1. Explain the difference between FIFO and Capacity scheduler

|  |  |
| --- | --- |
| FIFO | Capacity scheduler |
| * The FIFO Scheduler has the merit of being simple to understand and not needing any configuration, but it’s not suitable for shared clusters. | * The Capacity Scheduler is designed to allow sharing a large cluster while giving each organization a minimum capacity guarantee. |
| * The FIFO Scheduler places applications in a queue and runs them in the order of submission (first in, first out). | * With the Capacity Scheduler, a separate dedicated queue allows the small job to start as soon as it is submitted. |
| * Fast execution i.e time taken for its execution is less. | * Large job finishes late when compared with using the FIFO Scheduler. |
| * Large applications will use all the resources in a cluster, so each application has to wait its turn ,so it does not work well on shared clusters. | * It was mainly designed to allow sharing a large cluster. |

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

* LIMITATIONS OF HADOOP 1.X OVERCOME IN HADOOP 2.X:
* Hadoop 1.x supports only one namespace for managing HDFS file system whereas Hadoop 2.x supports multiple namespaces.
* Hadoop 1.x supports only one programming model,MapReduce while Hadoop 2.x supports multiple programming models with YARN Component like Spark, Storm etc.
* Hadoop 1.x has lot of limitations in Scalability,while Hadoop 2.x has overcome that limitation with new architecture so as to increase scalability.
* In Hadoop 1.x ,HDFS uses fixed-size Slots mechanism for storage purpose whereas Hadoop 2.x uses variable-sized Containers.
* Hadoop 1.x supports maximum 4,000 nodes per cluster where Hadoop 2.x supports more than 10,000 nodes per cluster.
* In Hadoop 1.x, a map tasks can not run on reduce slot. So cluster utilization is low while in Hadoop 2.x, resources are dynamic and fine grained, this leads to better cluster utilization.