handles large data sets running on commodity hardware. It is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes.
Hadoop File System was developed using distributed file system design. It is run on commodity hardware unlike other distributed system, HDFS is highly fault tolerant and designed using low-cost hardware.

HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored accross multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure.

Architecture of HDFS



Features of HDFS :

-> It is suitable for distributed storage and processing

-> Hadoop provides a command interface to interact with HDFS.

-> The built in server of namenode and datanode help users to easily check the status of cluster.

-> Streaming access to file system data.

-> HDFS provides file permission and authentication

HDFS has following component :

Namenode :- The namenode is the commodity hardware that contains GNU/Linux operating system and the namenode software. For every node in a cluster, there will be a datanode. It does following tasks :-

-> Manages the file system namespace

-> Regulates client's access to files

-> It also executes file system operation.

Datanode :- The data node is a commodity hardware having GNU/Linux OS and datanode software. These nodes manages the data storage of their system.

-> Datanodes perform read-write operations on the file system, as per client request

-> They also perform operation such as block creation deletion and replication.

"Failure comes only when we forget our ideals and objectives and principles." - Jawaharlal Nehru