Module 5 : Hadoop Cluster: Planning and Managing

Assignment

edureka!

edureka!

Assignment

→ 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?

  » How many disk drives per data node you will have?

  » What is the replication factor taken into account?

  » What is the memory on data nodes and Namenode, that can address this large data set? What will be the Network bandwidth utilization, in case one node fails and the data is replicated to another node?