EXNO.:8

Install Hadoop single node cluster and run simple applications like word count.

Aim:

To Install Hadoop single node cluster and run simple Applications like word count.

Steps:

Install Hadoop

Step1: Click here to down load the Java 8 Package. Save this file in your home directory.

Step2: Extract the JavaTarFile.

***Command*:**tar -xvf jdk-8u101-linux-i586.tar.gz

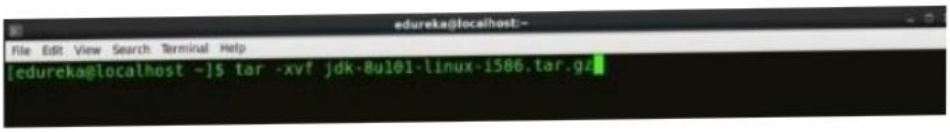


Fig:Hadoop Installation–Extracting Java Files

Step3:**DownloadtheHadoop2.7.3 Package.**

***Command*:**wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz

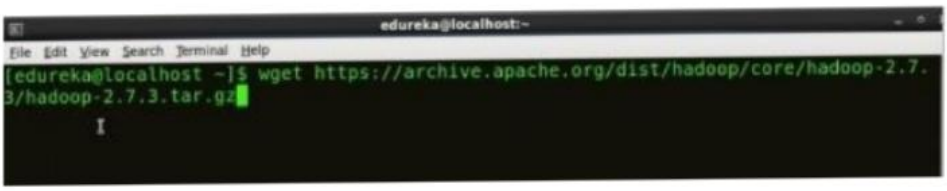


Fig: Hadoop Installation– Downloading Hadoop

Step4: Extract the Hadoop tar File.

Command:tar-xvfhadoop-2.7.3.tar.gz

Fig: Hadoop Installation – Extracting Hadoop Files **Step 5:**

Add the Hadoop and Java paths in the bash file (.bashrc). Open**.b ashrc** file.

Now, add Hadoop and Java Path as shown below.

***Command*:**vi.bashrc

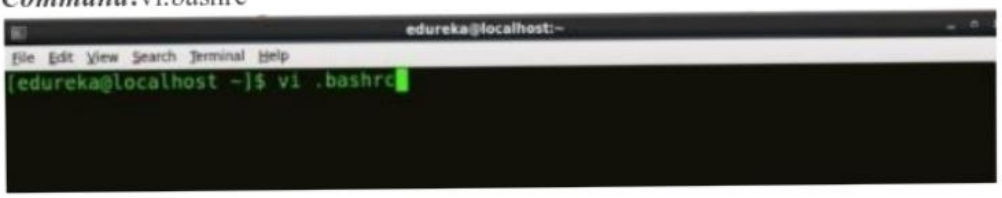


Fig:Hadoop Installation–Setting Environment Variable

Then, save the bash file and close it.

For apply in gall the sechan gesto the current Terminal, execute the source command.
***Command*:** source.bashrc

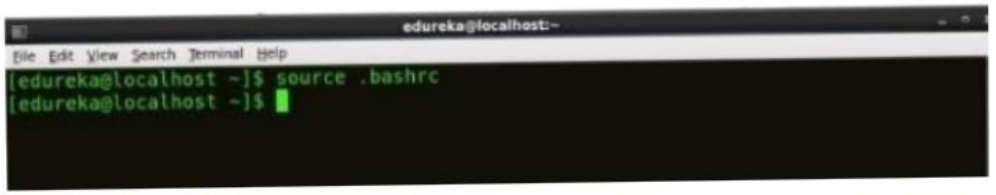
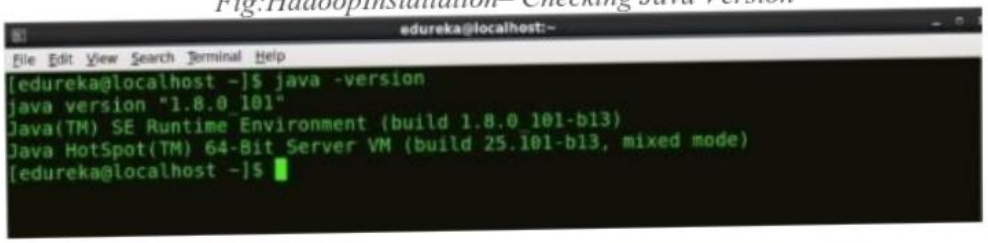


Fig:HadoopInstallation–Refreshing environment variables

To make sure that Java and Hadoop have been properly installed on your system and can be accessed through the Terminal, execute the java -version and hadoop version commands.

***Command*:**java-version

Fig:HadoopInstallation– Checking Java Version



***Command*:** hadoop version

```
edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hadoop version
Hadoop 2.7.3
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r baa91f7c6bc9cb92be
5982de4719c1c8af91ccff
Compiled by root on 2016-08-18T01:41Z
Compiled with protoc 2.5.0
From source with checksum 2e4ce5f957ea4db193bce3734ff29ff4
This command was run using /home/edureka/hadoop-2.7.3/share/hadoop/common/hadoop-co
mmon-2.7.3.jar
[edureka@localhost ~]$
```

Fig:Hadoop Installation– Checking Hadoop Version

Step6: Edit the **Hadoop Configuration files**.

Command: cdhadoop-2.7.3/etc/hadoop/

Command:ls

All the Hadoop configuration files are located in **hadoop-2.7.3/etc/hadoop** directory as you can see in the snapshot below:

```
edureka@localhost:~/hadoop-2.7.3/etc/hadoop
File Edit View Search Terminal Help
[edureka@localhost ~]$ cd hadoop-2.7.3/etc/hadoop/
[edureka@localhost hadoop]$ ls
capacity-scheduler.xml      httpfs-env.sh               mapred-env.sh
configuration.xsl           httpfs-log4j.properties     mapred-queues.xml.template
container-executor.cfg      httpfs-signature.secret     mapred-site.xml.template
core-site.xml               httpfs-site.xml             slaves
hadoop-env.cmd              kms-acls.xml                ssl-client.xml.example
hadoop-env.sh               kms-env.sh                  ssl-server.xml.example
hadoop-metrics2.properties  kms-log4j.properties        yarn-env.cmd
hadoop-metrics.properties   kms-site.xml                yarn-env.sh
hadoop-policy.xml           log4j.properties            yarn-site.xml
hdfs-site.xml               mapred-env.cmd
[edureka@localhost hadoop]$
```

Fig:Hadoop Installation– Hadoop Configuration Files

Step 7: Open *core-site.xml* and edit the property mentioned below inside configuration tag:

core-site.xml informs Hadoop daemon where Name Node runs in the cluster. It contains configuration settings of Hadoop core such as I/O settings that are common to HDFS & Map Reduce.

***Command*:**vicore-site.xml

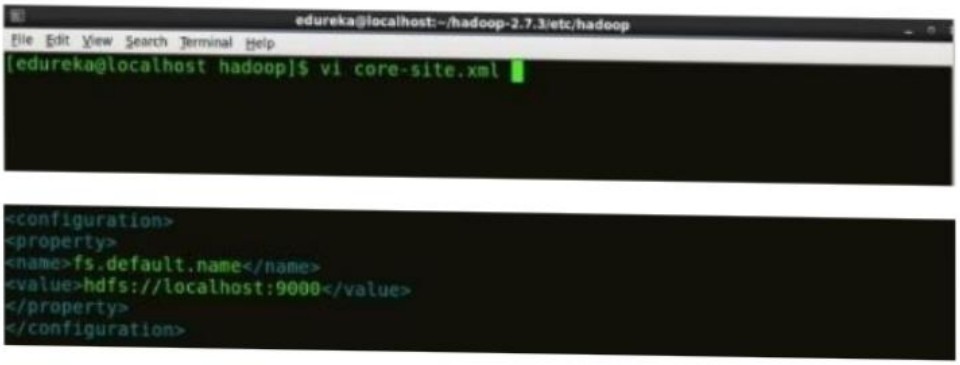


Fig:HadoopInstallation– Configuring core-site.xml

```
1
2                <?xmlversion="1.0"encoding="UTF-8"?>
3      <?xml-stylesheettype="text/xsl"href="configuration.xsl"?>
4                          <configuration>
5                            <property>
6                      <name>fs.default.name</name>
7                    <value>hdfs://localhost:9000</value>
                             </property>
                           </configuration>
```

Step 8: Edit *hdfs-site.xml* and edit the property mentioned below inside **configurationtag:**

hdfs-site.xml contains configuration setting so fHDFS daemons(i.e.NameNode,DataNode, Secondary Name Node). It also includes the replication factor and block size of HDFS.

***Command*:** vihdfs-site.xml

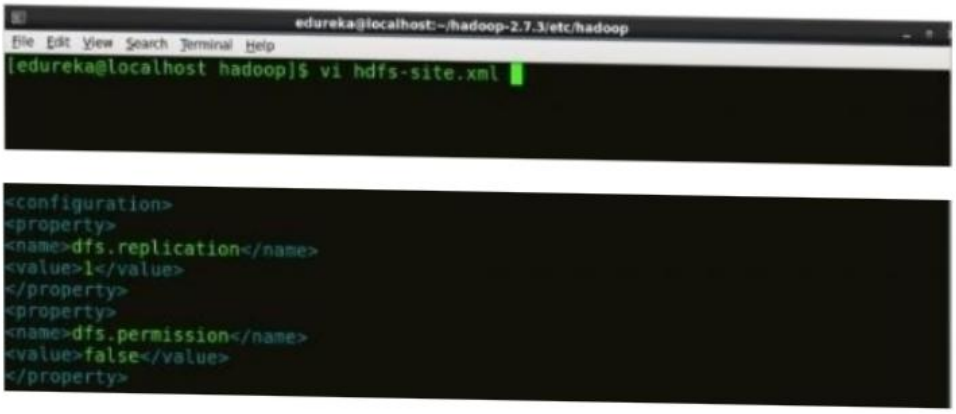


Fig:Hadoop Installation– Configuringhdfs-site.xml

```
<?xmlversion="1.0"encoding="UTF-8"?>
<?xml-stylesheettype="text/xsl"href="configuration.xsl"?>
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.permission</name>
<value>false</value>
</property>
</configuration>
```

Step9: Edit the *mapred-site.xml* file and edit the property mentioned below

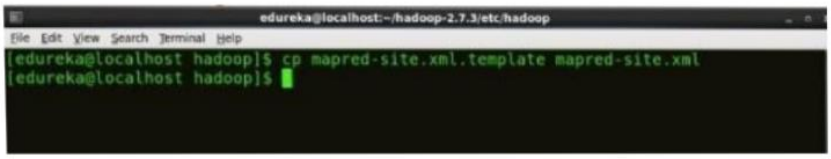
Inside configuration tag:

mapred-site.xml contains configuration settings of Map Reduce application like number of JVM that can run in parallel, the size of the mapper and the reducer process, CPU cores available for a process, etc.

In some cases, mapred-site.xml file is not available. So, we have to create the map red- site.xml file using mapred-site.xml template.

***Command*:** cpmapred-site.xml. template mapred-site.xml

***Command*:**vimapred-site.xml.



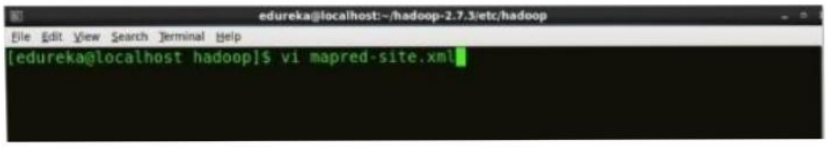


Fig:Hadoop Installation– Configuring mapred-site.xml

```
<?xmlversion="1.0"encoding="UTF-8"?>
<?xml-stylesheettype="text/xsl"href="configuration.xsl"?>
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

Step 10: Edit *yarn-site.xml* and edit the property mentioned below inside configuration tag:

yarn-site.xml contains configuration settings of Resource Manager and Node Manager like application memory management size, the operation needed on program & algorithm, etc.

***Command*:**viyarn-site.xml

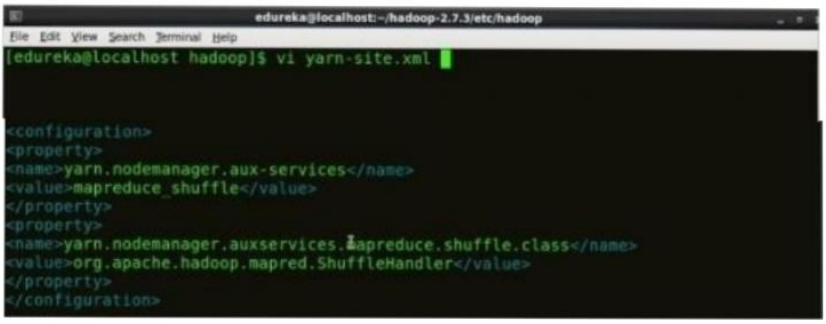


Fig:HadoopInstallation– Configuring yarn-site.xml

Step11: Edit *hadoop-env.sh* and add the JavaPath as mentioned below:

```
<?xmlversion="1.0">
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</
name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
```

hadoop-env.sh contains the environment variables that are used in the script to run Hadoop

like Java home path, etc.

***Command*:**vihadoop–env.sh

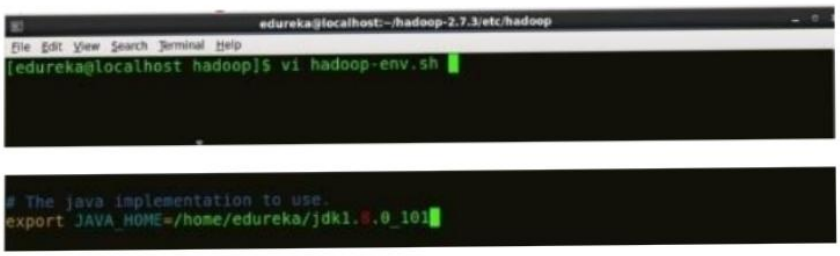


*Fig:HadoopInstallation–Configuringhadoop-env.sh***Step 12:**

Go to Hadoop home directory and format the Name Node.

***Command*:** cd

***Command*:** cd hadoop-2.7.3

***Command*:** bin/hadoop name node-format

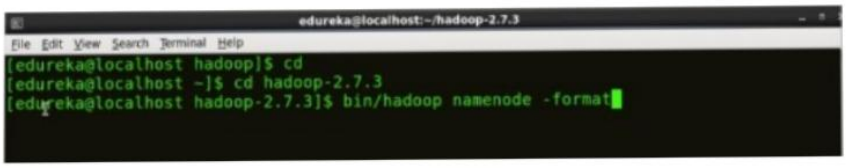


Fig:HadoopInstallation– Formatting NameNode

This formats the HDFS via Name Node.This command is only executed for the firsttime. Formatting the file system means initializing the directory specified by the dfs. name.dir variable.

Never format, up and running Hadoop file system. You will lose all your data stored in the HDFS.

Step13:**Once the Name Node is formatted, goto hadoop-2.7.3/s bin directory and start all the daemons.**

Command: cdhadoop-2.7.3/sbin

Either you can start all daemons with a single command or do it individually.

***Command*:**./start-all.sh

The above command is a combination of ***start-dfs.sh,start-yarn.sh&mr-jobhistory-daemon.sh***

Or you can run all the services individually as below:

Start Name Node:

The Name Node is the center piece of an HDFS file system. It keeps the directory tree of all files stored in the HDFS and tracks all the file stored across the cluster.

Command: ./hadoop-daemon.sh start name node

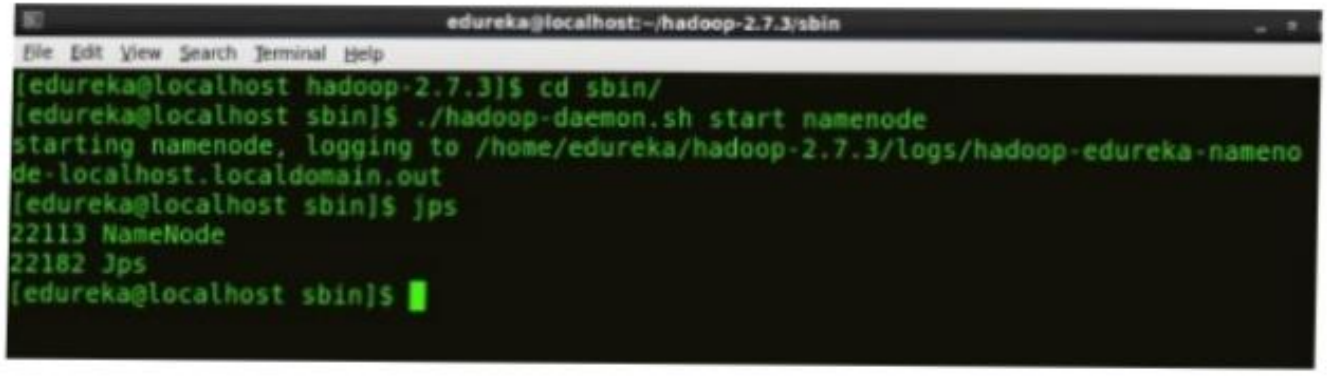


Fig: Hadoop Installation–Starting Name Node

Start Data Node:

On startup, a Data Node connects to the Name node and it responds to the requests from the Name node for different operations.

Command:./hadoop-daemon.sh start data node

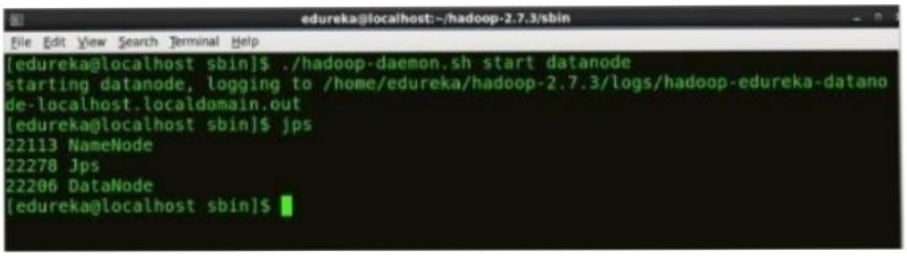


Fig: Hadoop Installation–StartingDataNode

Start Resource Manager:

Resource Manager is the master that arbitrates all the available cluster resources and thus helps in managing the distributed applications running on the YARN system.Its work is to manage each Node Managers and the each application's Application Master.

Command:./yarn-daemon.shstart resourcemanager

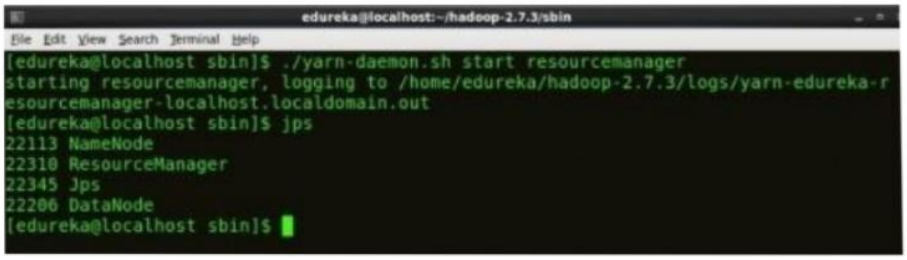


Fig:Hadoop Installation–StartingResourceManager

Start Node Manager:

The Node Manager in each machine framework is the agent which is responsible for managing containers, monitoring their resource usage and reporting the same to the Resource Manager.

Command:./yarn-daemon.sh start node manager

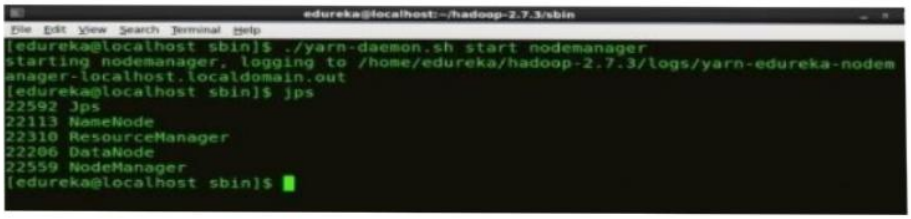


Fig:Hadoop Installation–Starting Node Manager

Start Job History Server:

Job History Server is responsible for servicing all job history related requests from client.

***Command*:**./mr-jobhistory-daemon.sh start history server

Step14:**To check that all the Hadoops ervices are up and running, run the below command.**

Command:jps

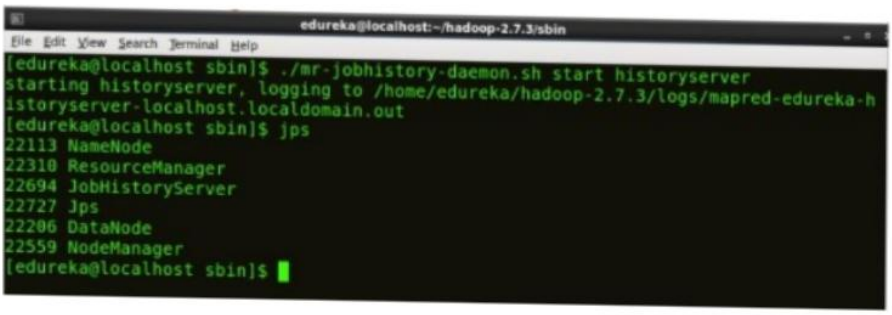


Fig:Hadoop Installation– Checking Daemons

Step15: Now open the Mozilla browser and go
To **localhost:50070/dfshealth.html** to check the NameNode interface.

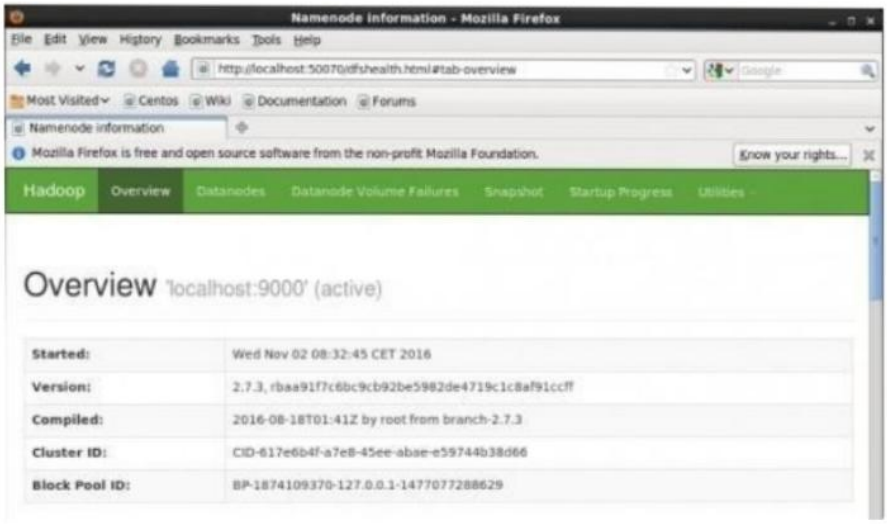


Fig: Hadoop Installation–StartingWebUI

Congratulations, you have successfully installed a single node Hadoop cluster

Result:

Thus the Hadoop one cluster was installed and simple applications executed successfully.