Components of Hadoop 2.x

These are the components of Hadoopn2.x :-

1 . High Availability

In this there are two namanode : Active Namenode and Stand-By Namenode

If Active namenode goes down Stand-By Namenode will act as Active Namenode and

whole cluster will be accessible, this is high availability. This was lacking in Hadoop 1.x

which was having one NameNode and it fails whole cluster was inaccessible.

Active NameNode and Stand-By NameNode both share the same Edit log file to have the

block report and other detail so that if one goes down other can act accordingly.

2 . HDFS Federation

In HDFS Federation there are multiple namespace.

As we can have more than two namenode so it is possible to have more than one namespace

As well . Suppose we have divided our HDFS in to two directory,i In this we can make one

namenode acting for one directory and other namenode acting for other directory to make

the system smooth running and there will be no overhead on one name node.

3. YARN(Yet Another Resource Negotiator)

Also known as MapReduce version 2.

Hadoop 2.x has remove old daemon which were there in Hadoop 1.x : Job Tracker and Task

Tracker

In Hadoop 2.x we have : Resource manager and Node manager

Resource Manager is the master of the infrastrucre, one for each cluster. It has many services

and resource scheduler is one of them and most important which decide how to assign the

resource.

Node manager is the slave of the infrastructure and many for one cluster.

It sends heart beat to Resource manager periodically to say that it is alive and block report for

the data. Some portaion od the space of the container is used for node manager for running

the program.