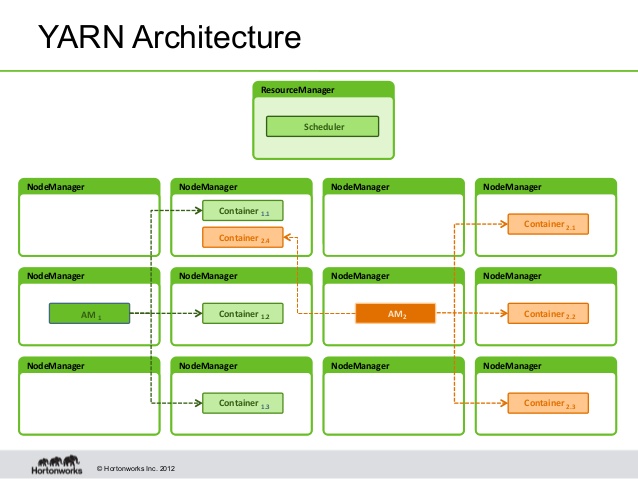
Assignment 8.3

**Explain in brief the architecture of Apache Hadoop Yarn.**



In order to overcome Single Point of Failure problem Hadoop 2.x splits up resource management and job scheduling/monitoring into separate daemons.

The job of resource management is done by YARN (Yet Another Resource Negotiator).It is global Resource Manager (RM).

In YARN, there is Application Master (AM) for every application. An application can be a single job or group of jobs.

The Resource Manager (RM) is responsible for arbitration of resources among all the applications in the system.

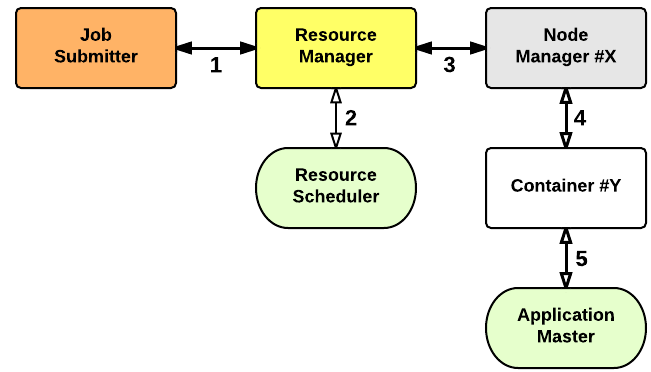
The Node Manager is an agent responsible for containers and monitoring their resource usage and report it to the scheduler (Resource Manager).

The Resource Manager (RM) has two main components: Scheduler and Applications Manager.

The Scheduler is responsible for allocating resources to the various running applications. Scheduler is pure scheduler in the sense that it performs no monitoring or tracking of status for the application.

It offers no guarantees about restarting failed tasks either due to application failure or hardware failures. The Scheduler performs its scheduling function based on the resource requirements of the applications; it does so based on the abstract notion of a resource Container which incorporates elements such as memory, CPU, disk, network etc.

The ApplicationsManager is responsible for accepting job-submissions, negotiating the first container for executing the application specific ApplicationMaster and provides the service for restarting the ApplicationMaster container on failure. The per-application ApplicationMaster has the responsibility of negotiating appropriate resource containers from the Scheduler, tracking their status and monitoring for progress.



The process of application startup can be explained as below:

Step 1: Client submits an application to the RM (Resource Manager).

Step 2: The Resource Manager allocates the container for the application.

Step 3: The Resource Manager contacts respective Node Manager.

Step 4: The Node Manager then launches the container.

Step 5: The container executes Application Master (AM).