

Question 1:

Yarn does the MapReduce by splitting the jortracker to take care of both job scheduling and task progress monitoring. These two roles are separated by YARN into two independent daemons: a resource manager and an application master. The resource manages to manage the use of resources across the cluster, while application master is to manage the lifecycle of applications running on the cluster.

When submit a new job to the MapReduce program, the job will get a new ID from from resource manager. The job clients checks the output specification and copies job resources to HDFS. At last, job is submitted by resource manager.

After resource manager receives the call, the scheduler allocates a container, and application master is launched under the node's manager's management.

Then the application master initializes jobs and coordinates the tasks running the mapreduce jobs and they are scheduled by the resource manager and managed by the node managers. And then retrieve job resources back to shared filesystem.