COMPONENTS OF HADOOP 1.X:

* HDFS stands for Hadoop Distributed File System. It is also know 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.

Hadoop 1.x Major Components components are:

* HDFS
* MapReduce.

They are also know as “Two Pillars” of Hadoop 1.x.

HDFS:

It is designed to work with Large DataSets.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.

Hadoop 1.x components interact each other and to work parallel in a reliable and fault-tolerant manner.

* `Both Master Node and Slave Nodes contain two Hadoop Components:
  1. HDFS Component
  2. MapReduce Component
* Master Node’s HDFS component is also known as “Name Node” is used to store Meta Data.
* Slave Node’s HDFS component is also known as “Data Node” is used to store actual our application Big Data.
* HDFS stores data by using 64MB size of “Data Slots” or “Data Blocks”.
* Master Node’s MapReduce component is also known as “Job Tracker” which take care assigning tasks to “Task Tracker” and receiving results from them.
* Slave Node’s MapReduce component is also known as “Task Tracker” which contains two MapReduce Tasks:
  1. Map Task
  2. Reduce Task
* Clients (one or more) submit their work to Hadoop System.
* When Hadoop System receives a Client Request, first it is received by a Master Node.
* Master Node’s MapReduce component “Job Tracker” is responsible for receiving Client Work and divides into manageable independent Tasks and assign them to Task Trackers.
* Slave Node’s MapReduce component “Task Tracker” receives those Tasks from “Job Tracker” and perform those tasks by using MapReduce components.
* Once all Task Trackers finished their job, Job Tracker takes those results and combines them into final result.
* Finally Hadoop System will send that final result to the Client.