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)?**

Let Available disk space denoted by d1=7 TB

Initial Data size ID=600 TB

Number of Data Nodes (n) = 600/7

= 85.7

Total number of data nodes required for 600 TB data size = 86 nodes.

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?**

Let assume we divide the 500 MB data file in 5 blocks of 100 MB.

Each 100 MB block is written on multiple data nodes and its replica is generated on multiple nodes.

If a client has written 100 MB of data file on a data node in a cluster then the client will write another 100 MB block into another data node in the cluster.

So this 100 MB block which is written first is available to read for the Client.

Hence the client can read the data block of 100 MB when write is in progress.