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

YARN

Hadoop Yet Another Resource Negotiator (YARN):

- This is a framework for job scheduling and cluster resource management.
- A resource management framework for scheduling and handling resource requests from distributed applications.

Why YARN?

- In Hadoop version 1.0 which is also referred to as MRV1 (MapReduce Version 1), MapReduce performed both processing and resource management functions.

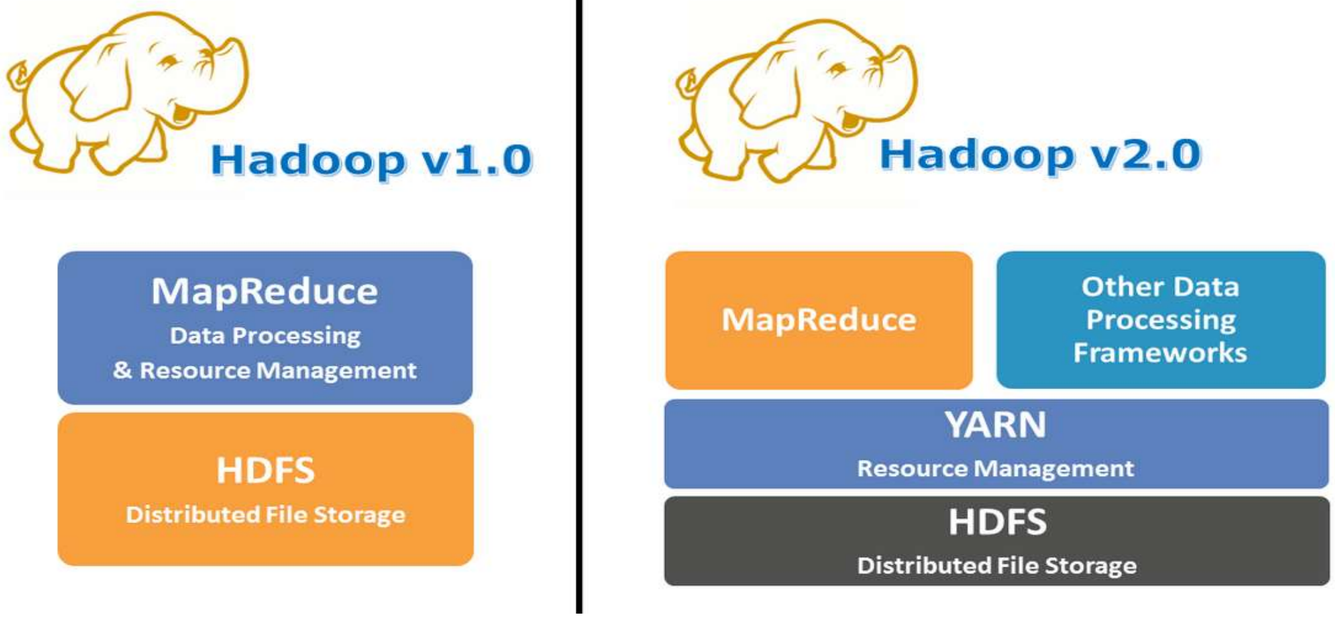
- It consisted of a Job Tracker which was the single master. This design resulted in scalability bottleneck due to a single Job Tracker.

- The practical limits of such a design are reached with a cluster of 5000 nodes and 40,000 tasks running concurrently.

- Apart from this limitation, the utilization of computational resources is inefficient in MRV1. Also, the Hadoop framework became limited only to MapReduce processing paradigm.

YARN

- To overcome all these issues, YARN was introduced in Hadoop version 2.0 in the year 2012 by Yahoo and Hortonworks.
- The basic idea behind YARN is to relieve MapReduce by taking over the responsibility of Resource Management and Job Scheduling.
- YARN started to give Hadoop the ability to run non-MapReduce jobs within the Hadoop framework.



YARN Architecture

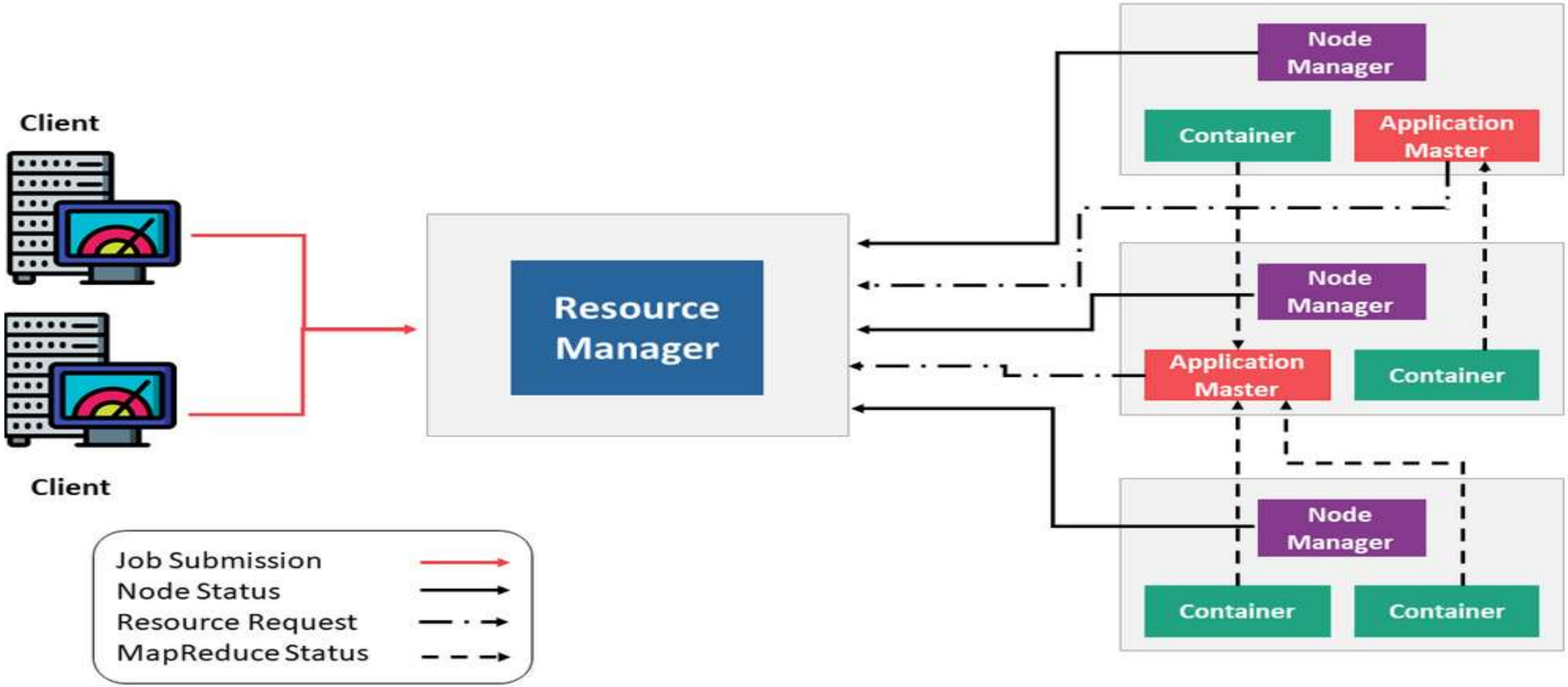
YARN Architecture

- Apart from Resource Management, YARN also performs Job Scheduling. YARN performs all your processing activities by allocating resources and scheduling tasks.

Apache Hadoop YARN Architecture consists of the following main components :

- Resource Manager: Runs on a master daemon and manages the resource allocation in the cluster.
- Node Manager: They run on the slave daemons and are responsible for the execution of a task on every single Data Node.
- Application Master: Manages the user job lifecycle and resource needs of individual applications. It works along with the Node Manager and monitors the execution of tasks.
- Container: Package of resources including RAM, CPU, Network, HDD etc on a single node.

YARN Architecture



YARN Architecture

YARN Architecture

- Apart from Resource Management, YARN also performs Job Scheduling. YARN performs all your processing activities by allocating resources and scheduling tasks.

Apache Hadoop YARN Architecture consists of the following main components :

- Resource Manager: Runs on a master daemon and manages the resource allocation in the cluster.
- Node Manager: They run on the slave daemons and are responsible for the execution of a task on every single Data Node.
- Application Master: Manages the user job lifecycle and resource needs of individual applications. It works along with the Node Manager and monitors the execution of tasks.
- Container: Package of resources including RAM, CPU, Network, HDD etc on a single node.

THANK YOU