• Name node

* NameNode is the centerpiece of HDFS.
* NameNode is also known as the Master
* NameNode only stores the metadata of HDFS – the directory tree of all files in the file system, and tracks the files across the cluster.
* NameNode does not store the actual data or the dataset. The data itself is actually stored in the DataNodes.
* NameNode knows the list of the blocks and its location for any given file in HDFS. With this information NameNode knows how to construct the file from blocks.
* NameNode is so critical to HDFS and when the NameNode is down, HDFS/Hadoop cluster is inaccessible and considered down.
* NameNode is a single point of failure in Hadoop cluster.
* NameNode is usually configured with a lot of memory (RAM). Because the block locations are help in main memory.

• Da**t**a node

* DataNode is responsible for storing the actual data in HDFS.
* DataNode is also known as the Slave
* NameNode and DataNode are in constant communication.
* When a DataNode starts up it announce itself to the NameNode along with the list of blocks it is responsible for.
* When a DataNode is down, it does not affect the availability of data or the cluster. NameNode will arrange for replication for the blocks managed by the DataNode that is not available.
* DataNode is usually configured with a lot of hard disk space. Because the actual data is stored in the DataNode.

• 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

NodeManagers take instructions from the ResourceManager and manage resources available on a single node.