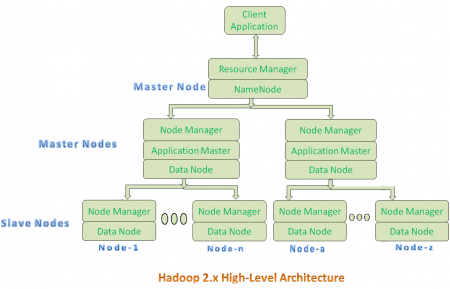
**Hadoop 2.x Major Components**

Hadoop 2.x has the following three Major Components:

* HDFS
* YARN
* MapReduce

These three are also known as Three Pillars of Hadoop 2. Here major key component change is YARN.

**HADOOP 2.X ARCHITECTURE**



**ARCHITECTURE DESCRIPTION**

* At the root of a YARN hierarchy is the ResourceManager. This entity governs an entire cluster and manages the assignment of applications to underlying compute resources.
* The ResourceManager orchestrates the division of resources (compute, memory, bandwidth, etc.) to underlying NodeManagers (YARN's per-node agent).
* The ResourceManager also works with ApplicationMasters to allocate resources and work with the NodeManagers to start and monitor their underlying application.
* In this context**, the ApplicationMaster has taken some of the role of the prior TaskTracker, and the ResourceManager has taken the role of the JobTracker.**
* An ApplicationMaster manages each instance of an application that runs within YARN.
* The **ApplicationMaster is responsible for negotiating resources from the ResourceManager** and, through the NodeManager, monitoring the execution and resource consumption of containers (resource allocations of CPU, memory, etc
* The NodeManager manages each node within a YARN cluster. The NodeManager provides per-node services within the cluster, from overseeing the management of a container over its life cycle to monitoring resources and tracking the health of its node.

**RESOURCE MANAGER:**

**ResourceManager (RM)** is the master that arbitrates all the available cluster resources and thus helps manage the distributed applications running on the YARN system. It works together with the per-node **NodeManagers (NMs)** and the per-application **ApplicationMasters (AMs)**.

**NODE MANAGER:**

The NodeManager (NM) is YARN’s per-node agent, and takes care of the individual compute nodes in a Hadoop cluster. This includes keeping up-to date with the ResourceManager (RM), overseeing containers’ life-cycle management; monitoring resource usage (memory, CPU) of individual containers, tracking node-health, log’s management and auxiliary services which may be exploited by different YARN applications.

**Application Master:**

* Application Master is a per-application level component. It is responsible for:
  1. Managing assigned Application Life cycle.
  2. It interacts with both Resource Manager’s Scheduler and Node Manager
  3. It interacts with Scheduler to acquire required resources.
  4. It interacts with Node Manager to execute assigned tasks and monitor those task’s status.

**Container:**

* Each Master Node or Slave Node contains set of Containers. In this diagram, Main Node’s Name Node is not showing the Containers. However, it also contains a set of Containers.
* Container is a portion of Memory in HDFS (Either Name Node or Data Node).
* In Hadoop 2.x, Container is a place where a unit of work occurs. For instance each MapReduce task(not the entire job) runs in one container.
* Set of system resources are allocated for each container, currently CPU core and RAM are supported. Each node in a Hadoop cluster can run several containers.
* In Hadoop 2.x, Container is similar to Data Slots in Hadoop 1.x. We will see the major differences between these two Components: Slots Vs Containers in my coming posts.

**NOTE:-**

* Resource Manager is Per-Cluster component where as Application Master is per-application component.
* Both Hadoop 1.x and Hadoop 2.x Architectures follow Master-Slave Architecture Model.