**ASSIGNMENT-3.6**

1. If 7TB is the available disk space per node (9 disks with 1 TB, 2 disk for operating system etc. were excluded.). Assuming initial data size is 600 TB. How will you estimate the number of data nodes (n)?

ANS:

The number of data nodes = available data size /disk space available per node

data size is 600 TB

available disk space per node is 7TB

Number of data nodes: = 600/7

                                    = 85

Therefore, Number of data nodes= 85

1. Imagine that you are uploading a file of 500MB into HDFS.100MB of data is successfully uploaded into HDFS and another client wants to read the uploaded data while the upload is still in progress. What will happen in such a scenario, will the 100 MB of data that is uploaded will it be displayed?

ANS:

No, the client will not be able to display the data until it completely finishes the process.

Process:

Initially, to upload a file, the client reaches the name node to locate the data node and to store the data and replicated copies. Once client is aware about the datanode, it will start copying. So, during this process if 1st block is written to HDFS and the next block has been started by the client to store then 1st block will be visible to readers. Only the current block being written will not be visible by the readers.