**Session3 Assignment1**

List the components of Hadoop 2.x and explain each component in detail

---- >

Components of Hadoop 2.x are:

* NameNode
* DataNode
* Resource Manager
* Node Manager
* Application Master
* Secondary Namenode

1. **NameNode:**

* NameNode holds the meta data for the HDFS like Namespace information, block information etc...
* There are two different files associated with the NameNode:
  + fsimage – It is the snapshot of the filesystem.
  + Edit logs – After the start of NameNode it maintains the every transaction that happened with NameNode.
* NameNode maintains the namespace tree and the mapping of blocks to Datanodes
* Client communicates with the NameNode and provides data to HDFS through it. HDFS then stores data as blocks inside DataNodes.
* NameNode maintains the metadata information of all the blocks present inside the hadoop cluster like permissions, modification, access times, namespace and disk space.
* NameNode is also known as master node.

1. **DataNode:**

* DataNode is the space where actual data is stored in hadoop cluster in a distributed manner.
* Every DataNode directly reports to the Namenode about the blocks which it is handling.
* DataNode communicates with Namenode by sending Heartbeats.
* During startup each DataNode connects to Namenode and performs a handshake to verify the namespace ID and the software version of the DataNode.

1. **Resource Manager:**

* RM assigns resources among applications for optimal resource utilization.
* One cluster has one instance of Resource Manager.
* RM is responsible for tracking the resources in a cluster and scheduling applications
* RM has Active / Standby architecture – at any point of time, one of the RMs is Active and one or more RMs are in standby mode waiting to take over if anything happen to the Active.

1. **Node Manager:**

* Node Manager runs on each node and communicates with Resource Manager about resource usage on the machine.
* Node Manager receives requests from resource manager about resource allocation to jobs and maintains life cycle of containers.

1. **Application Master:**

* Application Master is the actual instance which does processing.
* Application Master requests Resource Manager for resources and works with NodeManager to get those resources for task execution.
* Application Master could be processing framework.

1. **Secondary NameNode:**

* Secondary Namenode asks the namenode for its edit logs in regular intervals and copies them into the fsimage.
* After updating the fsimage, NameNode copy back that fsimage.
* Namenode uses this fsimage when it starts, this eventually will reduce the startup time.
* Secondary NameNode is used to maintain a checkpoint in HDFS.
* When a Namenode fails, Secondary NameNode helps NameNode in bringing back its data.
* Secondary Namenode is also called as checkpoint node in hadoop’s architecture.