Cluster

Cluster is a group of computers connected together so that they can share resources

It is used for load balancing and high avaiablity of resources.

Hadoop cluster

Hadoop cluster is a special cluster which is used in a distributed system for storing and analysing huge amount of both structured and un structured data. It runs on low cost commodity hardware.

One machine in the cluster is calle as name node and other are called as data node

It helps in boosting data analysis. They are highly scalable. Additional cluster nodes can be added to increase the throughput

Hadoop 1.x

The major components of hadoop 1.x are

Hdfs

Mapreduce

Hadoopecosystem

Hdfs

It consist of namenode and data nodes where data are stored in a distributed file system which are group of clusters

Map Reducer

It is a distributed data processing model. Map reducer is also divided as

Jobtracker :it is used for mapreduce task to task track in the cluster of nodes

Tasktracker: It executes the task assigned to it by job tracker

The map reduce maily runs on java . it consist of mapper and reducer

The mapper class is used to map the data and the reducer class is used to reduce the result.It also consist of shuffle and sort where the result from mapper is shuffled and sorted.