

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 turned 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:**

```
hadoop@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No user sessions are running outdated binaries.

No VM guests are running outdated hypervisor (qemu) binaries on this host.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-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 hadoop
Adding user 'hadoop' ...
Adding new group 'hadoop' (1000) ...
Adding new user 'hadoop' (1000) with group 'hadoop' ...
Creating home directory '/home/hadoop' ...
Copying files from '/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
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 - hadoop
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:~#
```

```
root@DESKTOP-49L1CF: /root/.bashrc
GNU nano 6.2
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
[ -z "$PS1" ] && return

# don't put duplicate lines in the history. See bash(1) for more options
# ... or force ignoredups and ignorespace
HISTCONTROL=ignoredups ignorespace

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

# check the window size after each command and, if necessary,
# update the values of LINES and COLUMNS.
shopt -s checkwinsize

# make less more friendly for non-text input files, see lesspipe(1)
[ -x /usr/bin/lesspipe ] && eval "$(SHELL=/bin/sh lesspipe)"

# set variable identifying the chroot you work in (used in the prompt below)
if [ -z "$debian_chroot" ] && [ -r /etc/debian_chroot ]; then
    debian_chroot=$(cat /etc/debian_chroot)
fi

# set a fancy prompt (non-color, unless we know we "want" color)
case "$TERM" in
    xterm-color) color_prompt=yes;;
esac

# uncomment for a colored prompt, if the terminal has the capability; turned
# off by default to not distract the user: the focus in a terminal window
# should be on the output of commands, not on the prompt
#force_color_prompt=yes

if [ -n "$force_color_prompt" ]; then
    if [ -x /usr/bin/tput ] && tput setaf 1 >/dev/null; then
        # We have color support; assume it's compliant with Ecma-48
        # (ISO/IEC-6429). (Lack of such support is extremely rare, and such
        # a case would tend to support setf rather than setaf.)
        color_prompt=yes
    fi
fi

Help      Write Out  Where Is  Cut       Execute   Location  Undo      Set Mark  To Bracket Previous  Back      Prev Word
Exit      Read File  Replace   Paste     Justify   Go To Line Redo      Copy      Where Was Next      Forward   Next Word
```

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
```



Overview 'localhost:9000' (✓active)

Started:	Mon Sep 19 14:31:41 +0530 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb
Compiled:	Fri Jul 29 18:02:00 +0530 2022 by stvel from branch-3.3.4
Cluster ID:	CID-9ab8fafc-9abf-44a3-82c4-372eb4debb72
Block Pool ID:	BP-145995662-127.0.1.1-1661323613442

Summary

Security is off.	
Safemode is off.	
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).	
Heap Memory used 86.29 MB of 126 MB Heap Memory. Max Heap Memory is 1.93 GB.	
Non Heap Memory used 52.94 MB of 55.94 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.	
Configured Capacity:	237.39 GB

View YARN Web Portal



Cluster
About
Nodes
Node Labels
Applications
NEW
NEW SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools

All Applications

Cluster Metrics

Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Used Resources

0
0
0
0
0
<memory:0 B, vCores:0>
<memory:8 B, vCores:0>

Cluster Nodes Metrics

Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes

1
0
0
0

Scheduler Metrics

Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation

Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU	Allocated Memory
No data available in table														

Showing 0 to 0 of 0 entries

Conclusion:

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

