Installation de Hadoop (Stand-Alone mode)

Step 1 - Installing Ubuntu on Windows 10:

Follow these steps to install ubuntu image as an application in windows 10:

https://answers.microsoft.com/en-us/insider/forum/all/how-to-enable-the-windows-subsystem-for-linux/16e8f2e8-4a6a-4325-a89a-fd28c7841775?auth=1

Step 2 - Installation of Java:

Open an Ubuntu Terminal and run the following commands:

```
$ sudo apt-get update
```

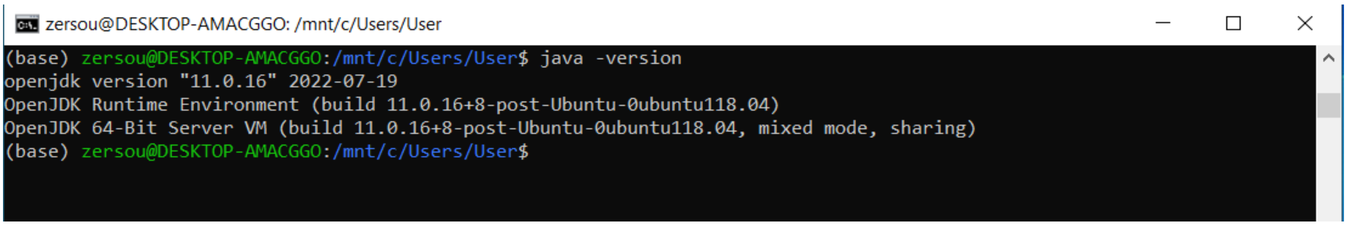
Install Java:

```
$ sudo apt-get install default-jdk
```

Check the Java version:

```
$ java -version
```

Output:



zersou@DESKTOP-AMACGGO: /mnt/c/Users/User

```
(base) zersou@DESKTOP-AMACGGO:/mnt/c/Users/User$ java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-0ubuntu118.04)
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-0ubuntu118.04, mixed mode, sharing)
(base) zersou@DESKTOP-AMACGGO:/mnt/c/Users/User$
```

Step 2 - Download and Install Hadoop:

Open an Ubuntu Terminal and run the following commands:

Download Hadoop:

```
$  sudo wget
https://archive.apache.org/dist/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.
tar.gz  --no-check-certificate
```

Extract Hadoop file:

```
$ sudo tar -xzvf hadoop-3.2.3.tar.gz
```

Move the installed file

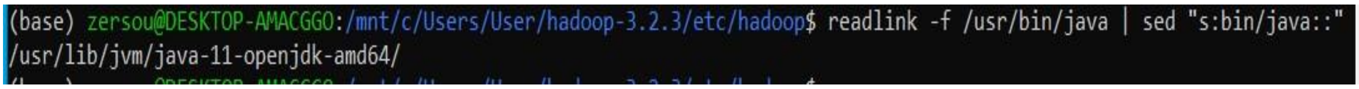
```
$ sudo mv hadoop-3.2.3 /usr/local/hadoop
```

Step 3 - Configure Hadoop and Java Home:

Look for the java installation root file:

```
$ readlink -f /usr/bin/java | sed "s:bin/java::"
```

Output:



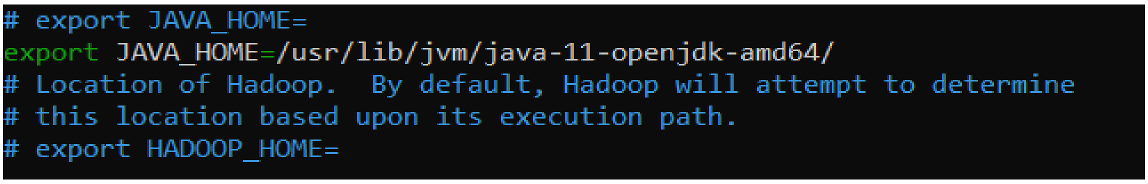
```
(base) zersou@DESKTOP-AMACGGO:/mnt/c/Users/User/hadoop-3.2.3/etc/hadoop$ readlink -f /usr/bin/java | sed "s:bin/java::"
/usr/lib/jvm/java-11-openjdk-amd64/
```

Open the file hadoop-env.sh:

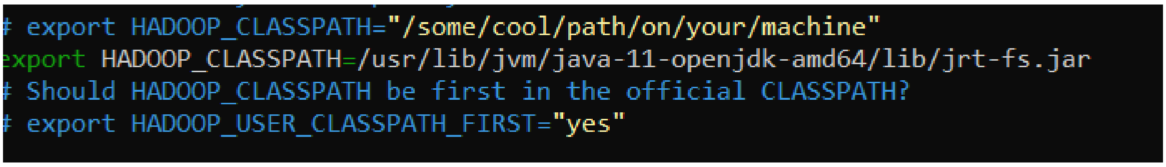
```
$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

In the hadoop-env.sh file, change JAVA_HOME and HADOOP_CLASSPATH as follows:

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/
export HADOOP_CLASSPATH=/usr/lib/jvm/java-11-openjdk-amd64/lib/jrt-fs.jar



```
# export JAVA_HOME=
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/
# Location of Hadoop.  By default, Hadoop will attempt to determine
# this location based upon its execution path.
# export HADOOP_HOME=
```



```
# export HADOOP_CLASSPATH="/some/cool/path/on/your/machine"
xport HADOOP_CLASSPATH=/usr/lib/jvm/java-11-openjdk-amd64/lib/jrt-fs.jar
# Should HADOOP_CLASSPATH be first in the official CLASSPATH?
# export HADOOP_USER_CLASSPATH_FIRST="yes"
```

Step 4: update HOME/.bashrc

```
$ nano ~/.bashrc
$ source ~/.bashrc
```

Add the following lines to the end of the $HOME/.bashrc file

```
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

Step 5: Formatting the HDFS filesystem via the NameNode

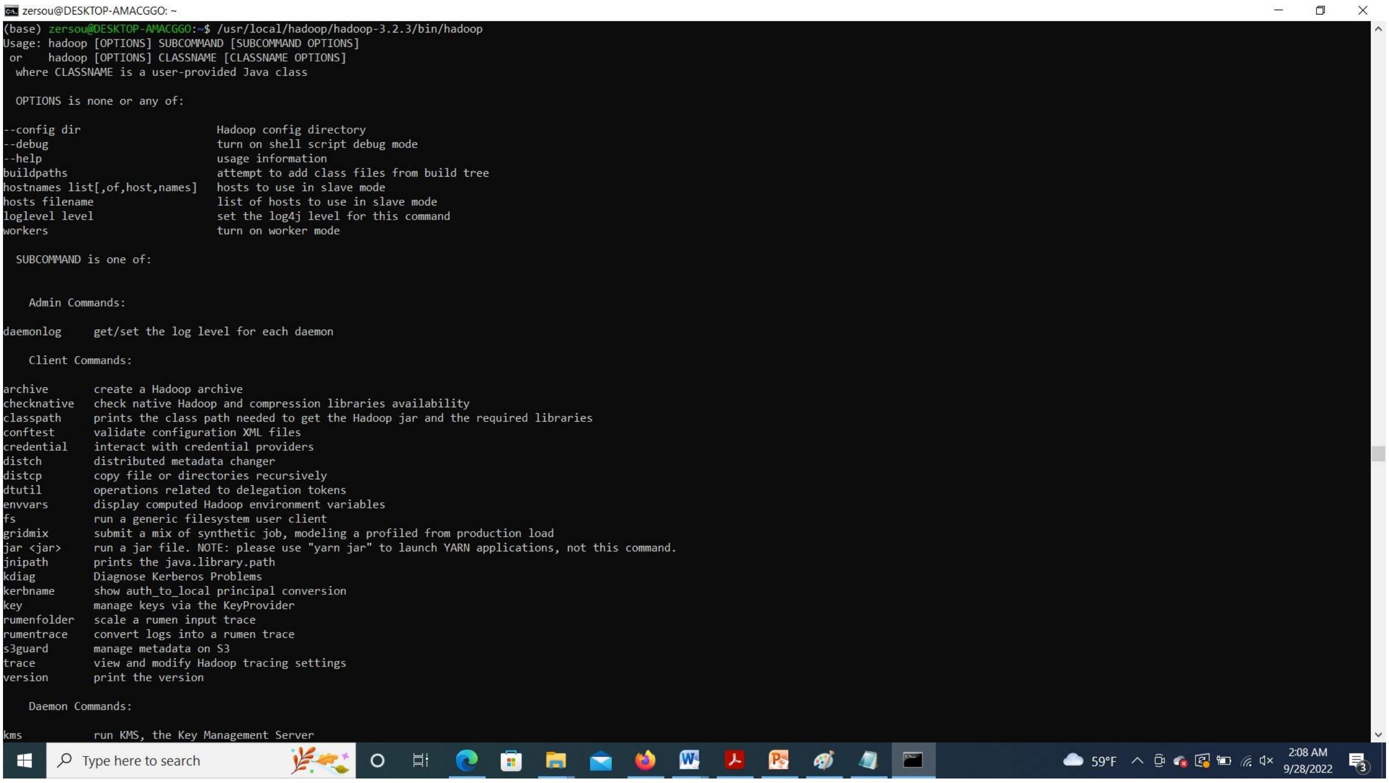
Run this command:

```
/usr/local/hadoop/bin/hadoop namenode -format
```

Step 6 - Test Hadoop

```
$ /usr/local/hadoop/bin/hadoop
```

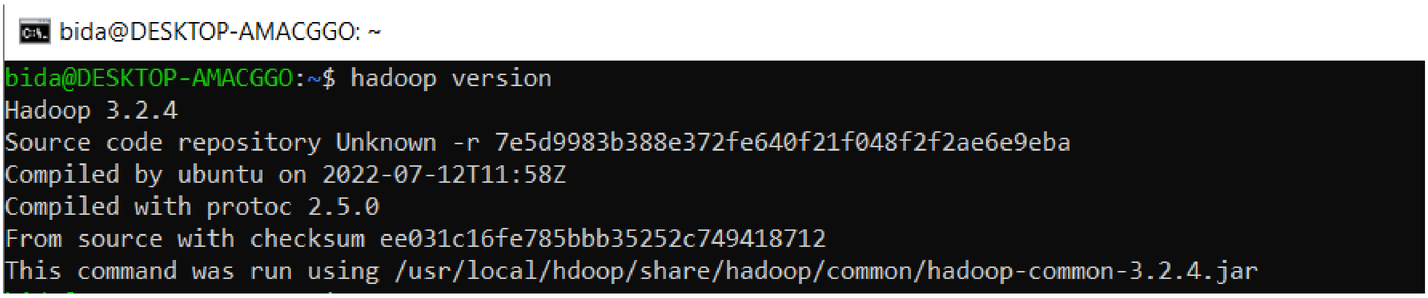
Output:



```
(base) zersou@DESKTOP-AMACGGO:~$ /usr/local/hadoop/hadoop-3.2.3/bin/hadoop
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
 or    hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
  where CLASSNAME is a user-provided Java class

  OPTIONS is none or any of:

--config dir                     Hadoop config directory
--debug                          turn on shell script debug mode
--help                           usage information
buildpaths                       attempt to add class files from build tree
hostnames list[,of,host,names]   hosts to use in slave mode
hosts filename                   list of hosts to use in slave mode
loglevel level                   set the log4j level for this command
workers                          turn on worker mode

  SUBCOMMAND is one of:


    Admin Commands:

daemonlog     get/set the log level for each daemon

    Client Commands:

archive       create a Hadoop archive
checknative   check native Hadoop and compression libraries availability
classpath     prints the class path needed to get the Hadoop jar and the required libraries
conftest      validate configuration XML files
credential    interact with credential providers
distch        distributed metadata changer
distcp        copy file or directories recursively
dtutil        operations related to delegation tokens
envvars       display computed Hadoop environment variables
fs            run a generic filesystem user client
gridmix       submit a mix of synthetic job, modeling a profiled from production load
jar <jar>     run a jar file. NOTE: please use "yarn jar" to launch YARN applications, not this command.
jnipath       prints the java.library.path
kdiag         Diagnose Kerberos Problems
kerbname      show auth_to_local principal conversion
key           manage keys via the KeyProvider
rumenfolder   scale a rumen input trace
rumentrace    convert logs into a rumen trace
s3guard       manage metadata on S3
trace         view and modify Hadoop tracing settings
version       print the version

    Daemon Commands:

kms           run KMS, the Key Management Server
```

Or check hadoop by this command:

```
$ hadoop version
```



```
bida@DESKTOP-AMACGGO:~$ hadoop version
Hadoop 3.2.4
Source code repository Unknown -r 7e5d9983b388e372fe640f21f048f2f2ae6e9eba
Compiled by ubuntu on 2022-07-12T11:58Z
Compiled with protoc 2.5.0
From source with checksum ee031c16fe785bbb35252c749418712
This command was run using /usr/local/hdoop/share/hadoop/common/hadoop-common-3.2.4.jar
```