**Hadoop 2.x Major Components**

Hadoop 2.x has the following three Major Components:

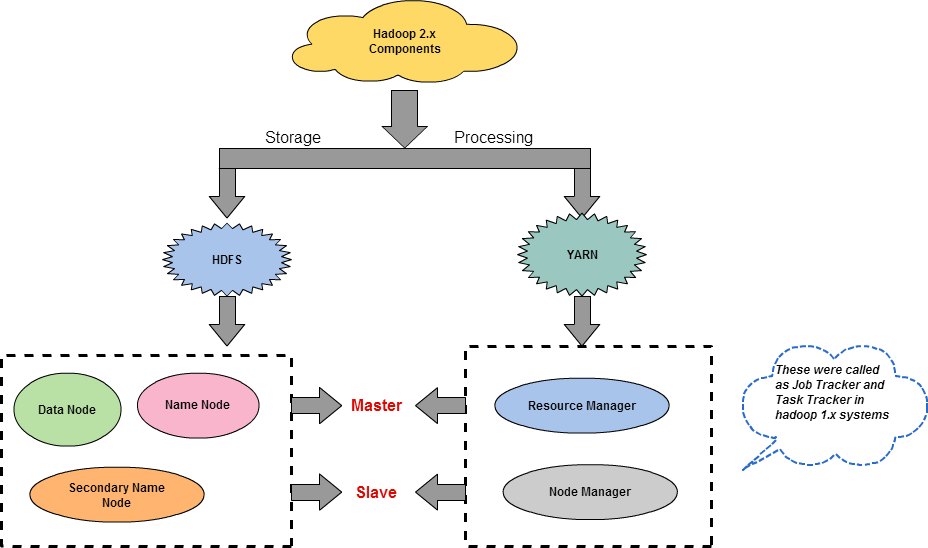
* **HDFS** :HDFS stands for Hadoop Distributed File System. It is also know as HDFS V2 as it is part of Hadoop 2.x with some enhanced features. It is used as a Distributed Storage System in Hadoop Architecture.
* **YARN** :YARN stands for Yet Another Resource Negotiator. It is new Component in Hadoop 2.x Architecture. It is also know as “MR V2”.
* **Map Reduce**: Map Reduce is a Batch Processing or Distributed Data Processing Module. It is also know as “MR V1” as it is part of Hadoop 1.x with some updated features.Map Reduce in hadoop-2.x maintains **API compatibility** with previous stable release (hadoop-1.x). This means that all Map Reduce jobs should still run unchanged on top of YARN with just a recompile.

These three are also known as Three Pillars of Hadoop 2. Here major key component change is YARN. It is really game changing component in Big Data Hadoop System.

**How Hadoop 2.x Major Components Works**

Hadoop 2.x components interact each other and to work parallel in a reliable, highly available and fault-tolerant manner.

**The below snap shot shows you the back end components of Hadoop System.**



**Apache Hadoop 2.x (MRv2) consists of the following daemons:**

* Name Node
* Secondary Name Node
* Data Node
* Resource Manager
* Node Manager

**Name Node** (Hadoop File System Component)**:**Stores Meta data only

The Name Node is the centerpiece of an HDFS file system. It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept. It does not store the data of these files itself.

**Secondary Name node** is optional now & Standby Name node has been to used for fail over process.

Standby Name Node will stay up-to-date with all the file system changes the Active Name Node makes .

**Date Node** (Hadoop File System Component): Stores Blocks from files

A Data Node stores the actual data in the HDFS. A functional file system typically have more than one Data Node in the cluster, with data replicated across them. On start up, a Data Node connects to the Name Node; spinning until that service comes up. It then responds to requests from the Name Node for file system operations.

**Resource Manager** (YARN Component)

The function of the Resource Manager is simple: Keeping track of available resources. One per cluster. It contains two main components: Scheduler and Applications Manager.  
The Scheduler is responsible for allocating resources to the various running applications.  
The Applications Manager is responsible for accepting job-submissions, negotiating the first container for Application Master and provides the service for restarting the Application Master container on failure.

**Node Manager** (YARN Component)

The Node Manager is the per-machine framework agent who creates container for each task. The containers can have variable resource sizes and the task can be any type of computations not just map/reduce tasks. It then monitors the resource usage (cpu, memory, disk, network) of the container and report them to the Resource Manager.

**Key Benefits of Hadoop 2.0 YARN Component-**

* It offers improved cluster utilization
* Highly scalable
* Beyond Java
* Novel programming models and services
* Agility