



# **SNS COLLEGE OF ENGINEERING**



**Kurumbapalayam(Po), Coimbatore – 641 107**

**Accredited by NAAC-UGC with 'A' Grade**

**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

## **Department of Artificial Intelligence and Data Science**

**Course Name – Big Data Analytics  
III Year / V Semester**

**Unit 3 – DATA ANALYTICAL FRAMEWORKS**

**Topic - Yarn Components**



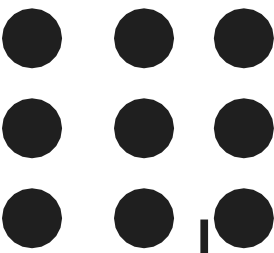


# YARN Components



## Components of YARN Resource Manager

- It is the ultimate authority in resource allocation.
- On receiving the processing requests, it passes parts of requests to corresponding node managers accordingly, where the actual processing takes place.
- It is the arbitrator of the cluster resources and decides the allocation of the available resources for competing applications.
- Optimizes the cluster utilization like keeping all resources in use all the time against various constraints such as capacity guarantees, fairness, and SLAs.
- It has two major components:
  - a) Scheduler
  - b) Application Manager



# YARN Components

## a) Scheduler

- The scheduler is responsible for allocating resources to the various running applications subject to constraints of capacities, queues etc.
- It is called a pure scheduler in ResourceManager, which means that it does not perform any monitoring or tracking of status for the applications.
- If there is an application failure or hardware failure, the Scheduler does not guarantee to restart the failed tasks.
- Performs scheduling based on the resource requirements of the applications.
- It has a pluggable policy plug-in, which is responsible for partitioning the cluster resources among the various applications.
- There are two such plug-ins: Capacity Scheduler and Fair Scheduler, which are currently used as Schedulers in ResourceManager.

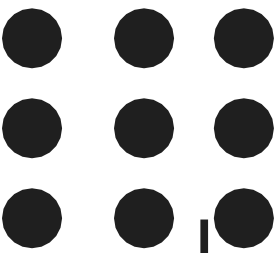


# YARN Components



## b) Application Manager

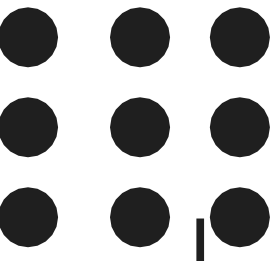
- It is responsible for accepting job submissions.
- Negotiates the first container from the Resource Manager for executing the application specific Application Master.
- Manages running the Application Masters in a cluster and provides service for restarting the Application Master container on failure.



# YARN Components

## Node Manager

- It takes care of individual nodes in a Hadoop cluster and manages user jobs and workflow on the given node.
- It registers with the Resource Manager and sends heartbeats with the health status of the node.
- Its primary goal is to manage application containers assigned to it by the resource manager.
- It keeps up-to-date with the Resource Manager.
- Application Master requests the assigned container from the Node Manager by sending it a Container Launch Context(CLC) which includes everything the application needs in order to run. The Node Manager creates the requested container process and starts it.
- Monitors resource usage (memory, CPU) of individual containers.
- Performs Log management.
- It also kills the container as directed by the Resource Manager.



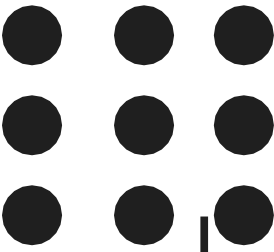
# YARN Components

## Application Master

- An application is a single job submitted to the framework. Each such application has a unique Application Master associated with it which is a framework specific entity.
- It is the process that coordinates an application's execution in the cluster and also manages faults.
- Its task is to negotiate resources from the Resource Manager and work with the Node Manager to execute and monitor the component tasks.
- It is responsible for negotiating appropriate resource containers from the ResourceManager, tracking their status and monitoring progress.
- Once started, it periodically sends heartbeats to the Resource Manager to affirm its health and to update the record of its resource demands.

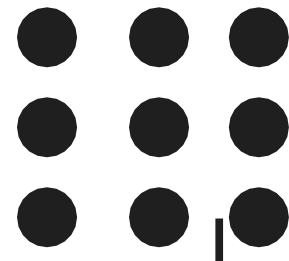


# YARN Components



## Container

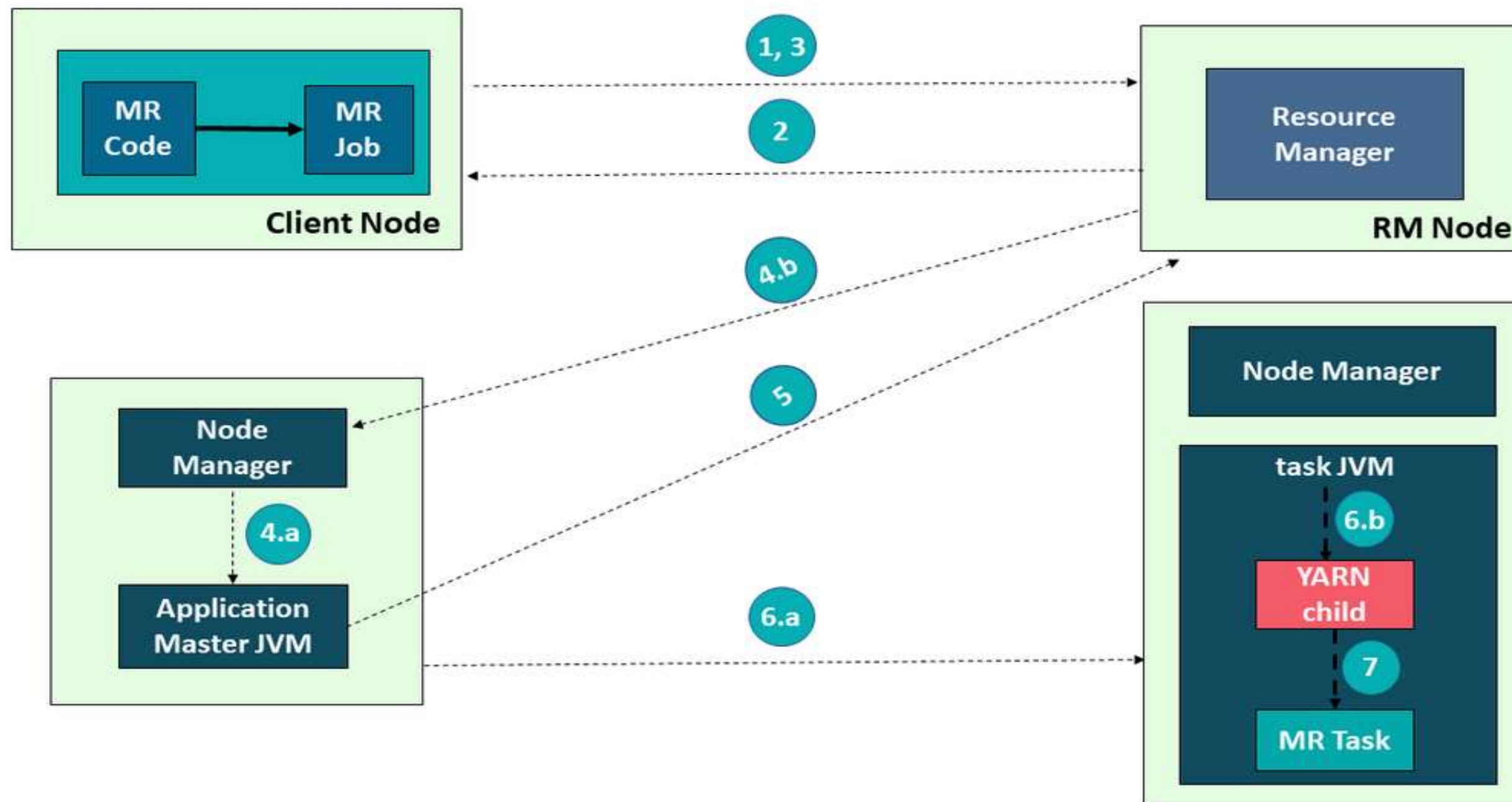
- It is a collection of physical resources such as RAM, CPU cores, and disks on a single node.
- YARN containers are managed by a container launch context which is container life-cycle (CLC).
- This record contains a map of environment variables, dependencies stored in a remotely accessible storage, security tokens, payload for Node Manager services and the command necessary to create the process.
- It grants rights to an application to use a specific amount of resources (memory, CPU etc.) on a specific host.



# YARN Components

## Application Workflow in Hadoop YARN

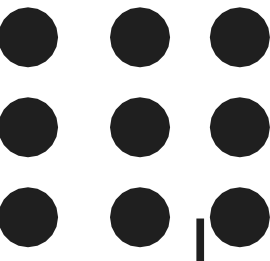
- Refer to the given image and see the following steps involved in Application workflow of Apache Hadoop YARN:







# YARN Components



## Application Workflow in Hadoop YARN

- Client submits an application
- Resource Manager allocates a container to start Application Manager
- Application Manager registers with Resource Manager
- Application Manager asks containers from Resource Manager
- Application Manager notifies Node Manager to launch containers
- Application code is executed in the container
- Client contacts Resource Manager/Application Manager to monitor application's status
- Application Manager unregisters with Resource Manager



**THANK YOU**