1. Explain the difference between FIFO and Capacity scheduler

2. Explain the difference between FIFO and Fair scheduler

3. Explain the difference between Capacity and Fair scheduler

There are different schedulers 1) FIFO Scheduler 2) Capacity Scheduler 3) Fair Scheduler.

Fair Scheduler - Fair scheduling is a method of assigning resources to applications such that all apps get, on average, an equal share of resources over time.

FIFO Scheduler – Scheduling is done according to the sequence of the request.

Capacity Scheduler - The Capacity Scheduler is designed to allow sharing a large cluster while giving each organization a minimum capacity guarantee.

FIFO Scheduler – Will execute every order in order, so large applications will take lots of time and the small applications will have to wait for longer period of time but the resources are fully used in this case.

Capacity Scheduler – Here the large applications will be queued differently than that of the small applications. Recourses will also be allotted differently. Here the problem is that, it may happen that only one queue may be crowded and the other queue may be free. Hence the efficiency is not 100 percent. Basically the scheduling is done according to the capacity.

Fair Scheduler- Fair scheduler is works according to the resources needed for the application and also the number of application. When there is a single app running, that app uses the entire cluster. When other apps are submitted, resources that free up are assigned to the new apps, so that each app eventually on gets roughly the same amount of resources. Fair Scheduler which forms a queue of apps, this lets short apps finish in shorter period of time.

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

Ans -

Hadoop 1.x also called as hadoop 1 where HDFS (Resource management and scheduling) and MapReduce(Programming Framework) hence non batch

applications cannot be run on hadoop 1. It can support only single name node. So it does no provide system availability and scalability.

Hadoop 2.x

- hadoop decouples the responsibilities into different components.

-This is done by YARN(Yet Another Resource Manager)

-By decoupling component’s responsibilities, it supports multiple namespace, Multi-tenancy, Higher Availability and Higher Scalability.

Hadoop 2.x has following benefits-

1) Highly Scalability.

2)Highly Availability.

3)Supports Multiple Programming Models.

4)Supports Multi-Tenancy.

5)Supports Multiple Namespaces.

6)Improved Cluster Utilization.

7)Supports Horizontal Scalability.

MRv2 (aka Hadoop 2) in this version of hadoop the resource management and scheduling tasks are separated from Map Reduce which is separated by YARN(Yet Another Resource Negotiator). The resource management and scheduling layer lies beneath the Map Reduce layer.

It also provides high system availability and scalability as we can create redundant Name Nodes.

The new feature of snapshot through which we can take backup of file systems which helps disaster recovery.

Note – Referenced from Assignment 8.2