IMPORTANCE OF NAMENODE IN CLUSTER

- The NameNode is the **centerpiece** of an HDFS file system. It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept.
- It does not store the data of these files itself. It stores the **METADATA** of the files stored on Hadoop.
- Client applications talk to the NameNode whenever they wish to locate a file, or when they want to add/copy/move/delete a file.
- The NameNode responds the successful requests by returning a list of relevant DataNode servers where the data lives.
- The NameNode is **a Single Point of Failure** for the HDFS Cluster. HDFS is not currently a High Availability system.
- When the NameNode goes down, the file system goes offline.
- There is an optional SecondaryNameNode that can be hosted on a separate machine. It only creates checkpoints of the namespace by merging the edits file into the fsimage file and does not provide any real redundancy.
- Thus, SecondaryNameNode can never act as NameNode.
- NameNode knows the list of the blocks and its location for any given file in HDFS. With this information NameNode knows how to construct the file from blocks.
- NameNode is so critical to HDFS and when the NameNode is down, HDFS/Hadoop cluster is inaccessible and considered down.

Hence NameNode is significant in hadoop cluster.