First Instruction

Manual Installation of Big Data Utilities (Hadoop, Sqoop, Pig, and Python)

via Bash Shell Commands

Step 1/8:

Open the terminal, type **`sudo su -`** and press the **`Enter`** key to switch to the root user.

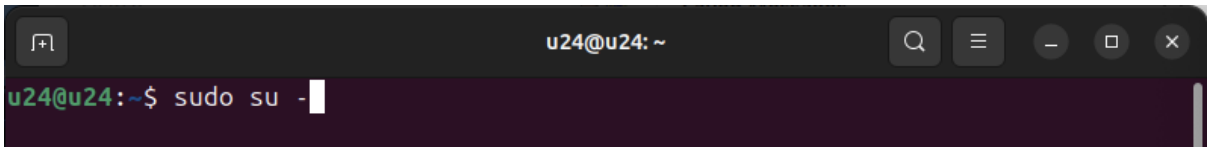

Figure 1. Switching to the root user.

Step 2/8:

Enter the password for the current user and press the **`Enter`** key.

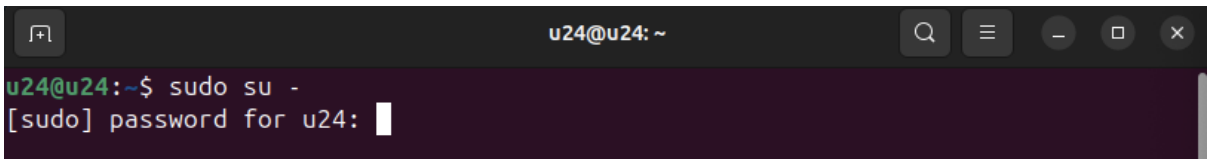

Figure 2. Filling the password for the current user.

Step 3/8:

Type the following commands to install Java openjdk version "11.0.26".

```
1: apt-get update -y

2: apt install openjdk-11-jdk -y

3: echo JAVA_HOME=\"/usr/lib/jvm/java-11-openjdk-amd64/\" >>
/etc/environment source /etc/environment

4: java -version
```

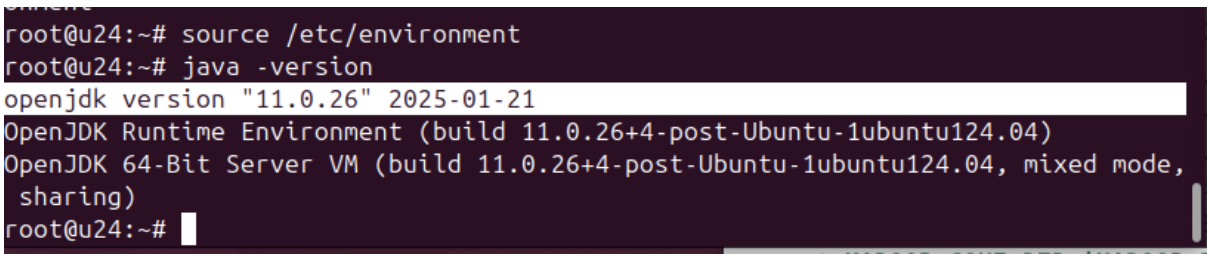

Figure 3. Java installation.

Step 4/8:

Type the following commands to install SSH.

```
1: apt-get install openssh-server -y

2: yes "" | ssh-keygen -t rsa -P ""

3: cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys

4: ssh-keyscan -H localhost >> ~/.ssh/known_hosts

5: ssh-keyscan -H 0.0.0.0 >> ~/.ssh/known_hosts

6: sudo systemctl restart ssh

7: sudo systemctl --no-pager status ssh
```

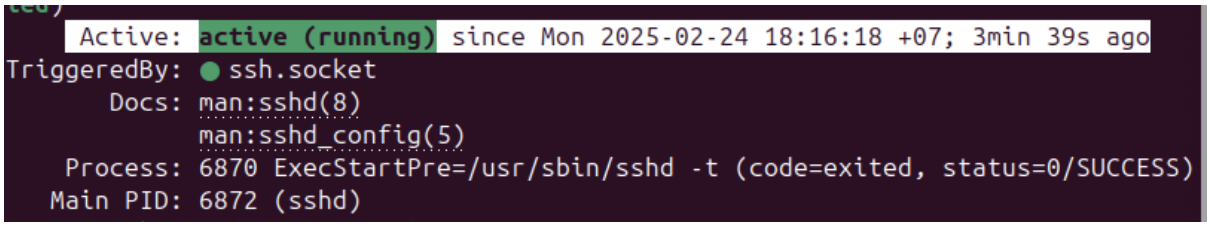

Figure 4. SSH installation.

Step 5/8:

Type the following commands to install Hadoop version 2.10.2.

```
1: wget https://downloads.apache.org/hadoop/common/hadoop-
2.10.2/hadoop-2.10.2.tar.gz

2: tar -zxvf hadoop-2.10.2.tar.gz

3: rm -rf hadoop-2.10.2.tar.gz

4: mkdir /usr/local/hadoop

5: mv hadoop-2.10.2/* /usr/local/hadoop/

6: rm -rf hadoop-2.10.2

7: echo -e "
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64

export HADOOP_INSTALL=/usr/local/hadoop

export PATH=\$PATH:\$HADOOP_INSTALL/bin
```

```
export PATH=\$PATH:\$HADOOP_INSTALL/sbin

export HADOOP_MAPRED_HOME=\$HADOOP_INSTALL

export HADOOP_COMMON_HOME=\$HADOOP_INSTALL

export HADOOP_HDFS_HOME=\$HADOOP_INSTALL

export YARN_HOME=\$HADOOP_INSTALL

export
HADOOP_COMMON_LIB_NATIVE_DIR=\$HADOOP_INSTALL/lib/native

export HADOOP_OPTS=\"-
Djava.library.path=\$HADOOP_INSTALL/lib\"

export HADOOP_PREFIX=\$HADOOP_INSTALL

export HADOOP_CONF_DIR=\$HADOOP_PREFIX/etc/hadoop" >>
~/.bashrc

8: source ~/.bashrc

9: mkdir -p /usr/local/hadoop_store/hdfs

10: cd /usr/local/hadoop_store/hdfs

11: mkdir namenode

12: mkdir datanode

13: echo -n > /usr/local/hadoop/etc/hadoop/hdfs-site.xml

14: echo -e "<?xml version=\"1.0\" encoding=\"UTF-8\"?>

<?xml-stylesheet type=\"text/xsl\"
href=\"configuration.xsl\"?>

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

<description>Default block replication.</description>

</property>
```

```
<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop_store/hdfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/local/hadoop_store/hdfs/datanode</value>

</property>

</configuration>" >> /usr/local/hadoop/etc/hadoop/hdfs-
site.xml

15: mkdir -p /app/hadoop/tmp

16: echo -n > /usr/local/hadoop/etc/hadoop/core-site.xml

17: echo -e "<?xml version=\"1.0\" encoding=\"UTF-8\"?>

<?xml-stylesheet type=\"text/xsl\"
href=\"configuration.xsl\"?>

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/app/hadoop/tmp</value>

<description>A base for other temporary
directories.</description>

</property>

<property>
```

```
<name>fs.default.name</name>

<value>hdfs://localhost:54310</value>

<description>The name of the default file
system.</description>

</property>

</configuration>" >> /usr/local/hadoop/etc/hadoop/core-
site.xml

18: cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template
/usr/local/hadoop/etc/hadoop/mapred-site.xml

19: echo -n > /usr/local/hadoop/etc/hadoop/mapred-site.xml

20: echo -e "<?xml version=\"1.0\"?>

<?xml-stylesheet type=\"text/xsl\"
href=\"configuration.xsl\"?>

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>" >> /usr/local/hadoop/etc/hadoop/mapred-
site.xml

21: echo -n > /usr/local/hadoop/etc/hadoop/yarn-site.xml

22: echo -e "<?xml version=\"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>

</configuration>" >> /usr/local/hadoop/etc/hadoop/yarn-site.xml

23: cd

24: hadoop namenode -format

25: /usr/local/hadoop/sbin/start-all.sh

26: jps

27: hadoop version
```

Figure 5. Hadoop installation.

Step 6/8:

Type the following commands to install Pig version 0.15.0.

```
1: wget https://archive.apache.org/dist/pig/pig-0.15.0/pig-0.15.0.tar.gz

2: tar -zxvf pig-0.15.0.tar.gz

3: rm -rf pig-0.15.0.tar.gz

4: mv pig-0.15.0 pig

5: mv pig /usr/local/
```

```
6: cd /usr/local/pig

7: echo -e "export PIG_HOME=/usr/local/pig export
PATH=\$PATH:/usr/local/pig/bin" >> /etc/profile

8: source /etc/profile

9: pig -version
```

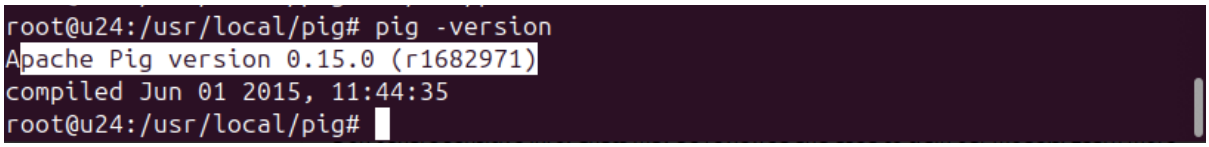

Figure 6. Pig installation.

Step 7/8:

Type the following commands to install Sqoop version 1.4.7.

```
1: wget https://archive.apache.org/dist/sqoop/1.4.7/

sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz

2: tar -xvzf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz

3: rm -rf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz

4: mv sqoop-1.4.7.bin__hadoop-2.6.0 sqoop

5: mv sqoop /usr/local/

6: echo -e "export SQOOP_HOME=/usr/local/sqoop

export PATH=\$PATH:\$SQOOP_HOME/bin" >> ~/.bashrc

7: source ~/.bashrc

8: cd /usr/local/sqoop/conf/

9: cp sqoop-env-template.sh sqoop-env.sh

10: echo -e "export HADOOP_COMMON_HOME=/usr/local/hadoop

11: export HADOOP_MAPRED_HOME=/usr/local/hadoop" >> sqoop-
env.sh

12: wget https://dlcdn.apache.org/commons/lang/binaries/

commons-lang-2.6-bin.tar.gz
```

```
13: tar -zxvf commons-lang-2.6-bin.tar.gz

14: rm -rf commons-lang-2.6-bin.tar.gz

15: cd commons-lang-2.6

16: mv commons-lang-2.6.jar /usr/local/sqoop/lib

17: cd

18: rm -rf commons-lang-2.6

19: wget https://cdn.mysql.com/archives/mysql-connector-java-
5.1/mysql-connector-java-5.1.34.tar.gz

20: tar -zxvf mysql-connector-java-5.1.34.tar.gz

21: rm -rf mysql-connector-java-5.1.34.tar.gz

22: cd mysql-connector-java-5.1.34

23: mv mysql-connector-java-5.1.34-bin.jar
/usr/local/sqoop/lib

24: sqoop version

25: ls /usr/local/sqoop/lib | grep commons-lang-2.6.jar

26: ls /usr/local/sqoop/lib | grep mysql-connector-java-
5.1.34-bin.jar
```

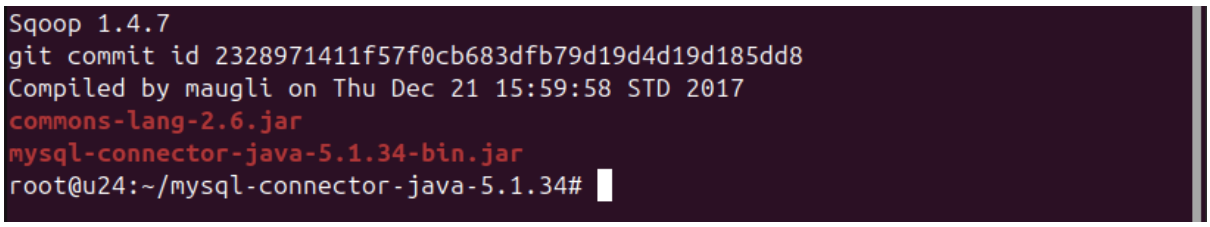

Figure 7. Sqoop installation.

Step 8/8:

Type the following commands to install Python version 3.12.3 and Pip version 24.0.

```
1: apt install -y python3

2: apt install -y python3-pip

3: python3 --version
```

4: pip3 –version

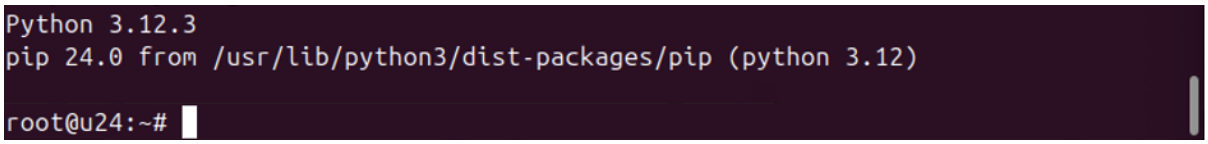

Figure 8. Python and Pip installation.