**1. Cluster and Hadoop Cluster:**

**Cluster:**

In general, Cluster is a group of similar things or people are grouped closely together.

In a computer system, a cluster is a group of servers or group of nodes and other resources that act like a single system and enable high availability.

In personal computer storage technology, a cluster is the logical unit of file storage on a hard disk. It is managed by the computer's operating system.

**Hadoop Cluster:**

A Hadoop cluster is a special type of computational cluster mainly designed for storing and analyzing huge amounts of unstructured data in a distributed computing environment using HDFS and Map Reduce.

Hadoop clusters are comprised of three different node types.

Master nodes

Slave nodes

Client nodes.

Master nodes perform operations the following operations, storing data in the Hadoop Distributed File System (HDFS) and running parallel computations on that data using Map reduce.

Slave nodes is used to storing the data and running computations.

Client nodes have Hadoop installed with all the cluster settings, but are neither master nor worker nodes. Instead, the client node loads data into the cluster, submits Map Reduce jobs describing how that data should be processed, and then retrieves or views the results of the job when processing is finished.

**2.Components of Hadoop1.x**

Hadoop1.x has two major components.

1.HDFS

2.Map Reduce

**1.HDFS**

HDFS stands for Hadoop Distributed File System. Actually our huge amount of data is stored in nodes using commodity hardware. Commodity hardware is less expensive.

It has two sub components:

Name node

Data node

**Name node:**

Name node is the master which is used to store the meta data.  
The meta data is the data about data node like how many blocks are stored in data node, slave node details, data node location, time stamp.

**Data node:**

Data node is the slave node. It is used to store the actual data in blocks. The file details are available in data node.

**2.Map Reduce**

Map reduce is used to processing the data. Map reduce also uses the commodity hardware to process the high volume of data at high velocity rate.

Map reduce is also divided into two sub components:

Job Tracker

Task Tracker

**Job Tracker:**

Job Tracker is used to assign the map reduce task to task trackers. It is also assign the same task to another task tracker, if previous task tracker is failed.

**Task Tracker:**

Task Tracker is used to do the task which is assigned by job tracker. It is also send the status about the task to job tracker.