**ASSIGNMENT-2.4**

1. ***Importance of Name node in Hadoop cluster:***

* NameNode is the centerpiece of HDFS.
* NameNode is also known as the Master.
* It does not store the data of these files itself.
* Namenode is the node which stores the filesystem metadata i.e. which file maps to what block locations and which blocks are stored on which datanode.
* 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](https://wiki.apache.org/hadoop/DataNode) servers where the data lives.
* NameNode is usually configured with a lot of memory (RAM). Because the block locations are help in main memory.
* NameNode does not store the actual data or the dataset. The data itself is actually stored in the DataNodes.
* 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.
* An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.
* There are two types of NameNodes in a cluster:

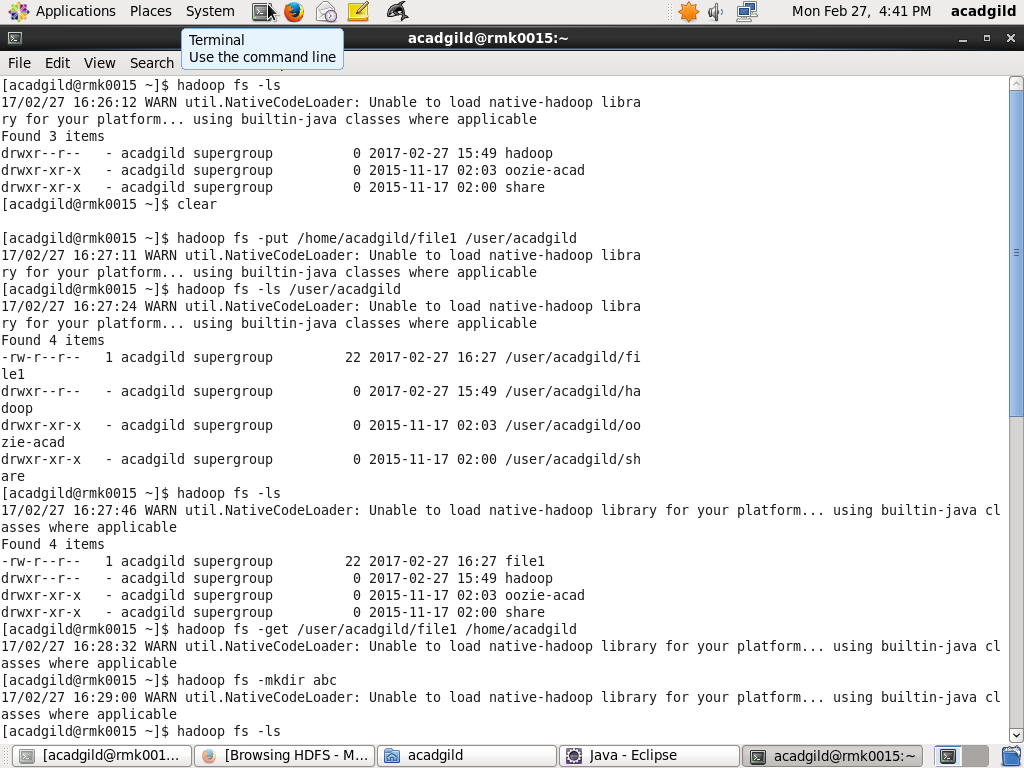
Active NameNode

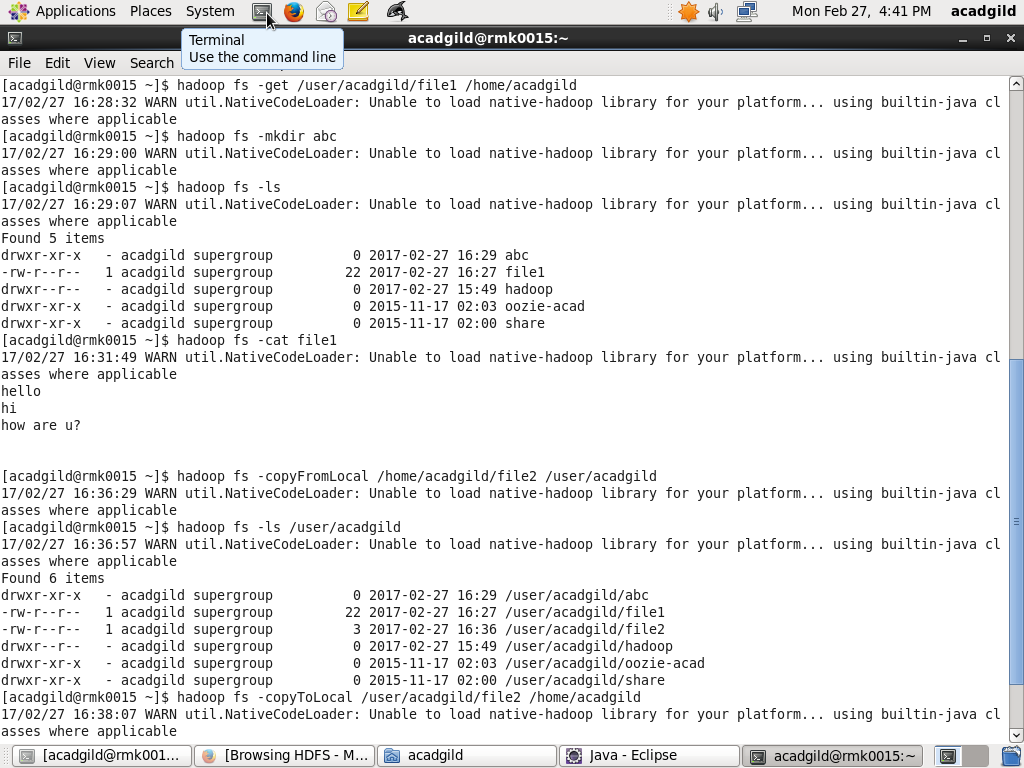
StandBy NameNode(Secondary NameNode)

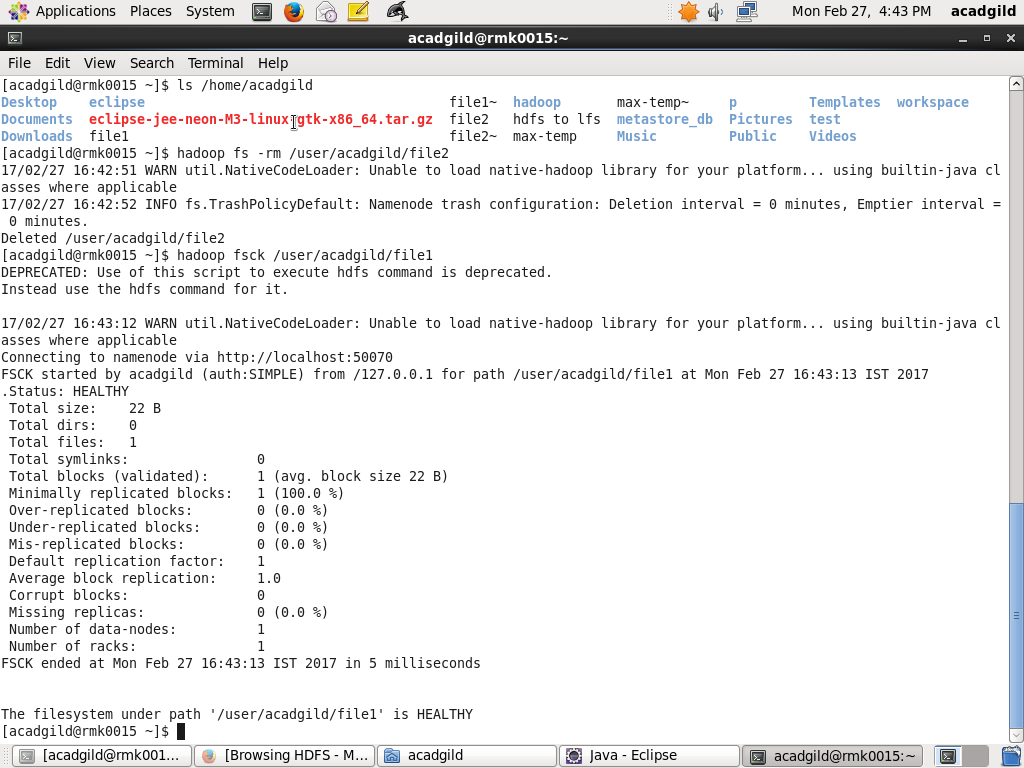
* The **Active NameNode** is responsible for all client operations in the cluster, while the **Standby NameNode** is simply acting as a slave
* NameNode is a single point of failure in **Hadoop** cluster.
* When the NameNode goes down, the file system will go offline. There is a secondary NameNode that can be hosted on a separate machine. So when active NameNode goes down, standby name node can be used as active NameNode and file system will be safe. This process is called failover.
* The namenode maintains two in-memory tables, one which maps the blocks to datanodes (one block maps to 3 datanodes for a replication value of 3) and a datanode to block number mapping.
* Whenever a datanode reports a disk corruption of a particular block, the first table gets updated and whenever a datanode is detected to be dead (because of a node/network failure) both the tables get updated.
* Failover semantics: The secondary namenode regularly connects to the primary namenode and keeps snap shooting the filesystem metadata into local/remote storage.

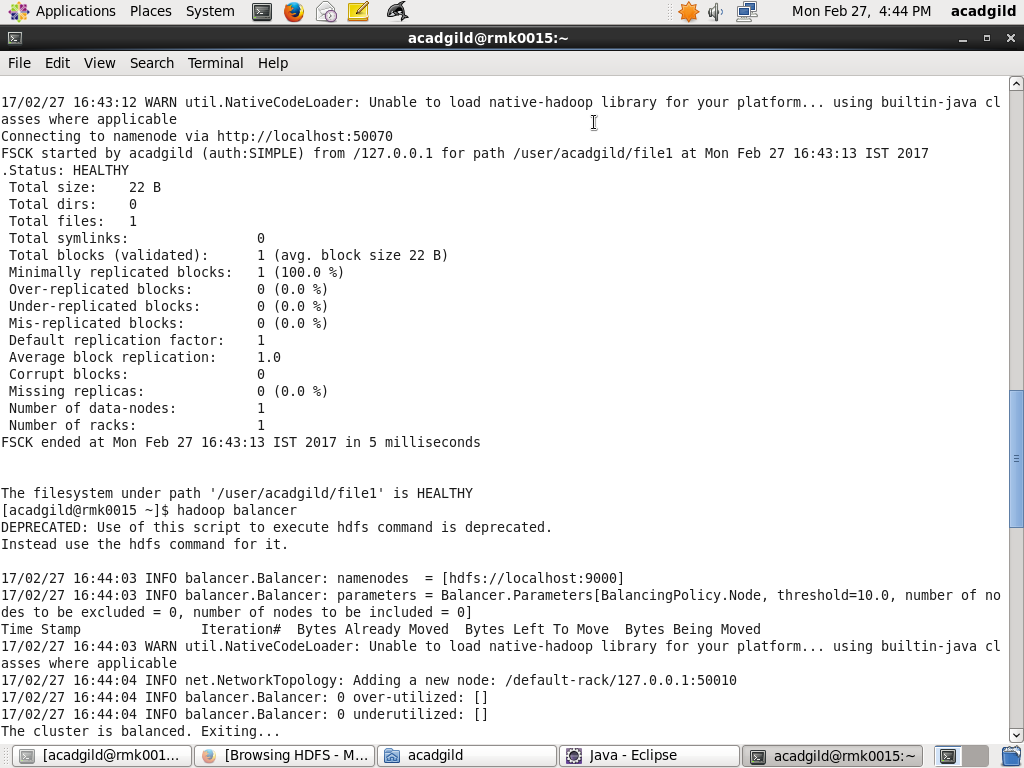
***Thus, NameNode is very important in hadoop cluster.***

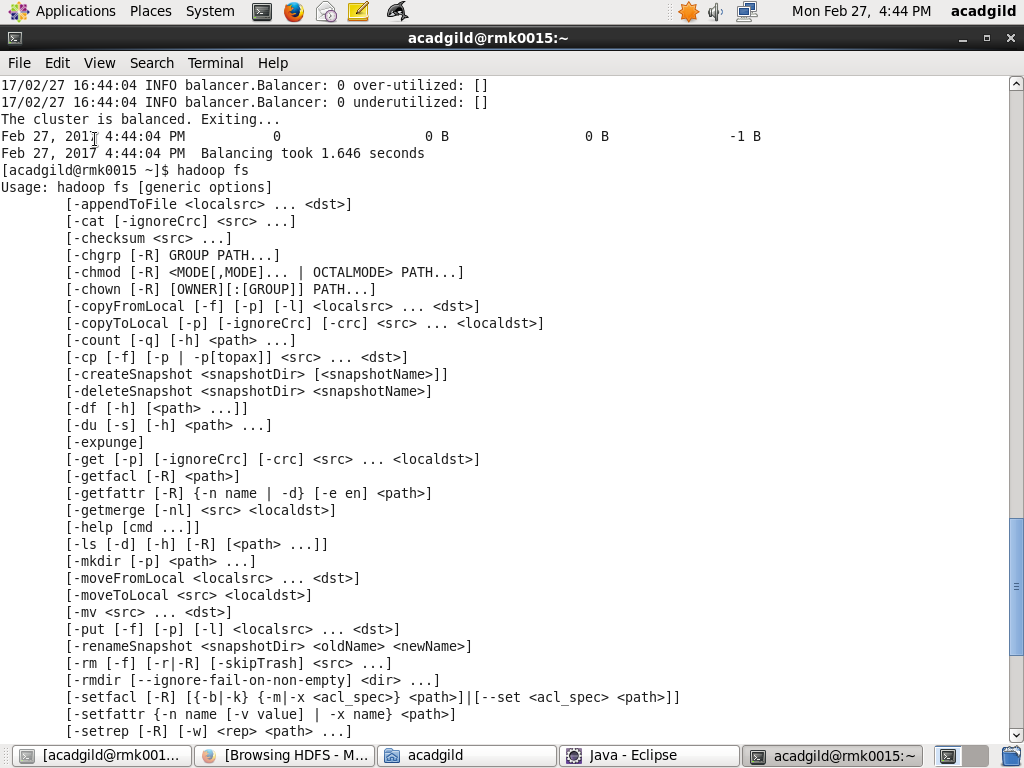
**2.Beginner commands for HDFS:**

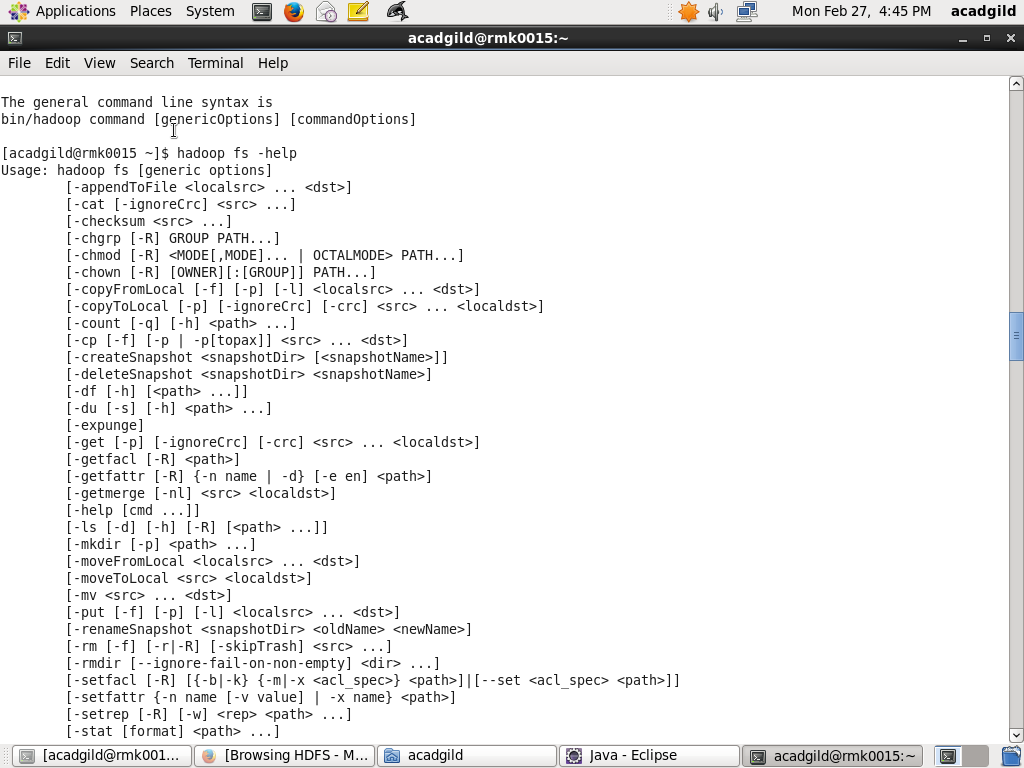
****

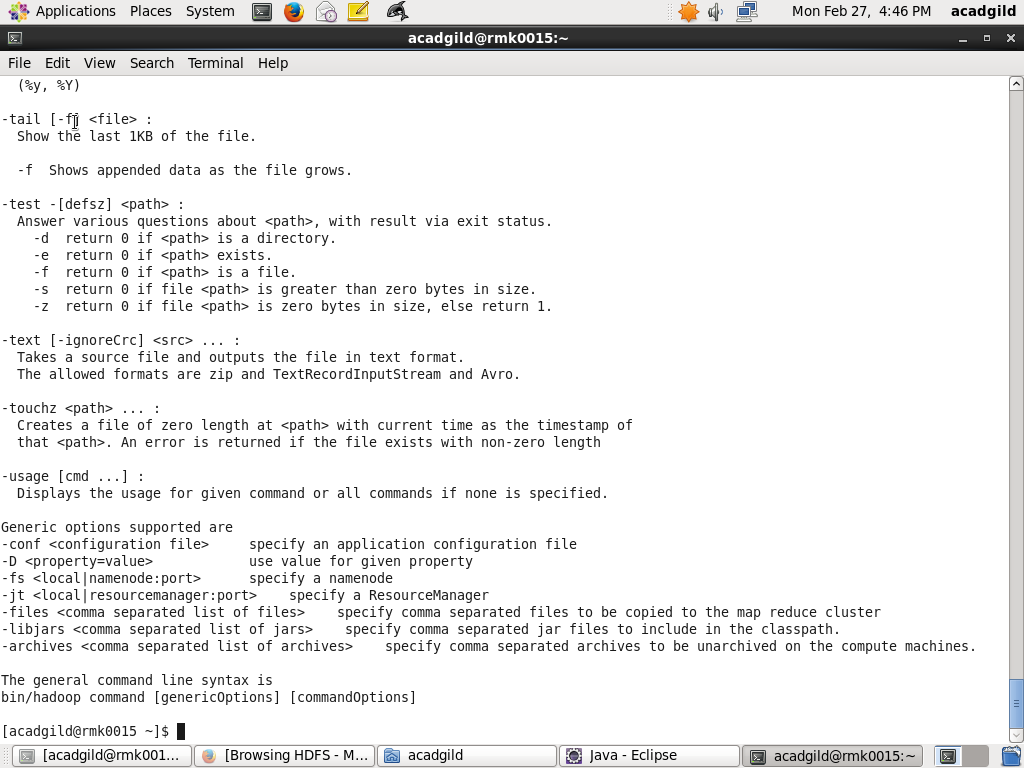
****

****

****

****

****

****