Name:Lisha Ingawale Roll No:2193140 Class: LY CSE IS 1

Assignment 3: Installation of Hadoop in standalone mode and distributed mode.

Theory:

In this Experiment we are installing **Hadoop 3.3.4** on Windows 10/11 - **Windows Subsystem for Linux (WSL)**.

Prerequisites:

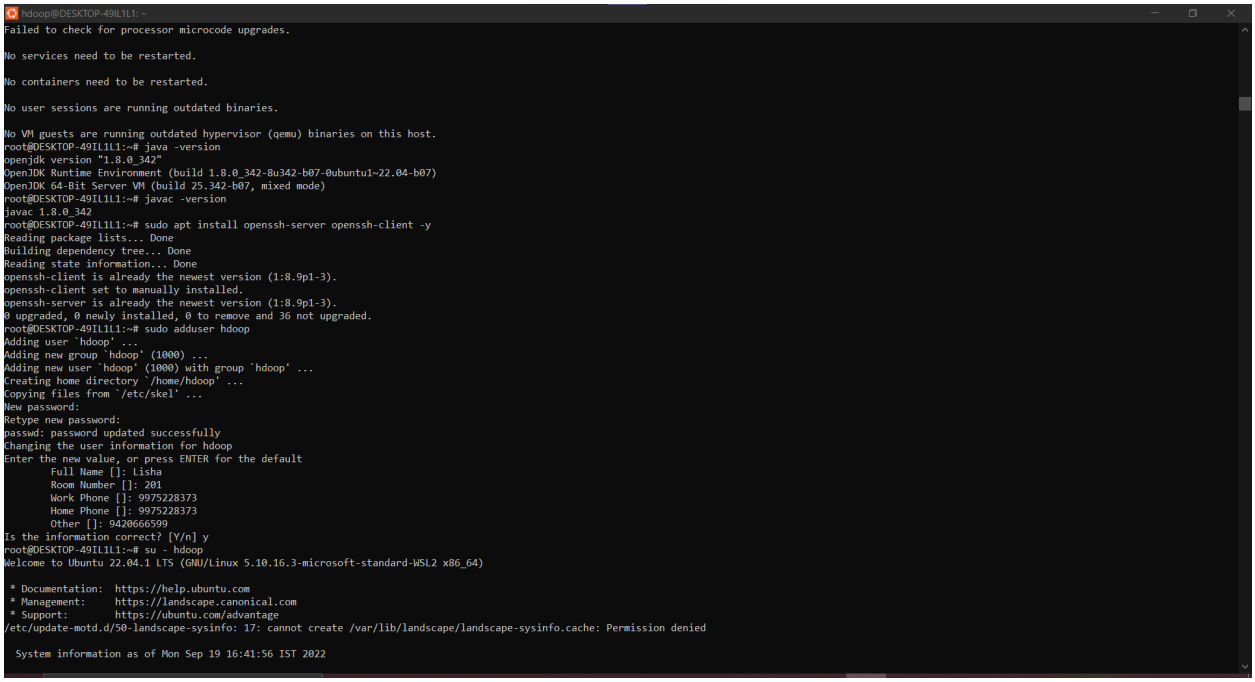
1. Windows 10/11 with Windows Subsystem for Linux turner on.
2. Any Linux Terminal application from **Microsoft Store** (Ubuntu – 20.04 or above).
3. JDK 8 or above installed in WSL.

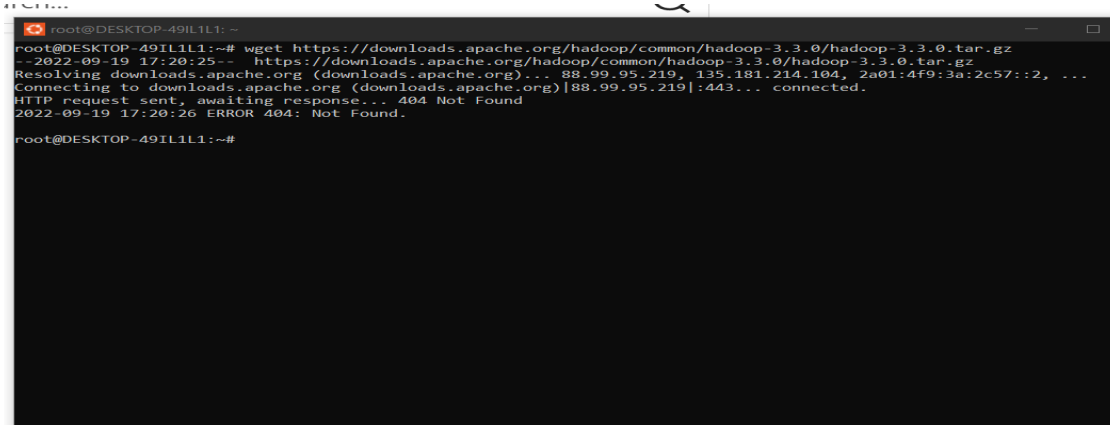
WSL:

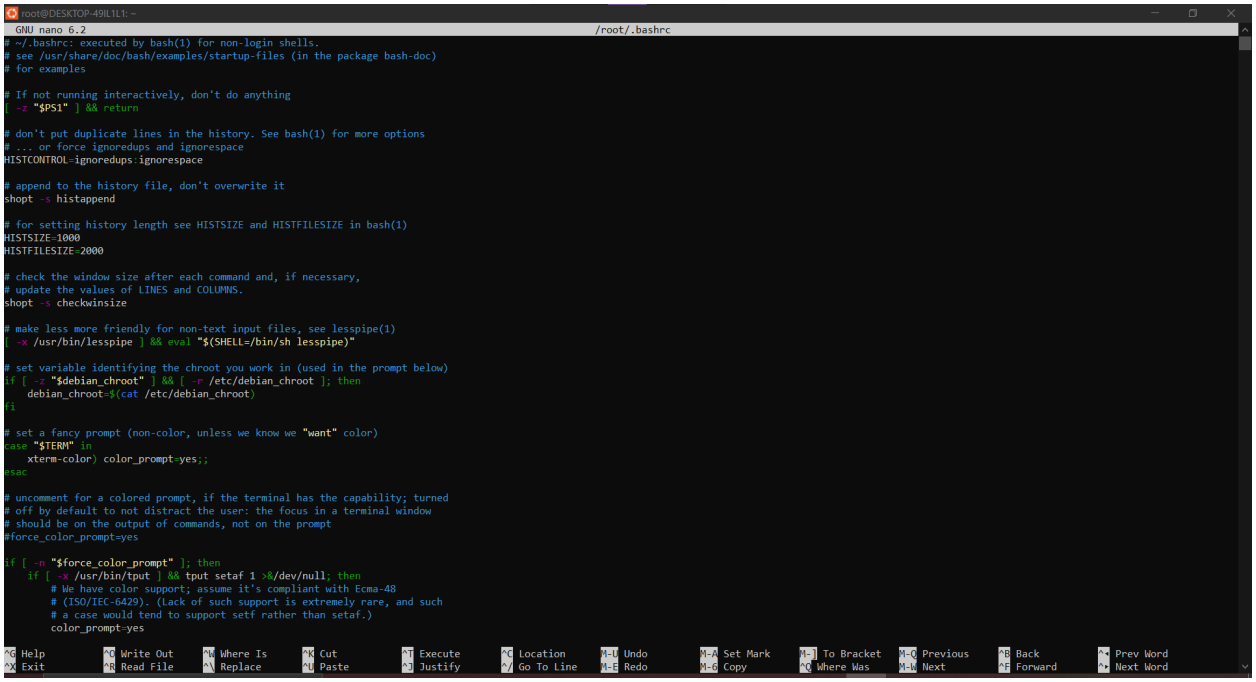
The Windows Subsystem for Linux lets developers run a GNU/Linux environment -- including most command-line tools, utilities, and applications -- directly on Windows, unmodified, without the overhead of a traditional virtual machine or dual boot setup.

Procedure:

1. Configure passphraseless ssh:


hdoop@DESKTOP-49IL1L1: ~
Failed to check for processor microcode upgrades.
No services need to be restarted.
No containers need to be restarted.
No user sessions are running outdated binaries.
No VM guests are running outdated hypervisor (qemu) binaries on this host.
root@DESKTOP-49IL1L1:~# java -version
openjdk version "1.8.0_342"
OpenJDK Runtime Environment (build 1.8.0_342-8u342-b07-0ubuntu1~22.04-b07)
OpenJDK 64-Bit Server VM (build 25.342-b07, mixed mode)
root@DESKTOP-49IL1L1:~# javac -version
javac 1.8.0_342
root@DESKTOP-49IL1L1:~# sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openssh-client is already the newest version (1:8.9p1-3).
openssh-client set to manually installed.
openssh-server is already the newest version (1:8.9p1-3).
0 upgraded, 0 newly installed, 0 to remove and 36 not upgraded.
root@DESKTOP-49IL1L1:~# sudo adduser hdoop
Adding user `hdoop' ...
Adding new group `hdoop' (1000) ...
Adding new user `hdoop' (1000) with group `hdoop' ...
Creating home directory `/home/hdoop' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hdoop
Enter the new value, or press ENTER for the default
Full Name []: Lisha
Room Number []: 201
Work Phone []: 9975228373
Home Phone []: 9975228373
Other []: 9420666599
Is the information correct? [Y/n] y
root@DESKTOP-49IL1L1:~# su - hdoop
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.10.16.3-microsoft-standard-WSL2 x86_64)
* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage
/etc/update-motd.d/50-landscape-sysinfo: 17: cannot create /var/lib/landscape/landscape-sysinfo.cache: Permission denied
System information as of Mon Sep 19 16:41:56 IST 2022


root@DESKTOP-49IL1L1: ~
root@DESKTOP-49IL1L1:~# wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
--2022-09-19 17:20:25-- https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... connected.
HTTP request sent, awaiting response... 404 Not Found
2022-09-19 17:20:26 ERROR 404: Not Found.
root@DESKTOP-49IL1L1:~#



1. Unzip Hadoop Binary:

```
mkdir ~/hadoop
```

```
tar -xvzf hadoop-3.3.4.tar.gz -C ~/hadoop
```

```
cd ~/hadoop/hadoop-3.3.4/
```

2. Setup Environment Variables:

```
nano ~/.bashrc
```

```
GNU nano 4.8                                    /home/hduser/.bashrc

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi


#Set Hadoop-related environment variables
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_HOME=~/hadoop/hadoop-3.3.4
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
```

```
source ~/.bashrc
```

Hadoop is Now installed.

We can check the version using:

1. Configure Hadoop in Single Node Mode:

```
cd ~/hadoop/hadoop-3.3.4/etc/hadoop
```

Edit file hadoop-env.sh

```
nano hadoop-env.sh
```

Set Java environment variable as,

```
  GNU nano 4.8                                                    hadoop-env.sh
# export HADOOP_ENABLE_BUILD_PATHS="true"

#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs


###
# Registry DNS specific parameters
###
# For privileged registry DNS, user to run as after dropping privileges
# This will replace the hadoop.id.str Java property in secure mode.
# export HADOOP_REGISTRYDNS_SECURE_USER=yarn

# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="-jvm server"

# Java path
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

Edit file core-site.xml

```
nano core-site.xml
```

Add the following configuration

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
</configuration>
```

Edit file hdfs-site.xml

```
nano hdfs-site.xml
```

Add the following configuration

```
<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
</configuration>
```

Edit file mapred-site.xml

```
nano mapred-site.xml
```

Add the following configuration

```
<configuration>
        <property>
                <name>mapreduce.framework.name</name>
                <value>yarn</value>
        </property>
        <property>
                <name>mapreduce.application.classpath</name>
                <value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
        </property>
</configuration>
```

Edit file yarn-site.xml

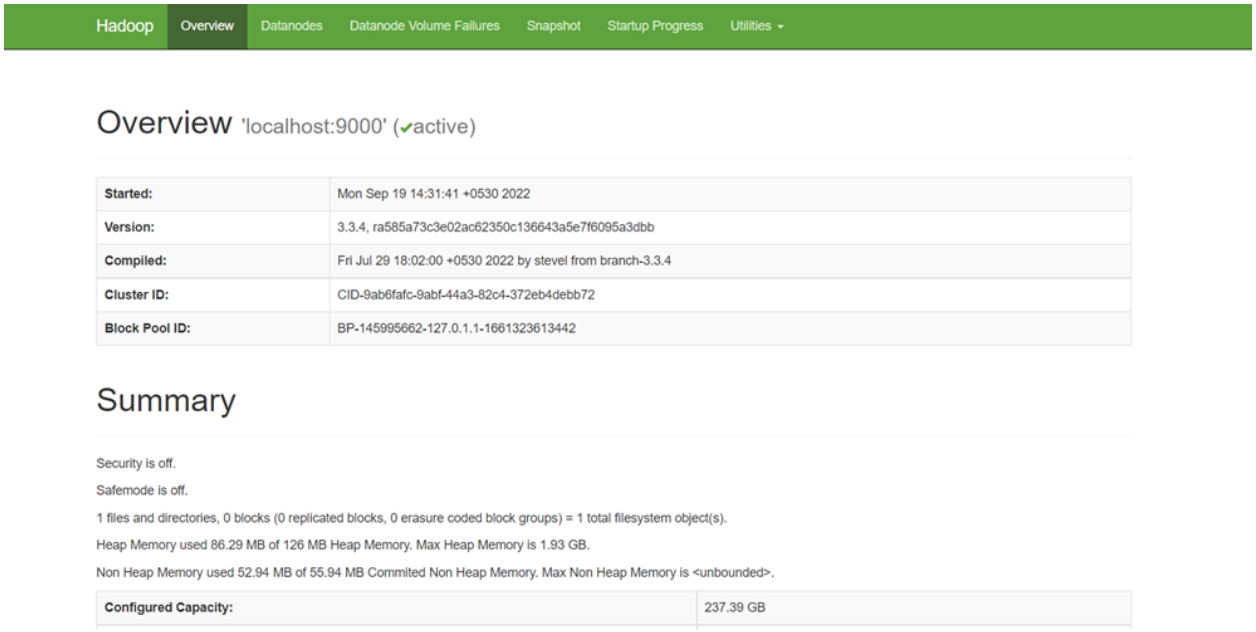
```
nano yarn-site.xml
```

Add the following configuration

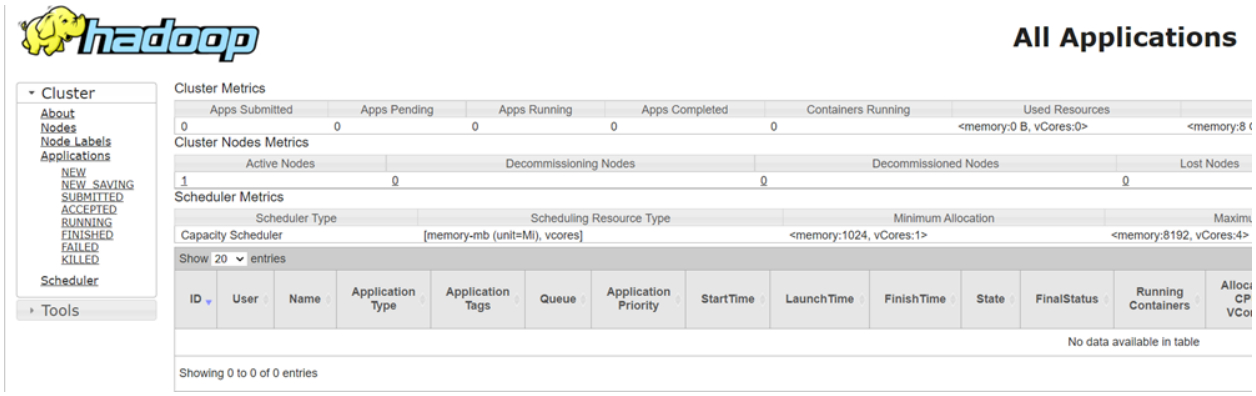
```
<configuration>

<!-- Site specific YARN configuration properties -->

        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
        <property>
                <name>yarn.nodemanager.env-whitelist</name>
                <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
        </property>

</configuration>
```

2. Format Namenode:

```
cd ~/hadoop/hadoop-3.3.4
bin/hdfs namenode -format
```



View YARN Web Portal



Conclusion:

We have Successfully Installed and Configured Hadoop in Single Node Mode.