

Hadoop-2 Single Node Cluster Creation

Installation

1. Download Hadoop and Java

```
$ tar -zxvf hadoop-2.9.1.tar.gz
```

(Extract the tar file)

```
$ tar -zxvf jdk-8u45-linux-x64.tar
```

(Extract the tar file)

```
$sudo apt-get install vim
```

(Install USER Friendly Editor)

```
$ vi .bashrc
```

(Set the java Path in your Home Path)

```
export JAVA_HOME=/home/username/jdk1.8.0_45
```

```
export PATH=$HOME/bin:$JAVA_HOME/bin:$PATH
```

```
$ source .bashrc
```

(Execute the bashrc file)

```
$ echo $JAVA_HOME
```

(Check the java path)

2. Modify Hadoop Configuration Files

NAMENODE	---->	core-site.xml
RESOURCE MANGER	---->	mapred-site.xml
SECONDARYNAMENODE	---->	
DATANODE	---->	slaves
NODEMANGER	---->	slaves & yarn-site.xml

Ports used by Hadoop Daemons

Daemon	RPC	WEB
NameNode	50000	50070
DataNode	50010	50075
SecondaryNameNode		50090
ResourceManger	8032	8088
NodeManger	8040	8042

\$ vi etc/hadoop/core-site.xml

```
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:50000</value>
</property>
```

\$ vi etc/hadoop/yarn-site.xml

```
<property>
<name>yarn.node.manager.aux-services</name> <value>mapreduce_shuffle</value>
</property>

<property>
<name>yarn.node.manager.aux-services.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>

<property>
<description>The hostname of the RM.</description>
<name>yarn.resourcemanager.hostname</name>
<value>localhost</value>
</property>

<property>
<description>The address of the applications manager interface in the RM.</description>
<name>yarn.resourcemanager.address</name>
<value>${ yarn.resourcemanager.hostname }:8032</value>
</property>
```

\$ vi etc/hadoop/hdfs-site.xml

```
<property>
<name>dfs.namenode.name.dir</name>
<value>/home/username/hadoop2-dir/namenode-dir</value>
</property>

<property>
<name>dfs.datanode.data.dir</name>
<value>/home/username/hadoop2-dir/datanode-dir</value>
</property>
```

\$ vi etc/hadoop/mapred-site.xml

```
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
```

\$ vi etc/hadoop/hadoop-env.sh

export JAVA_HOME=/home/username/jdk1.8.0_45

\$ vi etc/hadoop/mapred-env.sh

export JAVA_HOME=/home/username/jdk1.8.0_45

\$ vi etc/hadoop/yarn-env.sh

export JAVA_HOME=/home/username/jdk1.8.0_45

\$ vi etc/hadoop/slaves

localhost

Install the ssh key

(Generates, Manages and Converts Authentication keys)

\$ sudo apt-get install openssh-server

\$ ssh-keygen -t rsa

(Setup passwordless ssh to localhost and to slaves)

\$ cd .ssh

\$ ls

\$ cat id_rsa.pub >> authorized_keys (copy the .pub)

(Copy the id_rsa.pub from NameNode to authorized_keys in all machines)

\$ ssh localhost

(Asking No Password)

3. Format NameNode

\$ cd hadoop-2.9.1

\$ bin/hadoop namenode -format

(Your Hadoop File System Ready)



4. Start All Hadoop Related Services

\$ sbin/start-all.sh

(Starting Daemon's For DFS & YARN)

NameNode
DataNode
SecondaryNameNode
ResourceManager
NodeManager

(check the Browser Web GUI)

NameNode - <http://localhost:50070/>

Resource Manager - <http://localhost:8088/>

5. Stop All Hadoop and Yarn Related Services

\$ sbin/stop-all.sh