**Explain the difference between FIFO and Capacity scheduler**

|  |  |
| --- | --- |
| FIFO scheduler  1. The FIFO Scheduler places applications in a queue and runs them in the order of submission (first in, first out) which means application handled by the queue order. The first one runs first and then remaining runs as they are in the order.  2.Each application has to wait in the queue because Large applications will use all the resources in a cluster.    3. It is not suitable for shared clusters. | Capacity scheduler  1.In Capacity Scheduler, a individual queue is allocated to prefer the smaller jobs to complete first.  2. It takes more time to complete the large jobs when compared with the FIFO Scheduler  3.The Capacity Scheduler is suitable for sharing clusters. |

**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.