Assignment 3.1

Components of Hadoop 2.x:

1. **Master Node**
   - **1.1. HDFS HA Master Node [HIGH AVAILABILITY]**
     - Active Name node
     - Standby Name node
     - Secondary Name node, if there is no standby name node

2. **YARN Master Node**
   - **2.1. Resource Manager**
     - Scheduler
     - Application Manager
   - **2.2. Slave Node**
     - Application Master
     - Node manager

1. **Master Node:**
   It contains the information about the meta data

   1.1. HDFS HA Master Node
   - Hadoop 2.x supports two name node at a time, while one node is active and another is in standby mode which acts as a backup node.
   - Active Name Node handles the client operations in the cluster
   - Standby Name Node manages metadata same as Secondary Name Node in Hadoop 1.x
   - When Active Name Node is down, Standby Name Node takes over and will handle the client operations then after

2. **YARN Master Node**
   YARN has divided the responsibilities of Job Tracker to two processes Resource Manager and Application Master and instead of Task Tracker is using Node Manager daemon for map reduce task execution**.**

2.1. Resource Manager

- This daemon process resides on the Master Node (not necessarily on Name Node of Hadoop)
- Responsible for, managing resources scheduling for different compute applications in an optimum way
- Coordinating with two process on master node, Scheduler and Application Manager

Scheduler

- This daemon process resides on the Master Node (runs along with Resource Manager daemon)
- Responsible for, scheduling the job execution as per submission request received by Resource Manager
- Allocating resources to applications submitted to the cluster
- Coordinating with Application Manager daemon and keeping track of resources of running applications

Application Manager

- This daemon process resides on the Master Node (runs along with Resource Manager daemon)
- Responsible for, helping Scheduler daemon to keeps track of running application by coordination
- Accepting job submissions from client
- Negotiating first container for executing application specific task with suitable Application Master on slave node

2.2. Node Manager

- This daemon process resides on the slave nodes (runs along with Data Node daemon)

- Responsible for, managing and executing containers
- Monitoring resource usage (i.e. usage of memory, cpu, network etc..) and reporting it back to Resource Manager daemon

Application Master

- This daemon process runs on the slave node (along with the Node Manager daemon)
- It is per application specific library works with Node Manager to execute the task
- The instance of this daemon is per application, which means in case of multiple jobs submitted on cluster, it may have more than one instances of Application Master on slave nodes
- Negotiating suitable resource containers on slave node from Resource Manager
- Working with one or multiple Node Managers to monitor task execution on slave nodes