Components of Hadoop 2.x

1. Namenode :

* It’s the high performing master node which maintains and manages all the datanodes
* It contains Hadoop file system tree and metadata information for files and directories
* Contains the in-memory mapping of blocks to datanodes
* Receives heartbeat and block reports from all the datanodes on the cluster

The namenode contains two important files on its hard disk which are

1. Fsimage which contains directory structure of HDFS, replication level of files, block info what constitutes files, access permissions of files and directories and also the access and modify times of the same and block size of the files etc
2. Edits – When a directory structure or a file gets modified, these modifications are stored in memory as well as in edits file on hard disk.
3. When the existing fsimage gets merged with edits, the updated fsimage file is obtained and this process is called checkpointing
4. Secondary namenode :

* This is not a back-up of name node.
* It does the house keeping activities for the namenode called checkpointing which has been explained above.
* Though this is not the back-up for namenode and not designed to act as one, this could work as namenode once it gets the latest edits file from the namenode and merges it with the fsimage it has and also receives blocks mapping from all the datanodes after its communication with them.

1. Datanode :

* Datanodes are commodity machines which store the actual data and act as slaves to the namenode
* They send periodic signals to the namenode to verify they are active and this is called heartbeat
* Sends block report to the namenode at the cluster start up and periodically as well at the 10th heartbeat
* They are the workhorse of the system, They receive instructions from namenode as to where and how to put the blocks.
* They perform all the block operation including periodic checksum.