**Big Data using HADOOP**

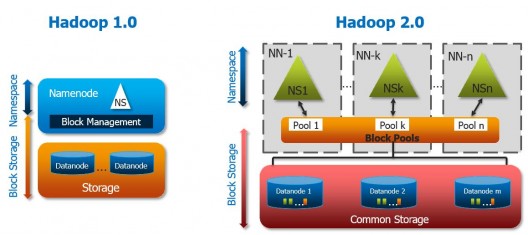
**Hadoop V1 vs Hadoop V2**

**Hadoop V1 -**

1. Batch processing  only supported  i.e  only map reduce processing is achieved
2. Single point of failure due to name node down
3. External data storage is needed for real time processing or graph analysis
4. Doesn't support multi-tenancy   (means can't processing multiple jobs at the same time)
5. can't run more 4000 cluster with better performance

**Hadoop V2 –**

1. HDFS  federation
2. multiple namenode
3. for Map reducing here YARN
4. For better  processing control
5. support for non mapreduce type processing
6. support for multi-tenancy

****

**Setting up HDFS Cluster –**

1. Install JAVA 8..

* rpm -ivh <http://monalisa.cern.ch/MONALISA/download/java/jdk-8-linux-x64.rpm>
* rpm -ql jdk | grep jps
* copy this -> /usr/java/jdk1.8….. (this is the path of the java)

1. Download Hadoop tar file ..

* wget url ..
* tar -xzvf Hadoop….
* Move this folder in /opt/hadoop2

1. Setting up Path ..

* vim /root/.bashrc

HADOOP\_HOME = /opt/hadoop2

JAVA\_HOME=/usr/java/jdk1.8….

PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

export PATH

1. Reload the bashrc file –

**source /root/.bashrc**

1. **Setting up HDFS –**
2. Edit Java path in -

**/opt/hadoop2/etc/hadoop/hadoop-env.sh**

**export JAVA\_HOME=/usr/java/jdk1.8….**

1. Configure **core-site.xml** in both DATANODE AND NAMENODE-

**vim /opt/hadoop2/etc/hadoop/core-site.xml**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://namenodeip:anyport</value>

</property>

</configuration>

1. Configure **hdfs-site.xml** in datanode

**HDFS –**

After setting up the cluster and starting the datanode and namenode ---

1. To List the available directories in HDFS –

**hdfs dfs -ls /**

1. To Create a folder in HDFS –

**hdfs dfs -mkdir /harsh**

1. To upload a file in HDFS –

Create a file using (for learning) –

**yes “hello world” > abc.txt**

Now copying this file into HDFS (from local system to HDFS) –

**hdfs dfs -copyFromLocal /path/abc.txt /harsh/**

OR

**hdfs dfs -put /path/abc.txt /path/bcd.txt /harsh**/

(you can put multiple files using put)

1. To Copy files in HDFS (from HDFS to HDFS) –

**hdfs dfs -cp /harsh/demo.txt /harsh/demo1.txt**

1. To Copy files from HDFS to Local Machine –

**hdfs dfs -copyToLocal /harsh/demo.txt /destination\_address**

1. To Delete a file from HDFS –

**hdfs dfs -rm /harsh/demo.txt**

1. To Delete a directory from HDFS –

**hdfs dfs -rm -r /harsh**

(will delete even non-empty directory)

**hdfs dfs -rmdir /harsh**

(only delete empty directory)

1. View data of the file in HDFS –

**hdfs dfs -cat /harsh/demo.txt**

1. To view replication factor of HDFS –

**hdfs dfs -ls /harsh**

-rw-r--r-- **3** root supergroup 81932288 2019-06-29 05:37 /harsh/demo.txt

( 3 is the default replication factor)

1. To change the default replication factor –

**hdfs dfs -setrep 2 /harsh/demo.txt**

1. To change the replication factor at the time of upload –

**hdfs dfs -D dfs.replication=1 -put demo.txt /harsh**

1. To check number of directories and files in HDFS –

**hdfs dfs -count /harsh**

1. To turn off the SAFEMODE in NAMENODE –

**hdfs dfsadmin -safemode leave**