**ASSIGNMENT-2.5**

1. Cluster and Hadoop Cluster:

***Cluster:***

* In general, Cluster is a group of similar things or people positioned or occurring closely together.
* In a computer system, a cluster is a group of servers and other resources that act like a single system and enable high availability and, in some cases, load balancing and parallel processing.
* In personal computer storage technology, a cluster is the logical unit of file storage on a [hard disk](http://searchstorage.techtarget.com/definition/hard-disk); it's managed by the computer's [operating system](http://searchcio-midmarket.techtarget.com/definition/operating-system).
* In some products, a cluster is a group of terminals or workstations attached to a common control unit.

***Hadoop Cluster:***

* A **Hadoop cluster** is a special type of computational **cluster** designed specifically for storing and analyzing huge amounts of unstructured data in a distributed computing environment.
* Hadoop clusters are comprised of three different node types: master nodes, worker nodes, and client nodes.
* **Master nodes** perform operations that comprise: storing data in the Hadoop Distributed File System (HDFS) and running parallel computations on that data using Map reduce.
* **Worker nodes** make up the majority of virtual machines and perform the job of 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 Hadoop 1.x:**

* Name Node
* Secondary Name Node
* Data Node
* Job Tracker
* Task Tracker

***Name Node***: It 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” .

***Secondary Name Node:*** In Hadoop it is especially dedicated node in HDFS cluster whose main function is to take checkpoints of the file system metadata present onname node. It is not a backup name node. It just checkpoints name node’s file system namespace. It merges edit log and FSImage and produce new FsImage at regular check points.

***Data Node:*** Data Nodes are places in Slave Nodes. It is used to store our Application Actual Data. It stores data in Data Slots.

***Job Tracker:*** Job Tracker is used to assign Map Reduce 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.

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