**HADOOP SETUP ON UBUNTU**

**system hostname: myubuntu**

**ip address : 10.1.4.59**

**🡪 Before installing update your system by below command.**

**sudo apt-get update.**

**🡪 COMMANDS to install java and ssh**

**Open terminal execute the following commands**

**sudo apt-get install openjdk-7-jdk**

**sudo apt-get install ssh ( optional)**

**🡪 After completion of above steps then,**

**First download the Hadoop 2.6 version.**

**and also add one empty directory ( example : hadoop2data) in any location.**

**🡪 In order to configure Hadoop modify the following files found in**

**“ $HADOOP\_HOME/etc/hadoop “ folder.**

**1. Change the following files:**

**Core-site.xml**

**<?xml version="1.0" encoding="UTF-8"?>**

**<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>**

**<configuration>**

**<property>**

**<name>fs.defaultFS</name>**

**<value>hdfs://myubuntu:8020</value>**

**</property>**

**</configuration>**

**hdfs-site.xml**

**<?xml version="1.0" encoding="UTF-8"?>**

**<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>**

**<configuration>**

**<property>**

**<name>dfs.replication</name>**

**<value>1</value>**

**</property>**

**<property>**

**<name>dfs.namenode.name.dir</name>**

**<value>file:/home/hadoop/work/hadoop2data/dfs/name</value>**

**</property>**

**<property>**

**<name>dfs.datanode.data.dir</name>**

**<value>file:/home/hadoop/work/hadoop2data/dfs/data</value>**

**</property>**

**<property>**

**<name>dfs.namenode.checkpoint.dir</name>**

**<value>file:/home/hadoop/work/hadoop2data/dfs/namesecondary</value>**

**</property>**

**</configuration>**

**mapred-site.xml**

**<?xml version="1.0"?>**

**<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>**

**<configuration>**

**<property>**

**<name>mapreduce.framework.name</name>**

**<value>yarn</value>**

**</property>**

**</configuration>**

**yarn-site.xml**

**<?xml version="1.0"?>**

**<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>**

**<configuration>**

**<property>**

**<name>yarn.resourcemanager.hostname</name>**

**<value>myubuntu</value>**

**</property>**

**<property>**

**<name>yarn.nodemanager.aux-services</name>**

**<value>mapreduce\_shuffle</value>**

**</property>**

**<property>**

**<name>yarn.nodemanager.aux-services.mapreduce\_shuffle.class</name>**

**<value>org.apache.hadoop.mapred.ShuffleHandler</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.address</name>**

**<value>myubuntu:8021</value>**

**</property>**

**<property>**

**<name>yarn.log-aggregation-enable</name>**

**<value>true</value>**

**</property>**

**</configuration>**

**Export the java location:**

**mapred-env.sh**

**export JAVA\_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64**

**hadoop-env.sh**

**export JAVA\_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64**

**yarn-env.sh**

**export JAVA\_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64**

**2. open bashrc file and enter classpaths.**

**command: gedit ~/.bashrc**

**export JAVA\_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64**

**export HADOOP\_HOME=/home/msruser1/hadoopcluster/hadoop-2.6.0**

**export HADOOP\_COMMON\_HOME=$HADOOP\_HOME**

**export HADOOP\_HDFS\_HOME=$HADOOP\_HOME**

**export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME**

**export YARN\_HOME=$HADOOP\_HOME**

**export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop**

**export YAR$N\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop**

**export PATH=$JAVA\_HOME/bin:$PATH**

**export PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$PATH**

**🡪 Start the hadoop (pre process for hadoop setup)**

**Command: $ (only one time process)**

**🡪.Format Name node in Namenode machine only(msrcluster1)**

**(don’t start Hadoop in any machine )**

**Command: hadoop namenode -format**

**🡪 To Start Hadoop processes: type the following Command.**

**start-all.sh**

**or**

**$HADOOP\_HOME/bin/start-all.sh**

**🡪 Also start “JobHistoryServer” to solve problems in eco-systems like “ connection refused or connection failed etc.. )**

**🡪 cd /home/msruser1/hadoopcluster/hadoop-2.6.0/sbin/mr-jobhistory-daemon.sh start historyserver --config $HADOOP\_CONF\_DIR**

**🡪 DONE. (type jps to see hadoop is running or not )**