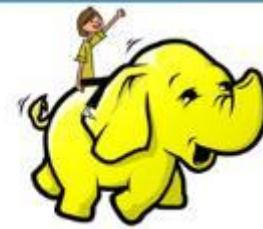


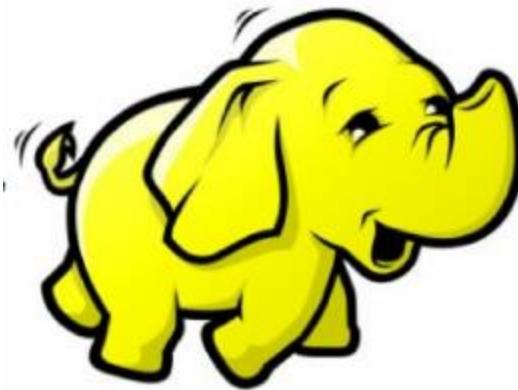
**edureka!**

Hadoop Administration



## Hadoop Administration

---



---

Module 3: Hadoop Cluster- Planning and Managing

## ✓ **Module 1**

- ✓ Understanding Big Data
- ✓ Hadoop Components

## ✓ **Module 2**

- ✓ Different Hadoop Server Roles
- ✓ Hadoop Cluster Configuration

## ✓ **Module 3**

- ✓ **Hadoop Cluster Planning**
- ✓ **Job Scheduling**

## ✓ **Module 4**

- ✓ Securing your Hadoop Cluster
- ✓ Backup and Recovery

## ✓ **Module 5**

- ✓ Hadoop 2.0 New Features
- ✓ HDFS High Availability

## ✓ **Module 6**

- ✓ Quorum Journal Manager (QJM)
- ✓ Hadoop 2.0 - YARN

## ✓ **Module 7**

- ✓ Oozie Workflow Scheduler
- ✓ Hive and Hbase Administration

## ✓ **Module 8**

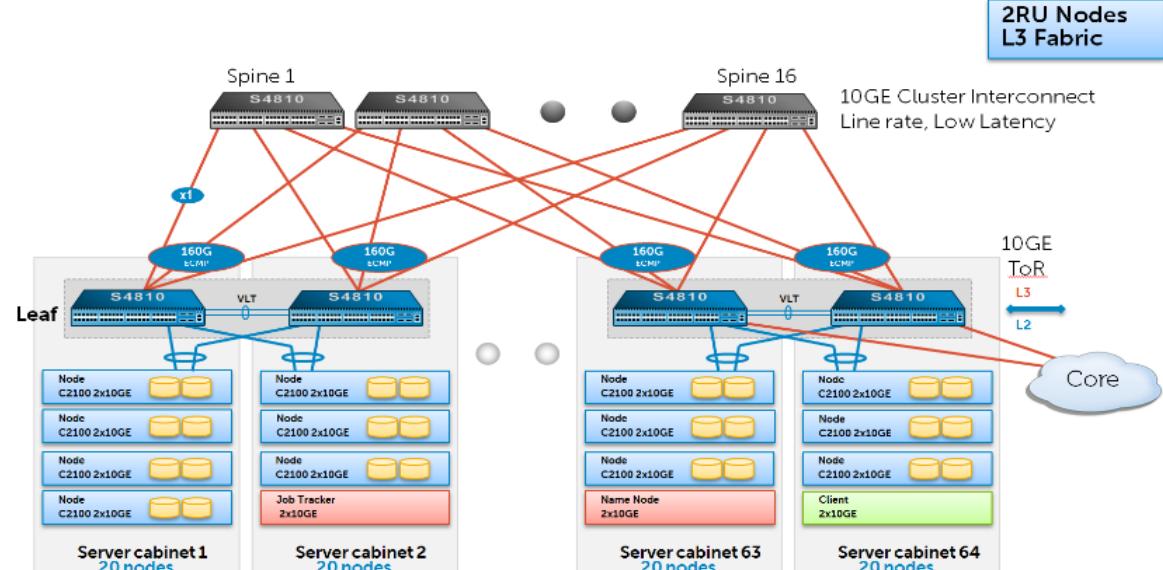
- ✓ Hadoop Cluster Case Study
- ✓ Hadoop Implementation

- ☛ **Let's Revise**
- ☛ **Plan your Hadoop Cluster: Hardware Considerations**
- ☛ **Cluster Network Configuration**
- ☛ **Plan your Hadoop Cluster: Software Considerations**
- ☛ **Popular Hadoop Distributions**
- ☛ **Monitoring Commands**
- ☛ **Schedulers**
- ☛ **FIFO Scheduler**
- ☛ **Fair Scheduler**
- ☛ **Configuring a Fair Scheduler**



- ✓ Hadoop Cluster Architecture
- ✓ Important Hadoop Configuration Files

## 10GE Hadoop Cluster – Scaled to 1280 nodes



- **2RU Nodes** 2 x 10GE (C2100)
- 20 Nodes per rack
- 64 racks, **1280 nodes**
- 2.5:1 oversubscription @ ToR
- **Expand as needed by adding Spine switches**
- Leaf QSFP+ optics, Spine SFP+ optics
- Leaf QSFP optical breakout cables
- 150m Leaf to Spine

<http://bradhedlund.com/2012/03/26/considering-10ge-hadoop-clusters-and-the-network/>

[www.edureka.in/hadoop-admin](http://www.edureka.in/hadoop-admin)



The data storage in Hadoop :

- a) Read Only
- b) Can be updated
- c) Is Write many, read once
- d) Is Write once, read many

Answer : Is Write once, Read many.





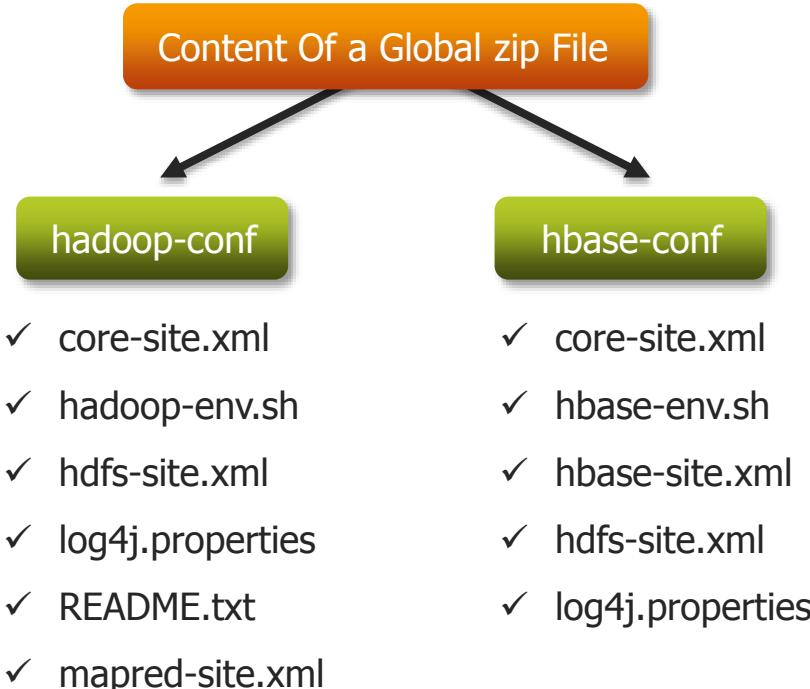
As the Hadoop administrator you can manually define the rack number of each slave Data Node in your cluster:

- a) True
- b) False

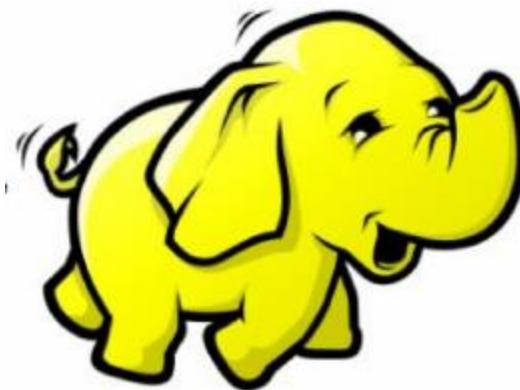
Answer : True



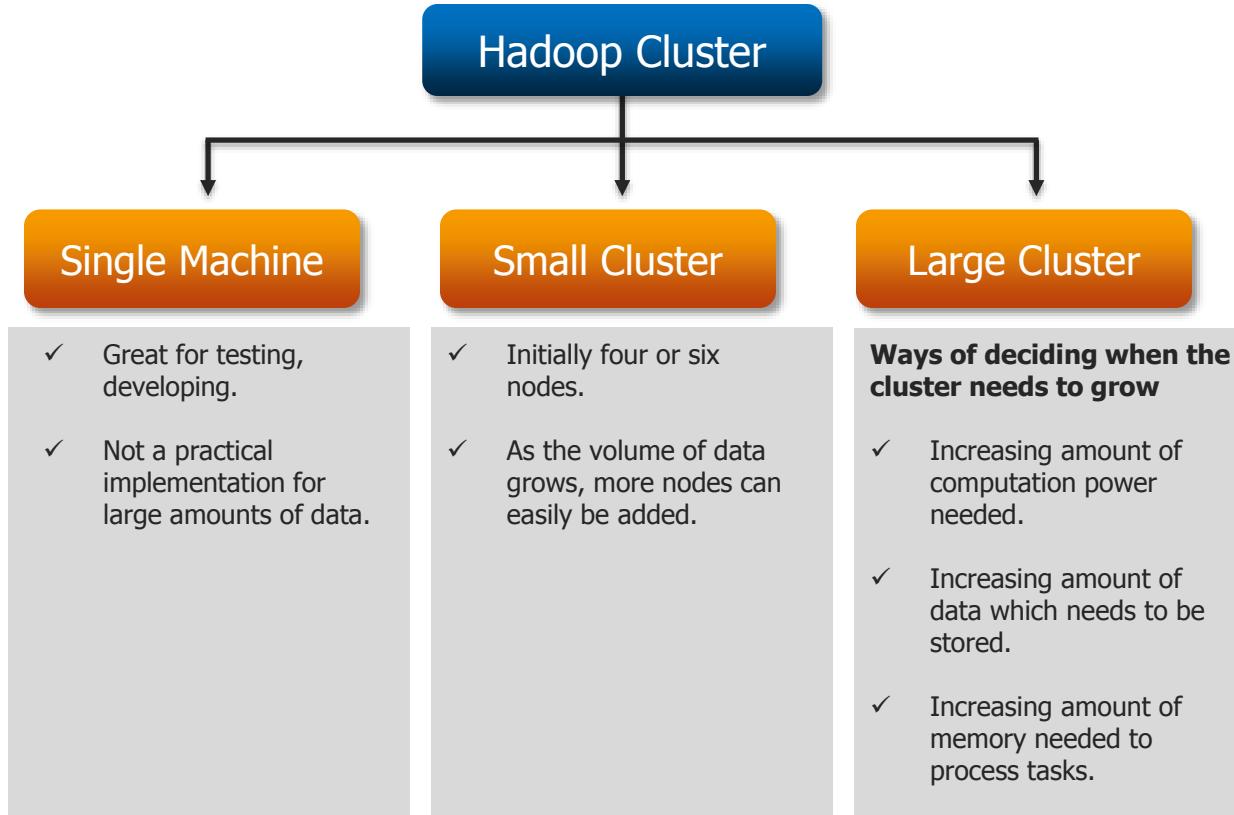
- ✓ To allow Hadoop client applications to work with the HDFS, MapReduce, and HBase services
- ✓ Create a zip file that contains all of the relevant configuration files with the settings for the cluster services such as HBase.
- ✓ Distribute the client configuration files to the users of a service.
- ✓ Either share the service-level client configuration files for the individual HDFS, MapReduce, and HBase services or create a global client configuration that contains client configuration files for all three types of services: [HDFS](#), [MapReduce](#), and [HBase](#).

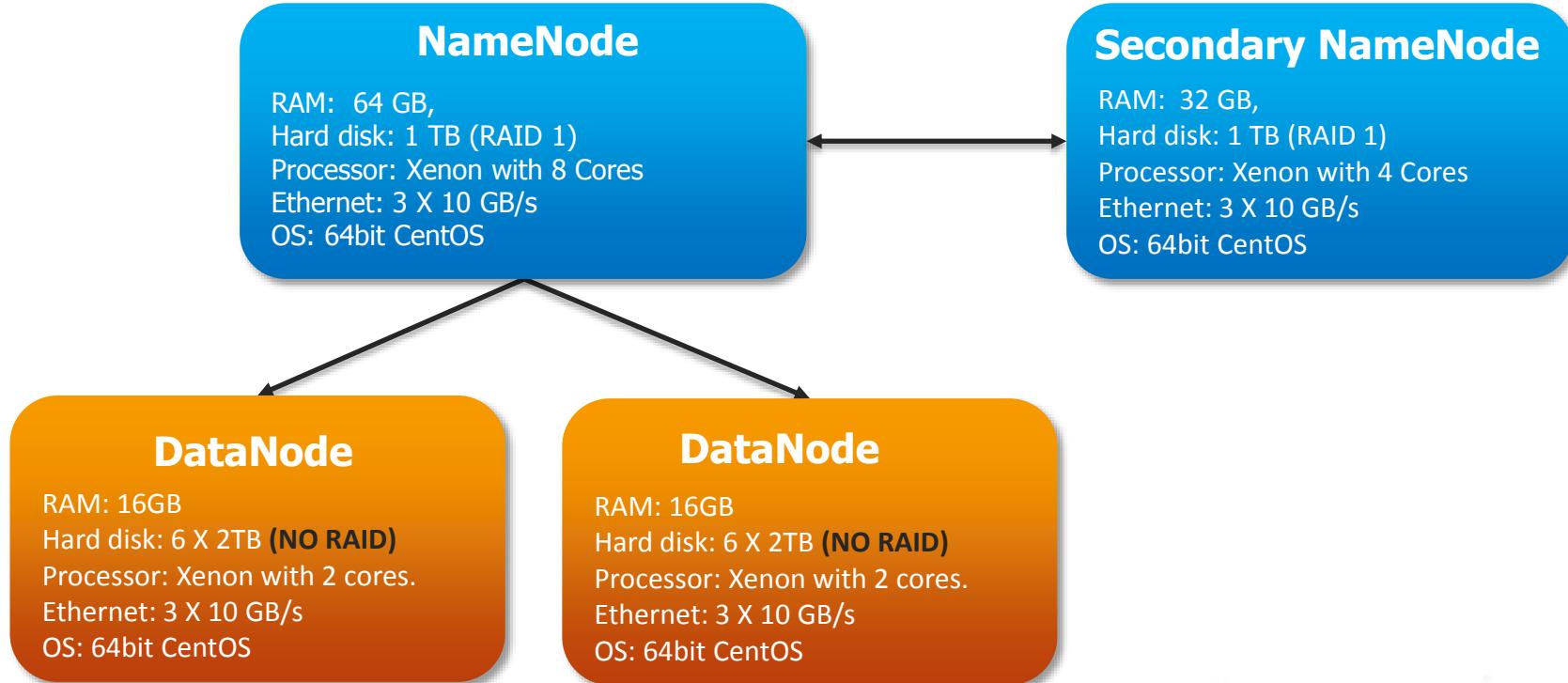


## Demo



Hadoop Client





## ✓ Master Hardware

### ✓ Namenode requirements

- ✓ RAM to fit metadata
- ✓ Modest but dedicated disk

### ✓ Secondary Namenode

- ✓ Almost identical to Namenode

### ✓ Resource Manager

- ✓ Retain Job Data, Memory Hungry
- ✓ Memory requirements can grow independent of cluster size

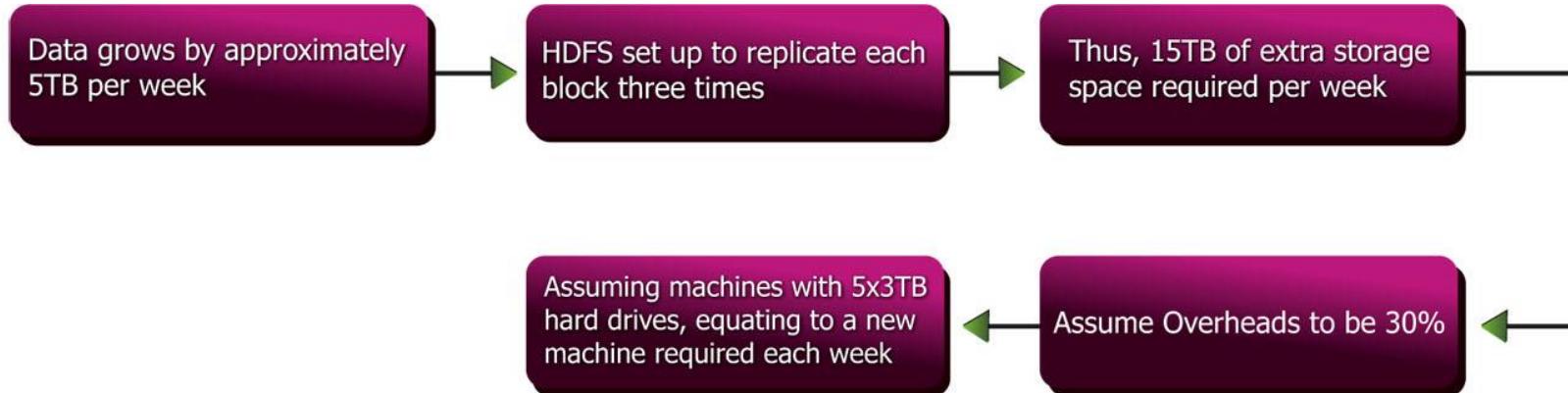
## ✓ Slave Hardware

- ✓ Storage
- ✓ Computation

### ✓ Cluster Sizing

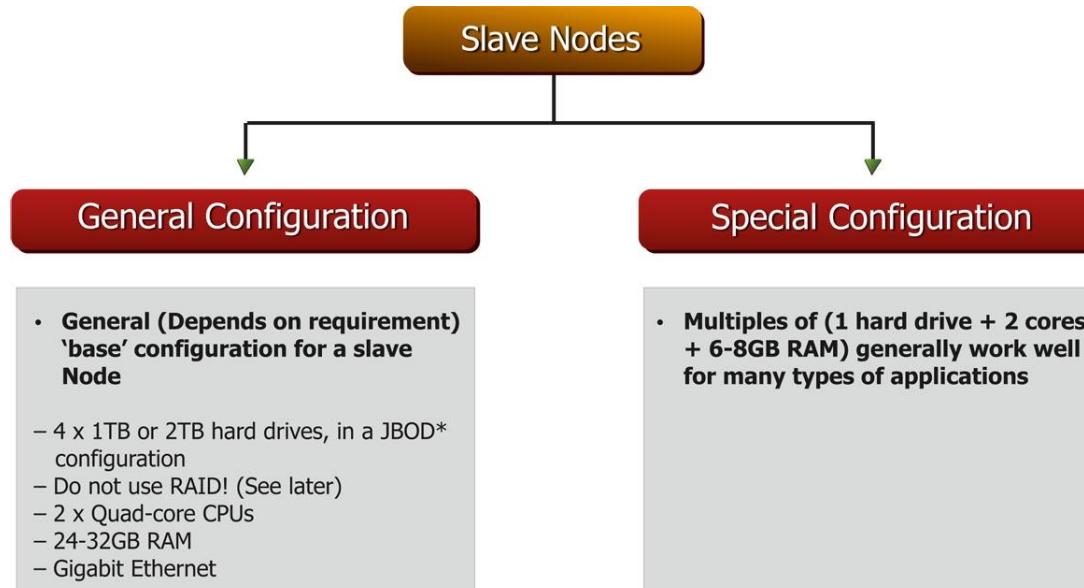
- ✓ Usage Pattern and Workloads
  - ✓ IO-bound or CPU-bound
- ✓ Consider requirements for additional components such as HBase

- ✓ **Seeking cluster growth on storage capacity is often a good method to use!**



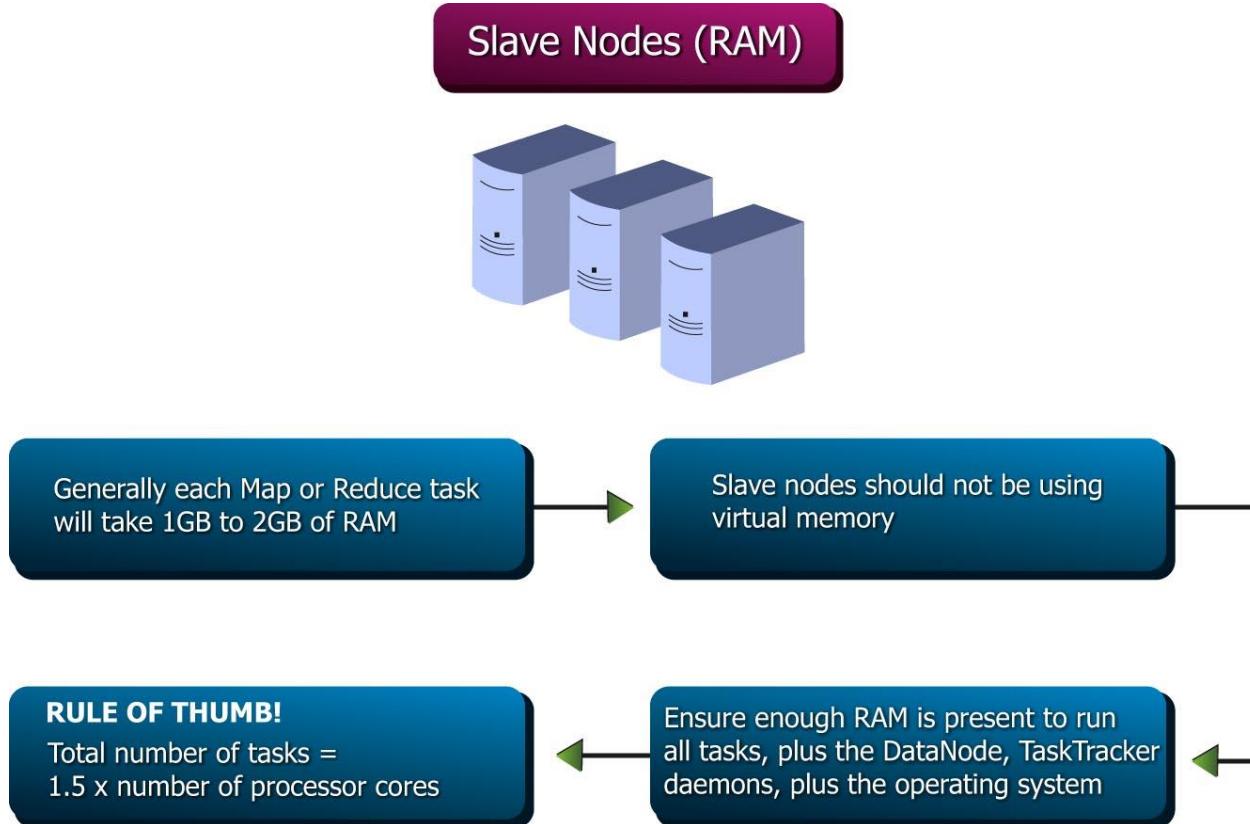
## Higher-performance vs lower performance components

Save the Money, Buy more Nodes!

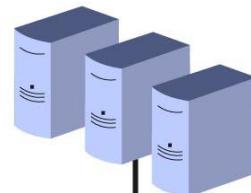


"A cluster with more nodes performs better than one with fewer, slightly faster nodes"

[www.edureka.in/hadoop-admin](http://www.edureka.in/hadoop-admin)



## Slave Nodes (Disk)



3.5" Disks

Faster, cheaper,  
higher capacity  
than 2.5" disks



8 x 1.5TB Drives

Different tasks  
are more likely  
to be accessing  
different disks

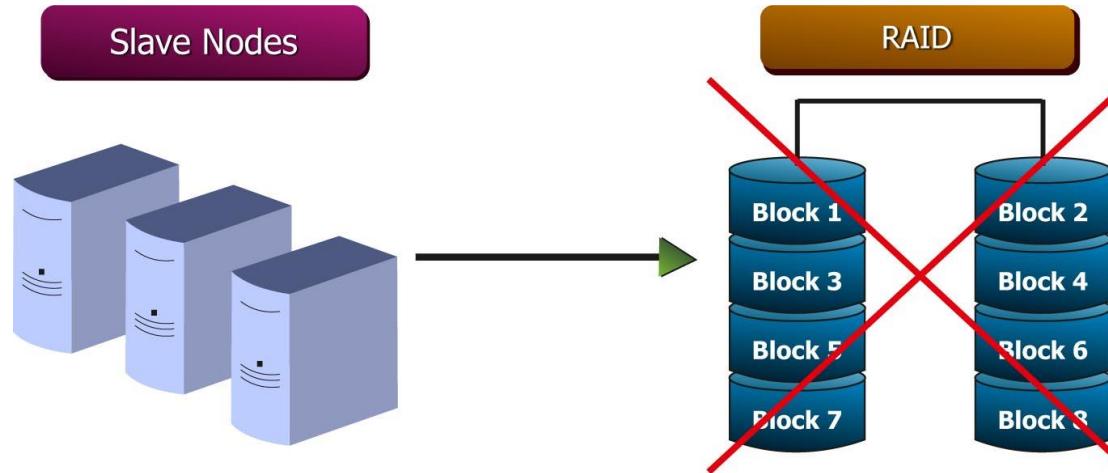


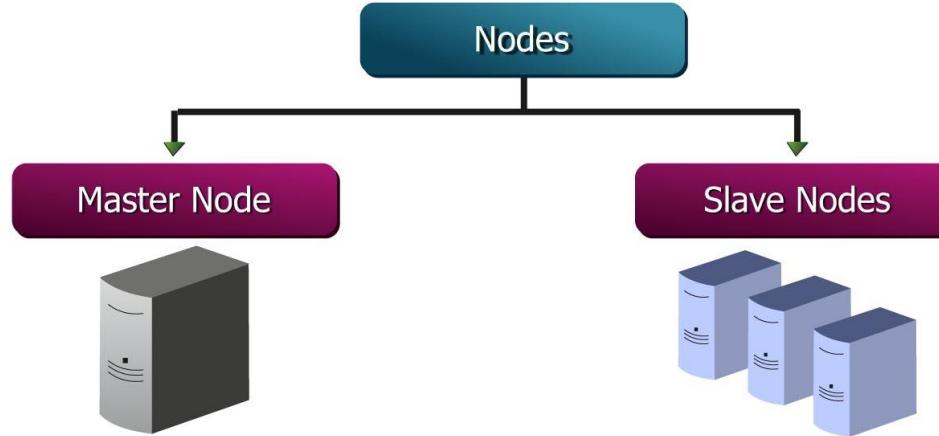
24TB per slave node

More than that will  
result in massive  
network traffic  
if a node dies and  
block re-replication  
must take place

- ✓ **Slave Nodes do not benefit from using RAID\* storage**

- ✓ HDFS provides redundancy by replicating blocks across multiple nodes
- ✓ RAID striping (RAID 0) is actually slower than the JBOD configuration used by HDFS
- ✓ RAID 0 read and write operations are limited by the speed of the slowest disk in the RAID array
- ✓ Disk operations on JBOD are independent, so the average speed is greater than that of the slowest disk



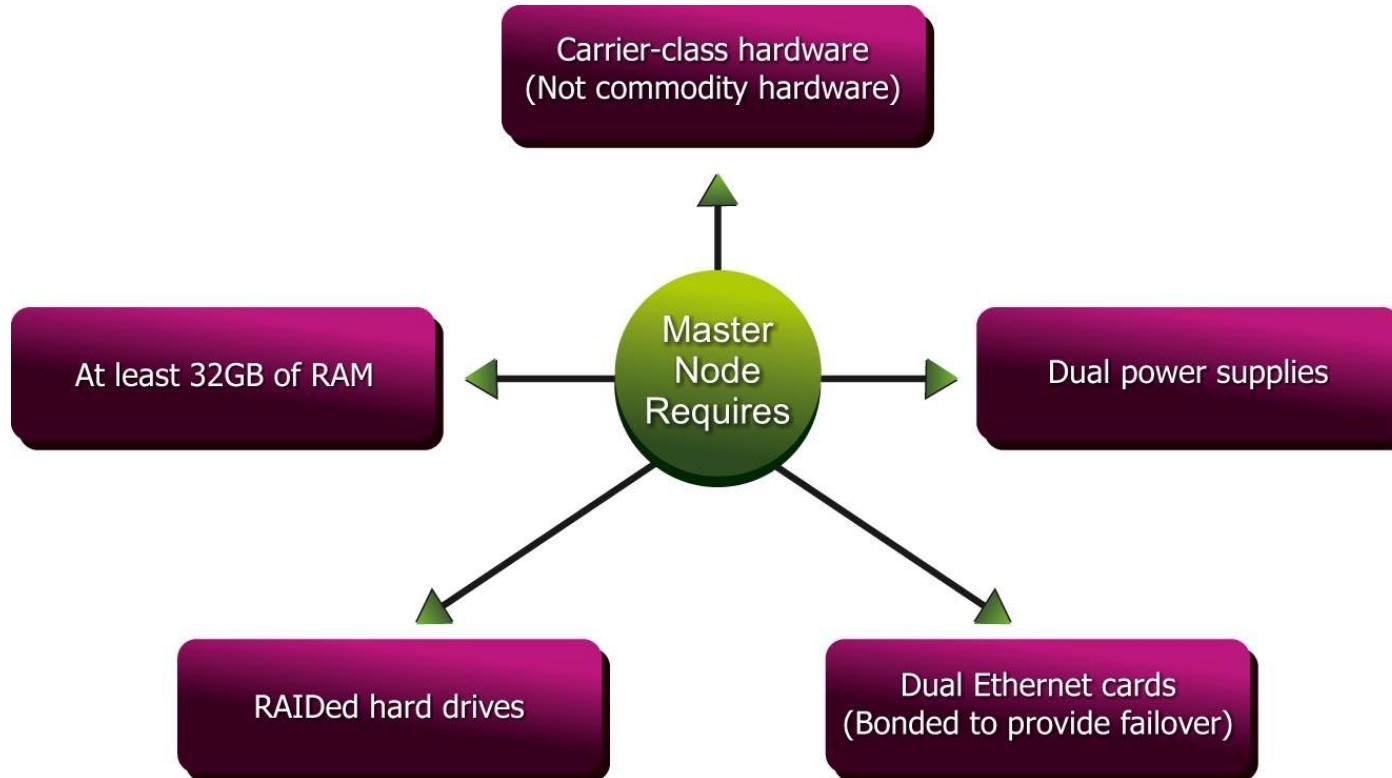


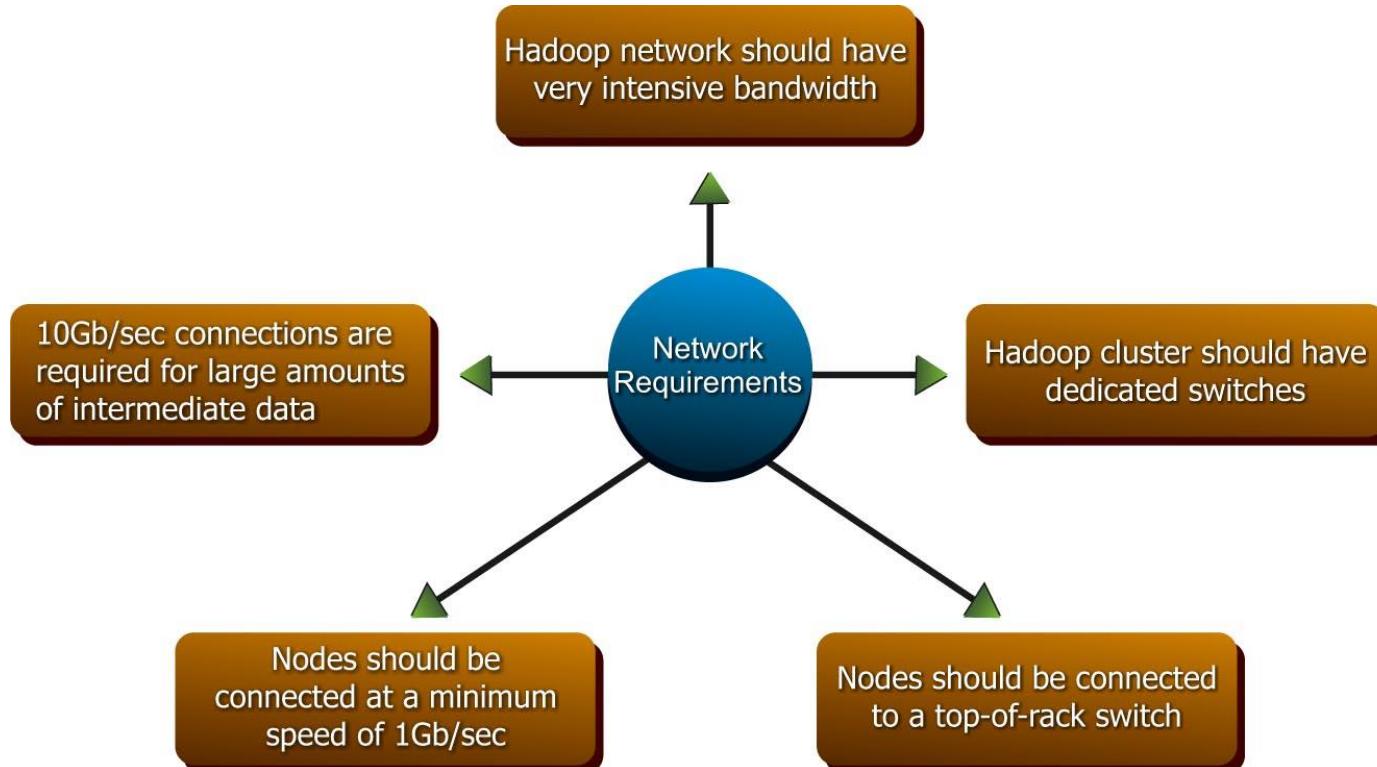
### **Master Node is a single point of failure**

- If the NameNode goes down, the cluster is inaccessible
- If the JobTracker goes down, no jobs can run on the cluster
- All currently running jobs will fail

### **Slave nodes are expected to fail at some point**

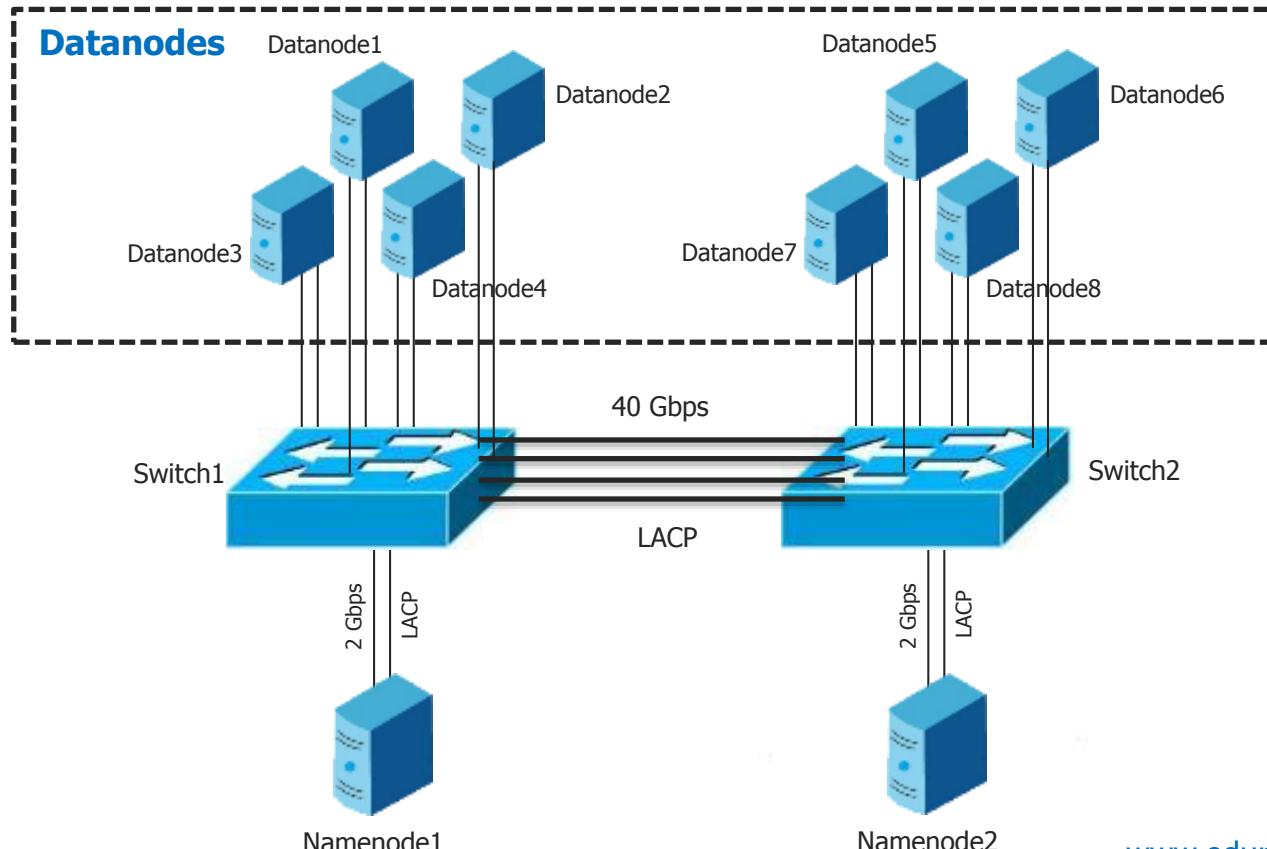
- This is an assumption built into Hadoop
- NameNode will automatically re-replicate blocks that were on the failed node to other nodes in the cluster, retaining the 3x replication requirement
- JobTracker will automatically re-assign tasks that were running on failed nodes





# Cluster Network Configuration

edureka!





A cluster with more nodes performs better than one with fewer, slightly faster nodes:

- a) True
- b) False

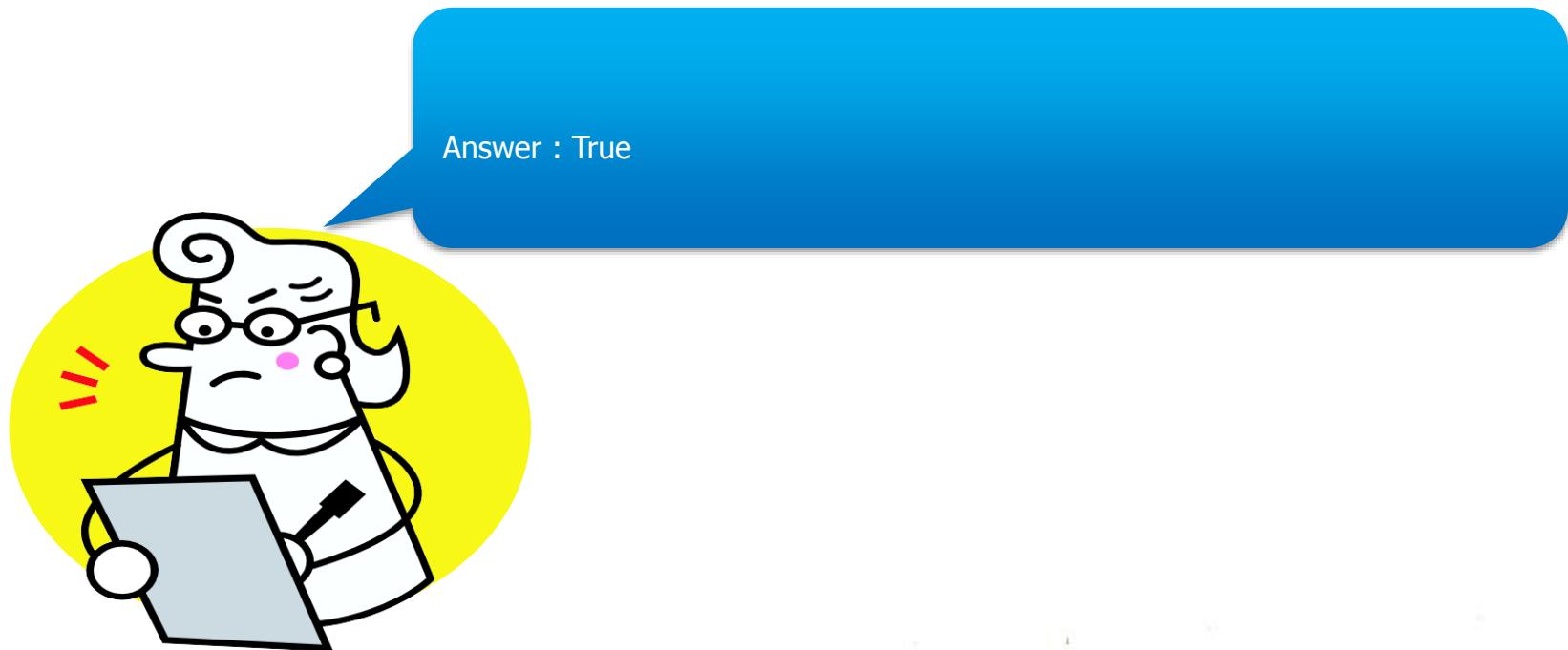
Answer : True





Slaves Nodes should not be using RAM memory:

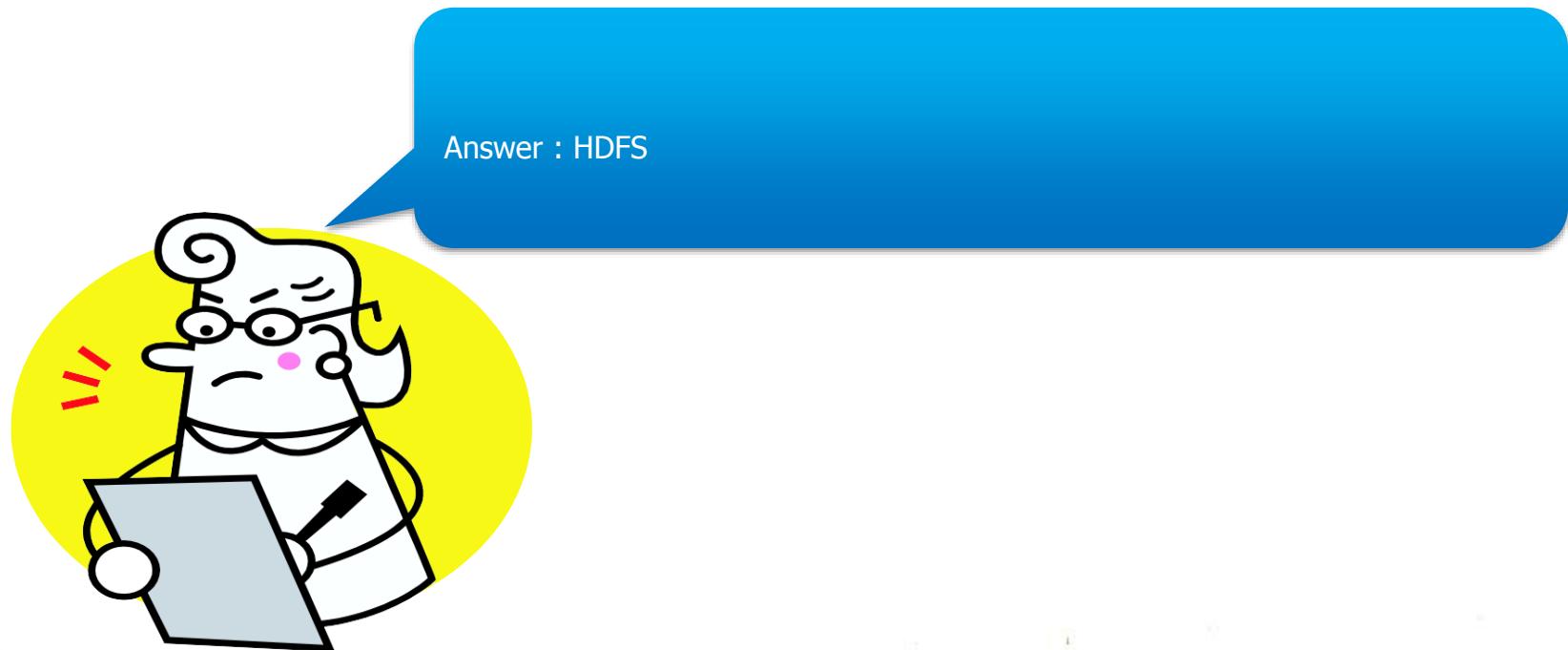
- a) True
- b) False





Which of these provides redundancy by replicating blocks across multiple nodes?

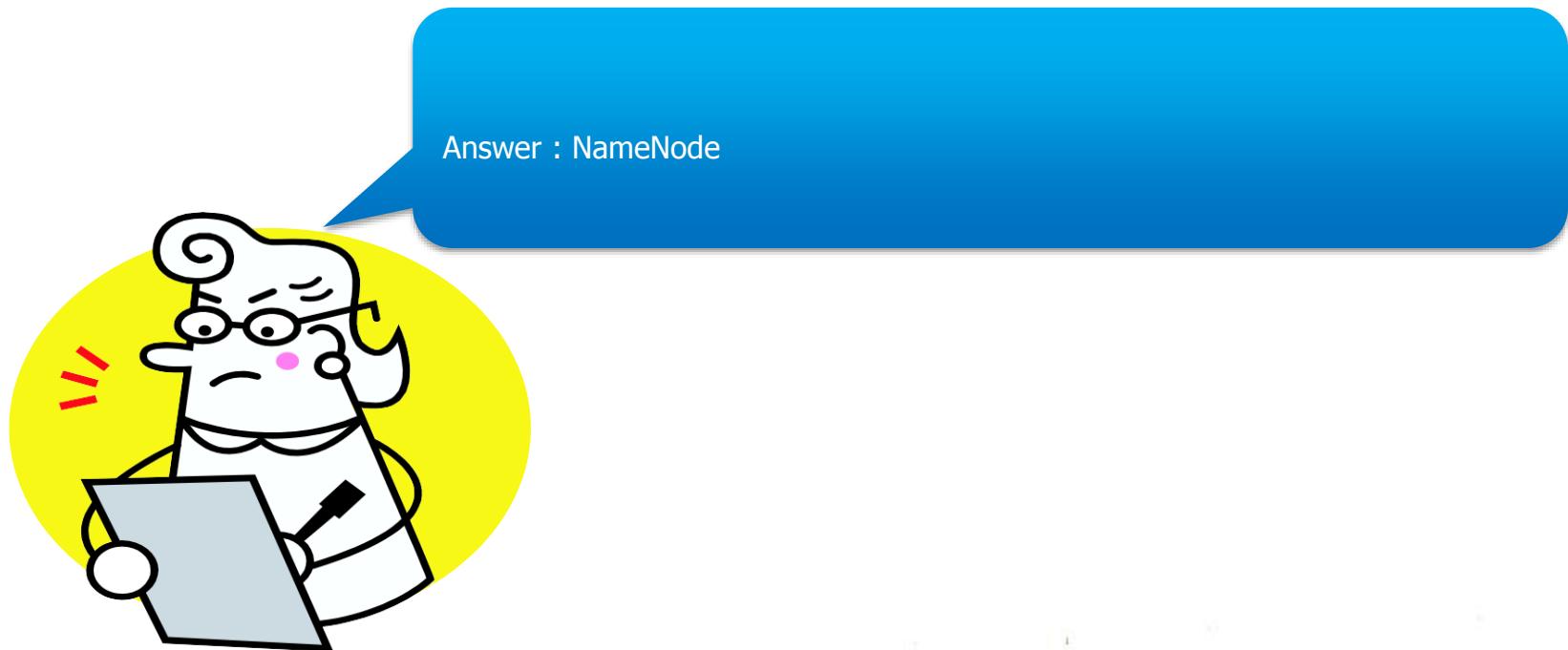
- a) HDFS
- b) NameNode
- c) DataNode

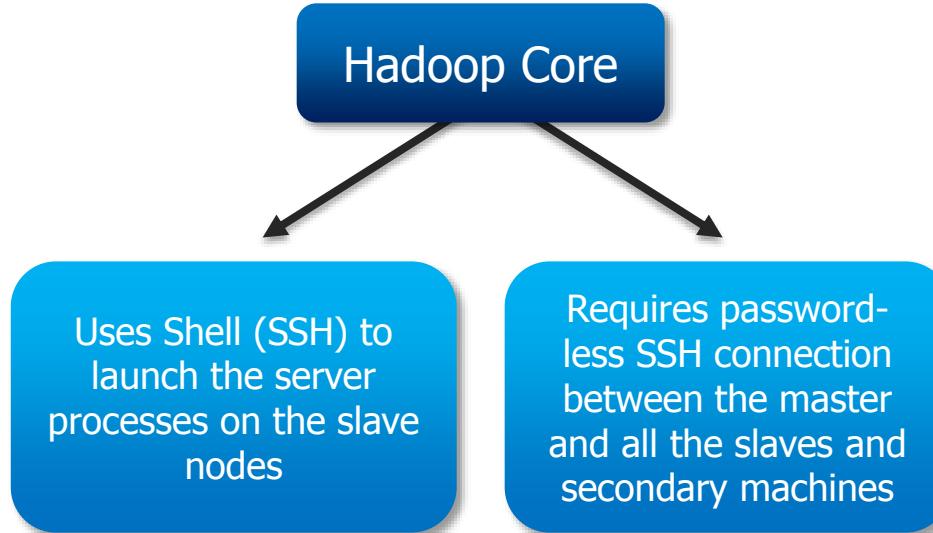




When a Client wants to retrieve a file from HDFS, it consults:

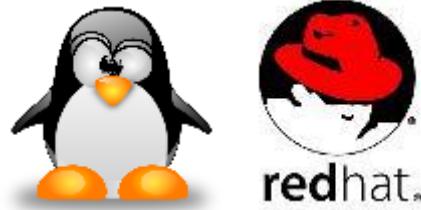
- a) DataNode
- b) NameNode
- c) Job Tracker





## ✓ Operating System

- ✓ Linux is the only production quality option today.
- ✓ A significant number run on RHEL.



## ✓ Java

- ✓ JDK - the most critical software
- ✓ List of tested JVMs:  
<http://wiki.apache.org/hadoop/HadoopJavaVersions>
- ✓ Java 1.6.x



## ✓ Operating System utilities

- ✓ ssh
- ✓ cron
- ✓ rsync
- ✓ ntp



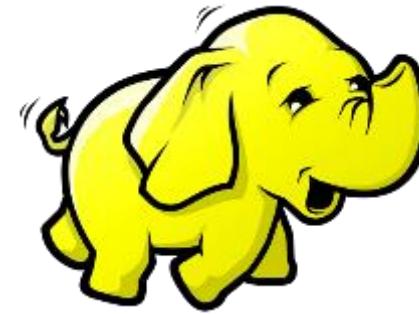
- ✓ **Choose a Distribution and Version of Hadoop**

- ✓ **Apache Hadoop**

- ✓ Complex Cluster setup
    - ✓ Manual install and Integration of Hadoop ecosystem components such as Pig, Hive, HBase etc
    - ✓ No commercial Support
    - ✓ Good for First try

- ✓ **Cloudera**

- ✓ Established distribution with many referenced deployments
    - ✓ Powerful tools for deployment, management and monitoring such as Cloudera Manager



cloudera

✓ **HortonWorks**

- ✓ Only distribution without any modification in Apache Hadoop
- ✓ HCatalog for metadata
- ✓ Stinger for Hive



✓ **MapR**

- ✓ Support native Unix filesystem
- ✓ HA features such as snapshots, mirroring or stateful failover
- ✓



✓ **Amazon Elastic Map Reduce (EMR)**

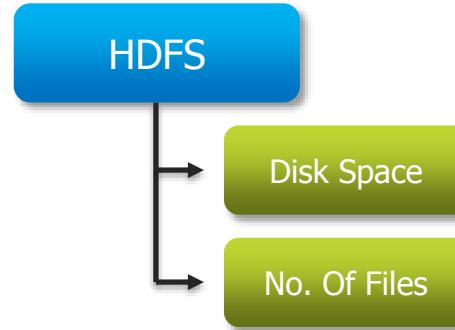
- ✓ Hosted Solution
- ✓ Only Pig and Hive are available as of now



Amazon Elastic  
Map Reduce (EMR)

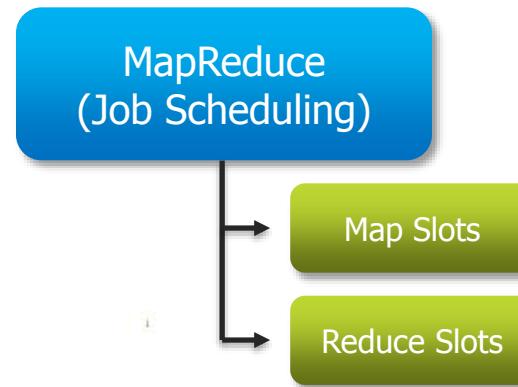
## ✓ Storage

- ✓ Disk Space
- ✓ No. Of Files



## ✓ Job scheduling

- ✓ Resource-aware Scheduling
- ✓ Data locality-aware Scheduling



To view applications running on the cluster

`hadoop job –list`

Displaying the Status of an Individual Job

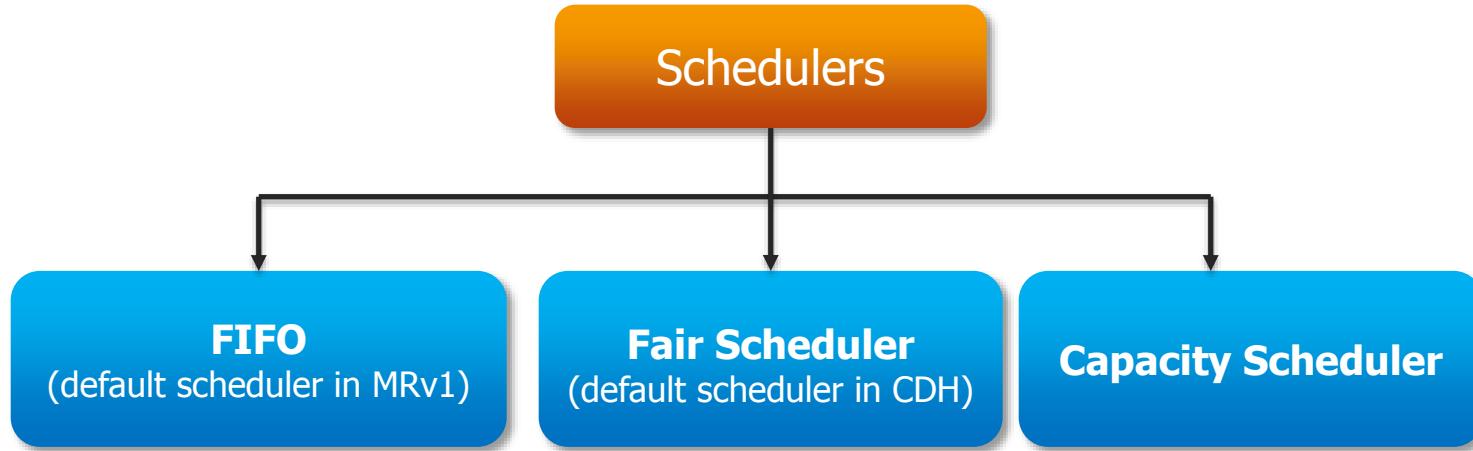
`hadoop job –status <job_id>`

To kill a job, use

`hadoop job –kill <job_id>`

<http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YarnCommands.html>  
[http://hadoop.apache.org/docs/stable/commands\\_manual.html](http://hadoop.apache.org/docs/stable/commands_manual.html)

- ✓ NameNode status: <http://localhost:50070/dfshealth.jsp>
- ✓ Job Tracker status: <http://localhost:50030/jobtracker.jsp>
- ✓ Task Tracker status: <http://localhost:50060/tasktracker.jsp>
- ✓ DataBlock Scanner Report: <http://localhost:50075/blockScannerReport>

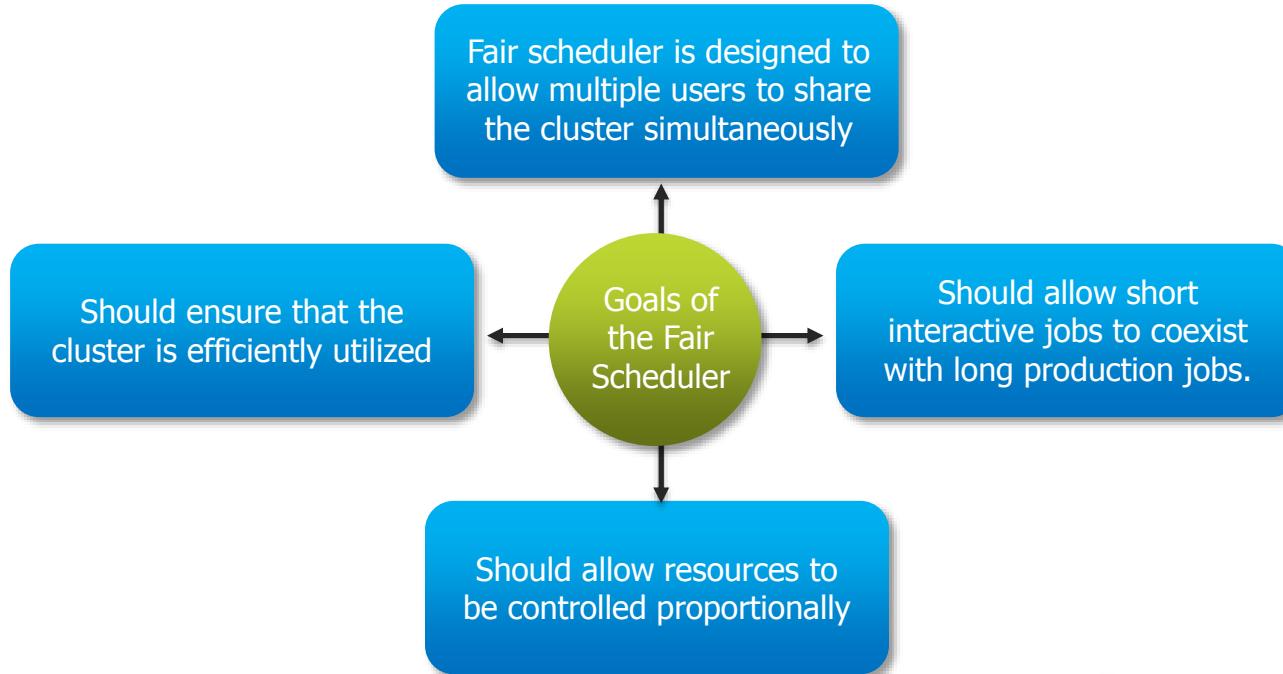


First In, First Out

Hadoop job –set-priority <job\_id> <priority>

All work in each queue is processed before moving on to the next.

## Resource-aware Scheduling



Stable Scheduler, Cloudera ship CDH with the Fair Share scheduler ON by default.

✓ **Each job is assigned to a pool.**

- ✓ Default assignment is one pool per username.

Total: 30 task slots

✓ **Jobs may be assigned to arbitrarily-named pools.**

- ✓ Such as "production".

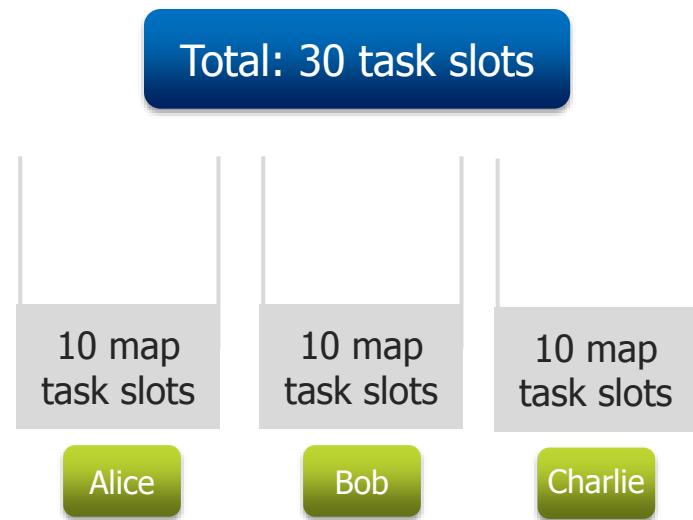
✓ **Physical slots are not bound to any specific pool.**

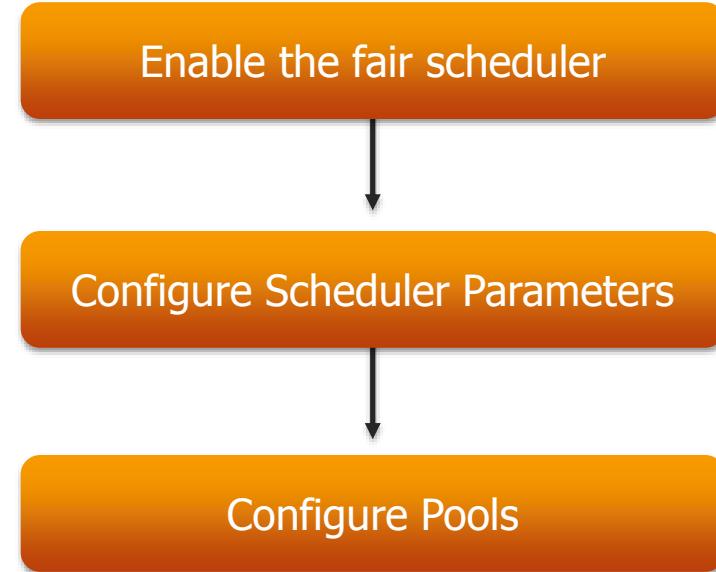
✓ **Each pool gets an even share of the available task slots.**



# Adding Pools Readjusts the Share of Slots

If client submits a job in a new pool, shares of slots are adjusted.





- ✓ In **mapred-site.xml** on the JobTracker, specify the scheduler to use:

```
<property>
<name>mapred.jobtracker.taskScheduler</name>
<value>org.apache.hadoop.mapred.FairScheduler</value>
</property>
```

- ✓ Identify the pool configuration file:

```
<property>
<name>mapred.fairscheduler.allocation.file</name>
<value>/etc/hadoop/conf/allocations.xml</value>
</property>
```

- ✓ The allocations configuration file must exist, and contain at least this:

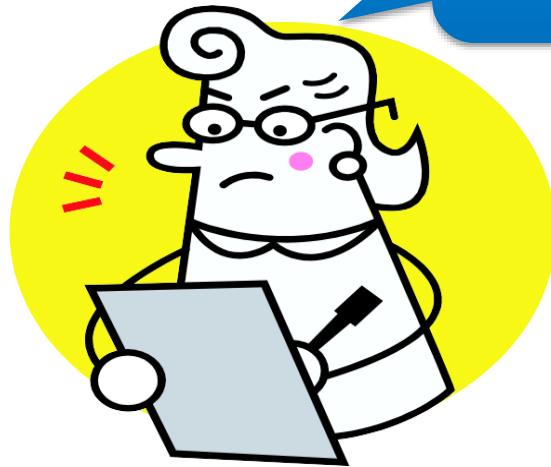
```
<?xml version="1.0"?>
<allocations>
</allocations>
```

```
<?xml version="1.0"?>
<allocations>
<userMaxJobsDefault>3</userMaxJobsDefault>
</allocations>
```

```
<?xml version="1.0"?>
<allocations>
<userMaxJobsDefault>3</userMaxJobsDefault>
<user name="bob">
<maxrunningJobs>6</maxrunningJobs>
</user>
</allocations>
```

```
<?xml version="1.0"?>
<allocations>
<userMaxJobsDefault>3</userMaxJobsDefault>
<pool name="production">
<minMaps>20</minMaps>
</minReduces>5</minReduces>
<weight>2.0</weight>
</pool>
</allocations>
```

Property, <schedulingMode>FAIR</schedulingMode> would use the Fair Scheduler.



Each job is assigned to a pool:

- a) True
- b) False

Answer : True





Tasks are scheduled by the:  
a) Job Tracker  
b) Task Tracker

Answer : Job Tracker



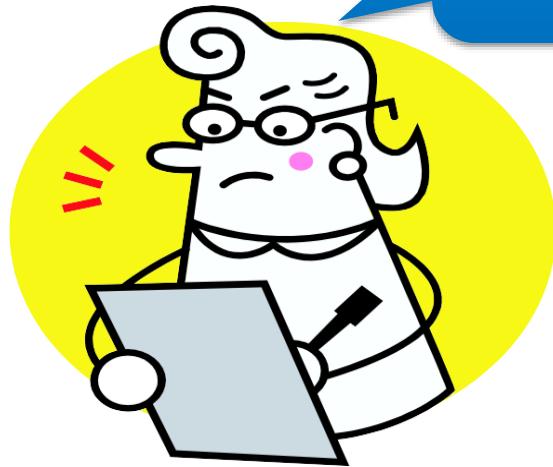


The schedulers can be defined in:

- a) hdfs-site.xml
- b) mapred-site.xml
- c) hadoop-env.sh

Answer : mapred-site.xml





Fair scheduler is designed to allow multiple users to share the cluster simultaneously:

- a) True
- b) False

Answer : True





- ✓ <http://hortonworks.com/blog/how-to-size-your-hadoop-cluster/>
- ✓ <http://hortonworks.com/resources/cluster-sizing-guide/>
- ✓ <https://blog.cloudera.com/blog/2013/08/how-to-select-the-right-hardware-for-your-new-hadoop-cluster/>

## Tasks for you



- 💡 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?

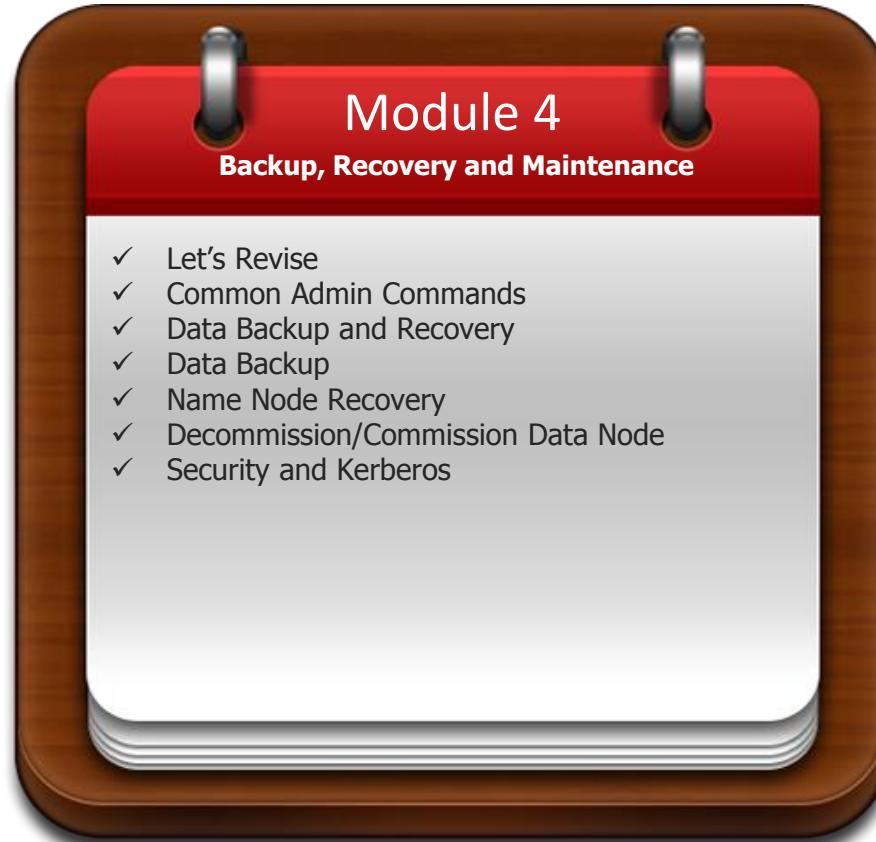




Go through this blog to know about commissioning and decommissioning of Data node

- <http://www.edureka.in/blog/commissioning-and-decommissioning-nodes-in-a-hadoop-cluster/>





**Module 3: Hadoop Cluster: Planning and Managing**

In this module, you will understand Planning and Managing a Hadoop Cluster, Hadoop Cluster Monitoring and Troubleshooting, Analysing logs, and Auditing. You will also understand Scheduling and Executing MapReduce Jobs, and different Schedulers.

 [Module 3 Recording](#)

 [Module 3 Presentation](#) [Download](#)

 [Configuration Files](#) [Download](#)

 [Hadoop Admin Assignment for Module 3](#) [Download](#)

**Recording of the Class** →  **Module 3 Recording**

**Presentation** →  **Module 3 Presentation**

**Configuration files** →  **Configuration Files**

**Assignment** →  **Hadoop Admin Assignment for Module 3**

Quiz

## Hadoop Admin Quiz for module 3 (8 Questions)

 30 MINUTES

This quiz is based on topics covered in Module-3; Planning the Hadoop Cluster, Cluster Size, Hardware and Software considerations, Managing and Scheduling Jobs, types of schedulers in Hadoop, Configuring the schedulers and run MapReduce jobs, Cluster Monitoring and Troubleshooting.

 Take Quiz

Pre-work

## Pre -work : Module 4 - Backup, Recovery and Maintenance

 Download

This document will help you to be prepared for the next class and understand the concept easily.

Further  
Reading

## Further Reading : Module 3 - Hadoop Cluster: Planning and Managing

 Download

This document contains links which will help you to know more about Hadoop Cluster: Planning and Managing.

**edureka!**

**Thank You**

See You in Class Next Week