Hadoop 2.x has some common Hadoop API which can easily be integrated with any third party applications to work with Hadoop

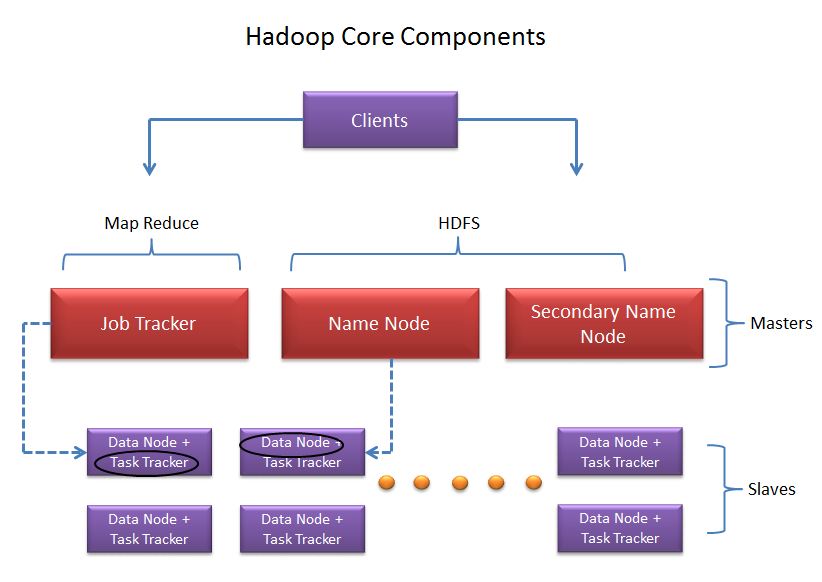
It has some new Java APIs and features in HDFS and MapReduce which are known as HDFS2 and MR2 respectively

New architecture has added the architectural features like HDFS High Availability and HDFS Federation

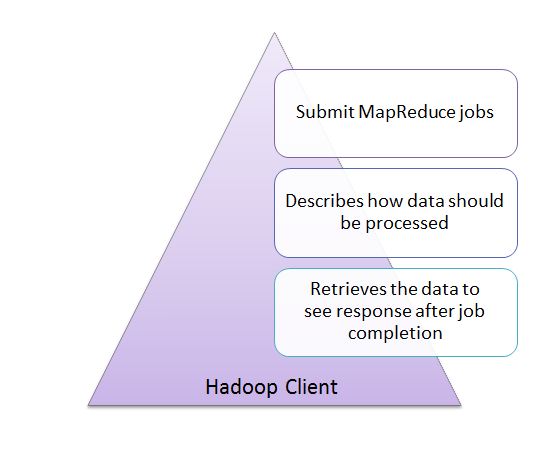
Hadoop 2.x not using Job Tracker and Task Tracker daemons for resource management now on-wards, it is using YARN (Yet Another Resource Negotiator) for Resource Management

Hadoop cluster has 3 components:

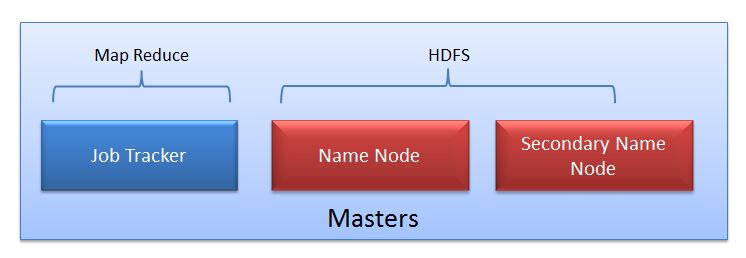
* 1. Client
  2. Master
  3. Slave

The role of each components are shown in the below image.   
  
 

* Client:

It is neither master nor slave, rather play a role of loading the data into cluster, submit MapReduce jobs describing how the data should be processed and then retrieve the data to see the response after job completion.   
  
 

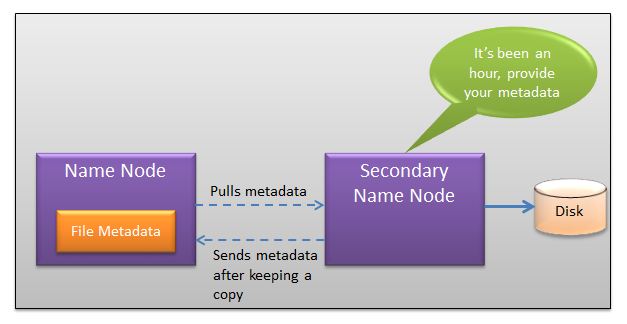
Masters:

The Masters consists of 3 components NameNode, Secondary Node name and JobTracker.   
  
   
  
**NameNode:**  
NameNode does NOT store the files but only the file's metadata.   
  
NameNode oversees the health of DataNode and coordinates access to the data stored in DataNode.   
Name node keeps track of all the file system related information such as to

* + Which section of file is saved in which part of the cluster
  + Last access time for the files
  + User permissions like which user have access to the file

**JobTracker:**  
JobTracker coordinates the parallel processing of data using MapReduce. 

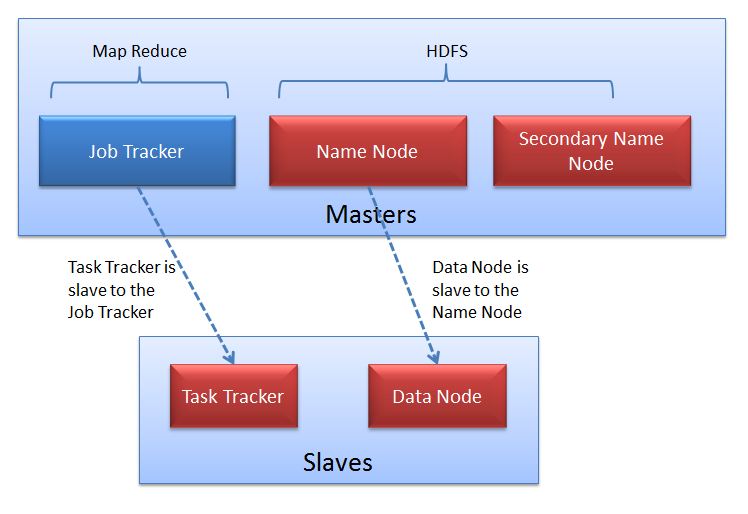
**Secondary Name Node:**

Don't get confused with the name "Secondary". Secondary Node is NOT the backup or high availability node for Name node.   
  
So what Secondary Node does?   
  
   
  
The job of Secondary Node is to contact NameNode in a periodic manner after certain time interval.  
NameNode which keeps all filesystem metadata in RAM has no capability to process that metadata on to disk. If NameNode crashes, you lose everything in RAM itself and you don't have any backup of filesystem. What secondary node does is it contacts NameNode in an hour and pulls copy of metadata information out of NameNode. It shuffles and merge this information into clean file folder and sent to back again to NameNode, while keeping a copy for itself. Hence Secondary Node is not the backup rather it does job of housekeeping.   
In case of NameNode failure, saved metadata can rebuild it easily. 

**Slaves:**

Slave nodes are the majority of machines in Hadoop Cluster and are responsible to

* + Store the data
  + Process the computation

   
  
Each slave runs both a DataNode and Task Tracker daemon which communicates to their masters. The Task Tracker daemon is a slave to the JobTracker and the DataNode daemon a slave to the NameNode