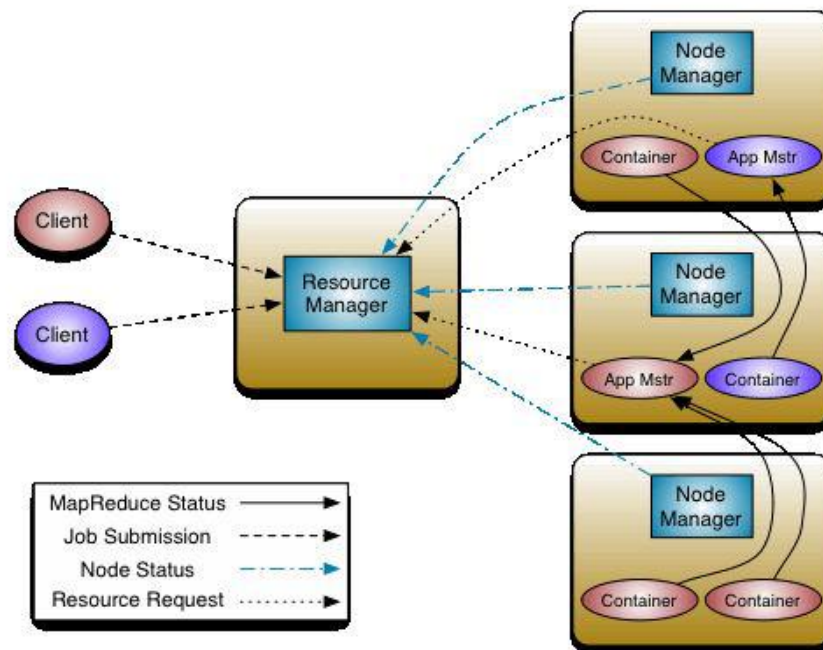


## Components of Hadoop 2.X (YARN)



---

### Components of YARN(Yet Another Resource Negotiator)

#### 1. Global Resource Manager

- Assigns resources among applications for optimal resource utilization.
- One cluster has one instance of resource manager.

#### 2. Node Manager

- Runs on each node and communicates with Resource Manager about the resource usage on machine.
- It receives requests resource manager about resource allocation to jobs and maintains life cycle of containers.

#### 3. Application Specific Application Master

- It is the actual instance which does processing.
- It requests Resource Manager for resources and works with Node Manager to get those resources for resources task execution. Application could be MapReduce or any other processing frame work.

#### 4. Scheduler

- It is plugged with Resource Manager to help in resource allocation. Different schedulers allocate resources using different algorithms.

## 5. Container

- It is set of allocated system resources (CPU and Memory). Containers are allocated and managed by Node Manager and are used by tasks.
- ❖ Resources is handled by Resource Manager and Node Manager.
- ❖ Processing is handled by Application Master (MapReduce one of many Application Master). So, processing other than MapReduce also possible.

### Execution Process

- Job/Application is submitted by the YARN Client application to the Resource Manager daemon along with the command to start the Application Master on any container at Node Manager.
- Application manager process on master node validates the job submission request and hands it over to scheduler process for resource allocation.
- Scheduler process assigns a container for Application Master on one slave node.
- Node Manager daemon starts Application master service within one of its container using the command mentioned in step1, hence Application master is considered to be the first container of any application.
- Application master negotiates the other containers from Resource manager by providing the details like location of data on slave nodes, required CPU, memory, cores etc.
- Resource Manager allocates the best suitable resources on slave nodes and responds to Application master with node details and other details.
- Then, Application Master sends request to Node Manager on suggested slave nodes to start the containers.
- Application master then manages the resources of requested containers while job execution and notifies the Resource Manager when execution is completed.
- Node Managers periodically notify the Resource Manager with the current status of available resources on the node as to what information can be used by scheduler to schedule new application on the clusters.
- In case of any failure of slave node, Resource Manager will try to allocate new container on other best suitable node so that Application Master can complete the process using new container.