**Workflow of MR job on YARN**

**Topic Progress:**

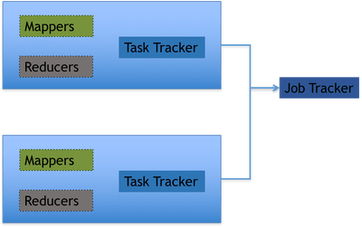
[← Back to Lesson](http://www.itversity.com/lessons/yarn-and-mrv2/)

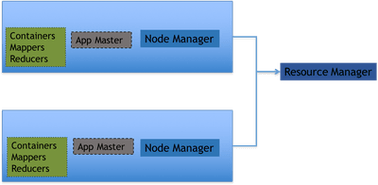
**Introduction**

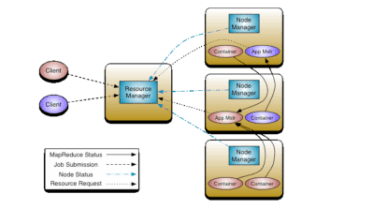
As part of this lesson following topics under YARN+MR2 are covered in detail

* Understand how upgrading a cluster from Hadoop 1 to Hadoop 2 affects cluster settings
* Understand how to deploy MapReduce v2 (MRv2 / YARN), including all YARN daemons
* Understand basic design strategy for MapReduce v2 (MRv2)
* Determine how YARN handles resource allocations
* **Identify the workflow of MapReduce job running on YARN**
* Determine which files you must change and how in order to migrate a cluster from MapReduce version 1 (MRv1) to MapReduce version 2 (MRv2) running on YARN.

**Cluster with YARN – Overview**







**Resource Manager**

* It manages nodes by tracking heartbeats from NodeManagers
* It manages containers
  + Handles application master requests for resources (like providing inputs for creation of containers)
  + De-allocates expired or completed containers
* It manages per job application masters
  + Creates containers for application masters and also tracks their heartbeats
* It also manages security (if Kerberos is enabled)

**Node Manager**

* Communicates with Resource Manager. It sends information about node resources, heartbeats, container status etc.
* Manages processes in containers
  + Launches Application Masters on request from Resource Manager
  + Launches containers (mappers/reducers) on request from Application Master
  + Monitors resource usage by containers(mappers/reducers
* Provides logging services to applications. It aggregates logs for an application and saves those logs to HDFS.
* Runs auxiliary services
* Maintains node level security (ACLs)

**Application Master**

* It will be created per job
* Keep track of progress of the job
* It plays vital role after it is created by Resource Manager
  + Negotiates with node manager to determine where containers should run for mappers and reducers
  + Keep track of progress of the job
  + Provide logs for running jobs
* Once the job execution is completed, logs will be copied to HDFS
* Job history server will provide logs for completed jobs (after application master is destroyed or killed)

**Workflow diagram**

