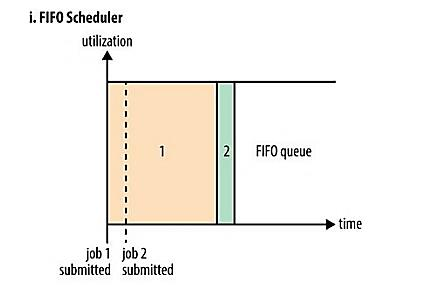
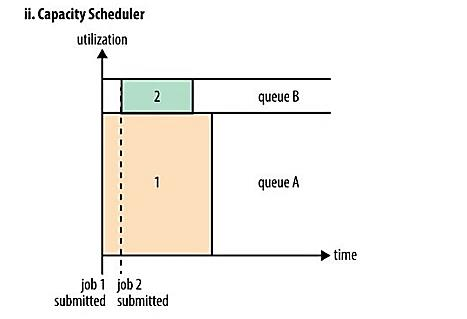
Assignment 6.6

1. Difference between FIFO and Capacity scheduler.

FIFO- First In First Out scheduler. It places the application in a queue and runs them in the order of submission. Initially the first application in the queue are allowed first. Ones it satisfied then the second application will execute. The very first application gets the preference first.



CAPACITY Scheduler – In the capacity scheduler it allows to run the separate queue which allows the smaller jobs finished first as soon as it is submitted. Larger jobs will finished at last as compared to the FIFO scheduler.



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

Limitations of Hadoop 1.x

* Name node is a single point of failure. Failure of name node, leads to inaccessible hadoop cluster. Name node must be high end machine which is less prone to failure.
* Secondary name node cannot act as a name node.
* Overburdened job Tracker.
* Not possible to run non-map reduce jobs.

In order to overcome those limitations hadoop 2.x introduces the HDFS federation and High availability.

**High availability:**

Name node is a single point of failure, need manual recovery using the secondary name node in case of failure. But this can be overcome in the Hadoop 2.x.

We can demonstrate the high availability from the below example:

Problem in 1.X: IF the name node fails then there is no access to the hadoop cluster.

Solution:

* + Hadoop 2.x supports two Name Nodes at a time one node is active and another is standby node
  + Standby Name Node manages metadata same as Secondary Name Node in Hadoop 1.x

**Federation:**

In the Hadoop 1.x federation contains the one name node and a namespace but in Hadoop 2.x we have the multiple name node and namespaces.

In order to scale the name service horizontally, JDFS federation uses multiple independent namenodes. The name nodes are federated, that is, the name node are independent and do not require coordination with each other.