**1.FORMULA TO CALCULATE THE NUMBER OF DATA NODES REQUIRED:**

**N=H/d**

Where H**-**Hadoop Storage, d-disk space available per node.

|  |
| --- |
| **Solution :** |
|  |
| H=600TB d=7TB (Given) |
| n=H/d =600/7 = 85.7 = 86 (approx) |
|  |
| So, 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?  **SOLUTION:** |
|  |
| Assume the block size here as 128 MB. |
| First the client consider the first block of 128 MB and will approach the namenode to get the datanode location to store this block. |
| After getting the datanode info the client will start copying the first block to the data node. |
| Then the client will get acknowledgement on the first block and it will start the same process for the second block. |
| Thus The reader can read the 100 MB uploaded data while the remaining upload is still in progress. |