**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)?**

Number of data nodes: Data size available/disk space per node

                                     =600/7

                                     =85

Number of data nodes= 85

**2. 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?**

No,the data will not be displayed to the client but during the process 100 mb of data is available to the client.

For this 500 mb data, these data will be stored in 5 blocks and for each block there will be 3 replicas.Each block has 100 mb.

The client will take 100 mb block of data and will approach namenode for datanode location to store this block and the replicated copies.

Once client is aware about the datanode information, it will directly reach out to datanode and start copying this block of data which will be simultaneously replicated to other 2 datanodes.

Once the block is copied and replicated to the datanodes, client will get the confirmation about the Block storage and then, it will initiate the same process for next block.