|  |
| --- |
| Cluster Planning |
|  |

|  |
| --- |
|  |



© Brain4ce Education Solutions Pvt. Ltd.

# Cluster Planning

As per the concepts covered today about cluster planning, layout a plan to address a data set of 8 to 16 TB

* How will the storage be distributed? How many data node you advice?

* As per recommendations, storage should be effectively utilized and data should be

evenly distributed. If total data accounts to 16 tb, we can consider totally storage required to

store 42 TB + 30% for local file system overhead

* Thus using 10 DN each of 6TB would be feasible
* How many disk drives per data node you will have?
* 3\*2TB or 1\*1TB disk drives would be optimum
* What is the replication factor taken into account?
* Replication factor =3
* What is the memory on data nodes and Namenode, that can address this large data set?
* RAM in Namenode = 32 GB ram
* RAM in Datanode = 8-16 GB ram
* What will be the Network bandwidth utilization, in case one node fails and the

data is replicated to another node?

* Network bandwidth depends on usage patter and workloads
* Min 1gb/sec - 10gb/sec would be good.(x3)