All components of Hadoop 1.x

* HDFS stands for Hadoop Distributed File System. It is also known as HDFS V1 as it is part of Hadoop 1.x. It is used as a Distributed Storage System in Hadoop Architecture.
* MapReduce is a Batch Processing or Distributed Data Processing Module. It is built by following Google’s MapReduce Algorithm. It is also known as “MR V1” or “Classic MapReduce” as it is part of Hadoop 1.x.
* Remaining all Hadoop Ecosystem components work on top of these two major components: HDFS and MapReduce..

Hadoop 1.x MapReduce is also know as “Classic MapReduce” as it was developed by following Google’s MapReduce Algorithm Tech Paper.

**Hadoop 1.x Major Components**

Hadoop 1.x Major Components components are: HDFS and MapReduce. They are also know as “Two Pillars” of Hadoop 1.x.

**HDFS:**  
HDFS is a Hadoop Distributed FileSystem, where our BigData is stored using Commodity Hardware. It is designed to work with Large DataSets with default block size is 64MB (We can change it as per our Project requirements).

HDFS component is again divided into two sub-components:

1. Name Node

Name Node is placed in Master Node. It used to store Meta Data about Data Nodes like “How many blocks are stored in Data Nodes, Which Data Nodes have data, Slave Node Details, Data Nodes locations, timestamps etc” .

1. Data Node

Data Nodes are places in Slave Nodes. It is used to store our Application Actual Data. It stores data in Data Slots of size 64MB by default.

**MapReduce:**  
MapReduce is a Distributed Data Processing or Batch Processing Programming Model. Like HDFS, MapReduce component also uses Commodity Hardware to process “High Volume of Variety of Data at High Velocity Rate” in a reliable and fault-tolerant manner.

MapReduce component is again divided into two sub-components:

1. Job Tracker

Job Tracker is used to assign MapReduce Tasks to Task Trackers in the Cluster of Nodes. Sometimes, it reassigns same tasks to other Task Trackers as previous Task Trackers are failed or shutdown scenarios.

Job Tracker maintains all the Task Trackers status like Up/running, Failed, Recovered etc.

1. Task Tracker

Task Tracker executes the Tasks which are assigned by Job Tracker and sends the status of those tasks to Job Tracker.

Secondary NameNode

• Performs house-keeping activities for Name Nodes, like the periodic merging of namespace and edits.

• This is not a back up for a Name Node

But where in handoop 2.x ,there is backup for master node when it stops working then control is taken by secondary namenode automatically

### Hadoop 1.x Limitations

Hadoop 1.x has many limitations or drawbacks. Main drawback of Hadoop 1.x is that MapReduce Component in it’s Architecture. That means it supports only MapReduce-based Batch/Data Processing Applications.

Hadoop 1.x has the following Limitations/Drawbacks:

* It is only suitable for Batch Processing of Huge amount of Data, which is already in Hadoop System.
* It is not suitable for Real-time Data Processing.
* It is not suitable for Data Streaming.
* It supports upto **4000 Nodes** per Cluster.
* It has a single component : JobTracker to perform many activities like Resource Management, Job Scheduling, Job Monitoring, Re-scheduling Jobs etc.
* JobTracker is the single point of failure.
* It does not support Multi-tenancy Support.
* It supports only one Name Node and One Namespace per Cluster.
* It does not support Horizontal Scalability.
* It runs only Map/Reduce jobs.
* It follows Slots concept in HDFS to allocate Resources (Memory, RAM, CPU). It has static Map and Reduce Slots. That means once it assigns resources to Map/Reduce jobs, it cannot re-use them even though some slots are idle.