**Apache Hadoop V.2.x has the following three major Components**

* HDFS V.2
* YARN (MR V2)
* MapReduce (MR V1)

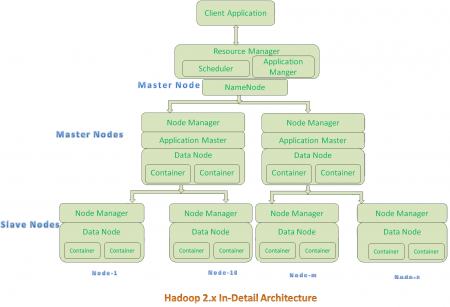
In Hadoop V.2.x, these two are also known as Three Pillars of Hadoop.

**YARN :**

The fundamental idea of YARN is to split up the functionalities of resource management and job scheduling/monitoring into separate daemons. The idea is to have a global ResourceManager (*RM*) and per-application ApplicationMaster (*AM*). An application is either a single job or a DAG of jobs.

The ResourceManager and the NodeManager form the data-computation framework. The ResourceManager is the ultimate authority that arbitrates resources among all the applications in the system. The NodeManager is the per-machine framework agent who is responsible for containers, monitoring their resource usage (cpu, memory, disk, network) and reporting the same to the ResourceManager/Scheduler.

The per-application ApplicationMaster is, in effect, a framework specific library and is tasked with negotiating resources from the ResourceManager and working with the NodeManager(s) to execute and monitor the tasks.



The ResourceManager has two main components: Scheduler and ApplicationsManager.

**MAP REDUCE:**

* MapReduce is a programming model designed for processing large volumes of
* data in parallel by dividing the work into a set of independent tasks.
* MapReduce sends code to distributed data, instead of bringing data to the actual code.
  + MapReduce works by breaking the processing into two phases:
  + the map phase
  + the reduce phase
  + Each phase has key-value pairs as input and output, the types of which may be chosenby the programmer.
  + The programmer also specifies two functions:
  + the map function
  + the reduce function
  + There is a shuffle and sort phase in between.
  + Map phase takes input in Key-Value pairs
  + It produces output in the form of Key-Value pair. Output from various Map tasks are grouped together on the basis of Key.
  + Key and its associated set of values are sent to the Reduce phase.
  + Reduce method operates on key and associated list of values.
  + Output of Reduce is written to HDFS.

**HDFS:**

.HDFS has mainly 3 components ,

1.Name Node (Job Tracker in 1.x)

2.Secondary NameNode

3.Data Node (Task Tracker in 1.x)

.Name Node:

--> This contains all the hadoop file system tree and other metadata information about files and the directories

--> This stores all the information of data that are stored at Data Nodes

.Secondary NameNode:

-->This does House-Keeping work to Name node like merging of data etc..

-->This is not back up for Name Node.

.Date Node:

-->This stores Actual data in its own local system disk in form of Blocks.

-->Sends heart beart to NameNode at regular intervals to inform NameNode that it is still Active and running.

-->It sends information to Name Node during Blockage at cluster startup and menawhile it sends signal at every 10th Heart Beat.

-->These are WorkHorse of the system.

-->It perfoms Check Sum operation(which is file size) and does other operations that are instructed by Name Node.

.Job Tracker: Controls overall execution of Map Reduce Jobs

.Task Tracker: This runs individual job runs in its data nodes present.And Communicates with Job Tracker periodiaclly to give updates(Heart Beat)

.Hadoop Cluster if it is connetced to large number of computers then they will be called as Nodes.

.NAME NODE:

--> This contains two sub parts that are presnt in HDFS.

1. fsimage

2. Edits

fsimage::

--> Contains all directory structure of HDFS.

--> has info about replication levels

--> modification,access times of file.

--> access permission to files & directories

Edits::

--> Any write operation takes place in HDFS then the directory structure gets modified.

--> These modifications are stored in memory as well as in Edits(Stored on Hard Disk).

--> If any fsimage file gets merged with edits then we will get updated fsimage file (which is entirely new)

--> This Process is called checkpointing and itis taken care by Secondary NameNode.