**ASSIGNMENT-6.6**

1. Explain the difference between FIFO and Capacity scheduler

|  |  |
| --- | --- |
| FIFO scheduler  • The FIFO Scheduler places applications in a queue and runs them in the order of submission (first in, first out).  • Large applications will use all the resources in a cluster, so each application has to wait its turn.  • The FIFO Scheduler is not suitable for shared clusters. | Capacity scheduler  • In Capacity Scheduler, a separate dedicated queue allows the small job to start as soon as it is submitted.  • Large job finishes late when compared with using the FIFO Scheduler  •The Capacity Scheduler is designed to allow sharing a largecluster while giving each organization capacity guarantee. |

1. What are the limitations of hadoop 1.x and how they were overcome in hadoop 2.x?

***Limitations of Hadoop 1.x:***

* NameNode is a single point of failure, failure of namenode leads to inaccessible hadoop cluster.
* Secondary NameNode cannot act as NameNode.
* NameNode must be a high-end machine which is less prone to failure.

***Hadoop 2.x:***

* HDFS Federation
* In HDFS federation, there are multiple NameNodes, each storing metadata and block mapping of files and directories contained in particular sub-directories
* If one NameNode fails, the namespace volume managed by the other NameNode is still accessible. So that the entire cluster doesn’t become inaccessible
* HDFS High Availability
* The NameNodes must use highly-available shared storage to share the edit log. Edit logs are read by Standby NameNode when it takes the responsibility of the Active NameNode.
* In the event of the failure of the active NameNode, the standby takes over its duties without a significant interruption.