**Hadoop 2.x Major Components**

Hadoop 2.x has the following three Major Components:

* HDFS
* YARN
* MapReduce

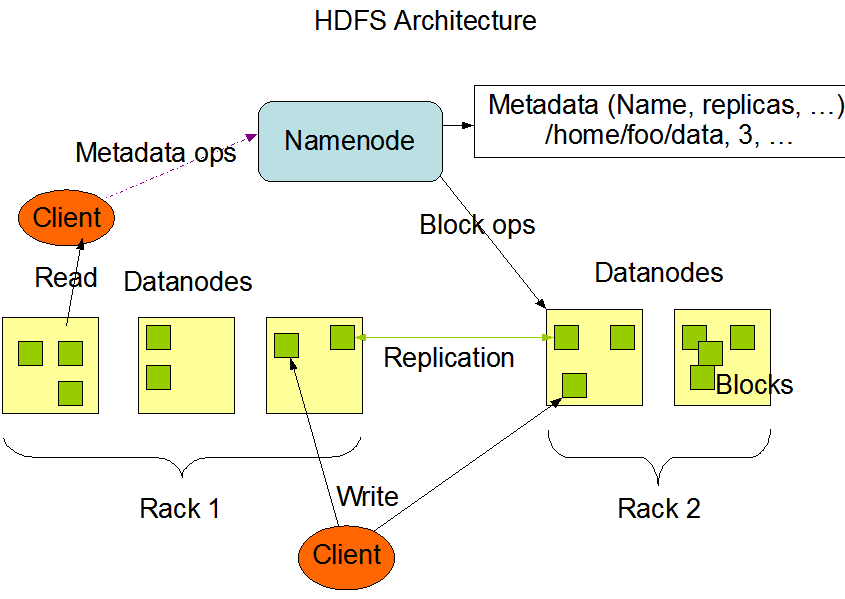
**1. HDFS**

**NameNode (Hadoop FileSystem Component):**Stores Metadata only

The NameNode 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.

**DateNode (Hadoop FileSystem Component):** Stores Blocks from files

A DataNode stores the actual data in the HDFS. A functional filesystem typically have more than one DataNode in the cluster, with data replicated across them. On startup, a DataNode connects to the NameNode; spinning until that service comes up. It then responds to requests from the NameNode for filesystem operations.



**2. Map reduce(YARN)**

In YARN, there are at least three actors:

1**. the Job Submitter** (the client)

2**. the Resource Manager** (the master)

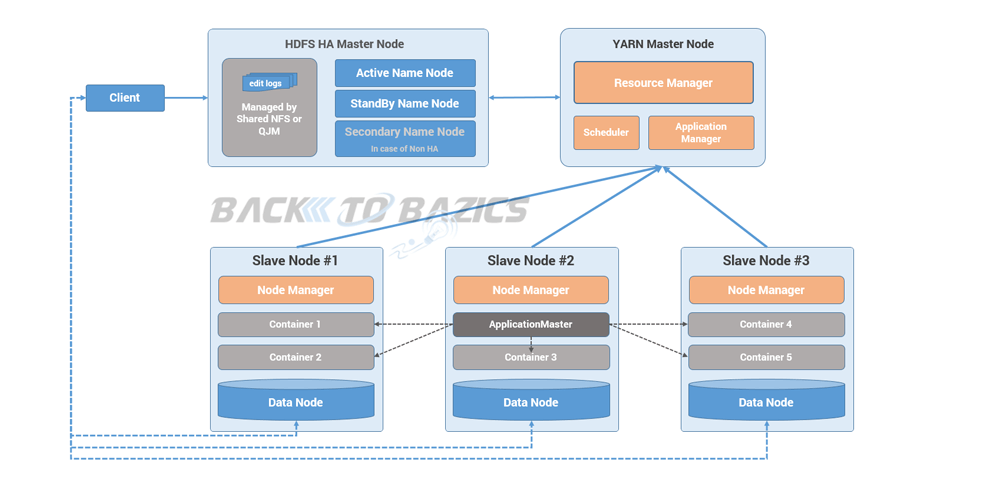
3**. the Node Manager** (the slave)

* **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 ApplicationsManager.  
The Scheduler is responsible for allocating resources to the various running applications.  
The ApplicationsManager is responsible for accepting job-submissions, negotiating the first container for ApplicationMaster and provides the service for restarting the ApplicationMaster container on failure.

* **Node Manager (YARN Component)**

The NodeManager 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 ResourceManager.

****