**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)**

the number of data nodes: n= H/d = c\*r\*S/(1-i)\*d

H,Data size and d is diskspace

H=600TB d=7TB

n=H/d =600/7 = 85.7

around 86 data nodes are needed.

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?**

the block size = 100 MB

replicate 5 blocks three times.

\*There are 5 blocks for a file , client , name node and data node .

client consider the first block and will approach name node for data node location to store this block .

When the client is aware of datanode information it will goto to data node and start copying the first block .

Now , the client will get status of first block and if success it will initiate the same process for the second block .

client can read the 100 MB uploaded data and the remaining upload is still in progress.