Assignment 8.2

**● Explain the core changes made in Hadoop 2.x.**

In Hadoop 1.x there is single point of failure as it has single namenode to manage the entire namespace. Hadoop 2.x has feature to overcome this Single Point of Failure Problem with a standby Namenode and in case of Namenode failure, it is configured for automatic recovery.It ensures the high availability.

The scalability in Hadoop 2.x is up to 10000 nodes per cluster. Previously it was limited to 400 nodes per cluster.

Hadoop 2.x has YARN.YARN is Yet Another Resource Negotiator .It does the cluster resource management and processing is done by other processing models.

Hadoop 2.x works on the concept of containers with the help of which they can run generic tasks. It overcomes the limitations of Hadoop 1.x.

**● Explain the difference between MapReduce 1 and MapReduce 2 / Yarn**

MapReduce 1 uses the JobTracker to create and assign tasks to data nodes, which can become a resource bottleneck when the cluster scales out far enough

MapReduce 2 /YARN has a Resource Manager for each cluster, and each data node runs a Node Manager. For each job, one slave node will act as the Application Master, monitoring resources/tasks, etc. MapReduce perform data processing via YARN.

YARN splits the two major functionalities of the JobTracker i.e. resource management and job scheduling/monitoring into 2 separate daemons (components).

* + Resource Manager
  + Node Manager(node specific )

Hadoop 2 using YARN for resource management. Besides that, hadoop support programming model which support parallel processing that we know as MapReduce. Before hadoop 2, hadoop already support MapReduce. In short, MapReduce run above YARN Architecture.