

In the name of God



Lesson : Massive of Data mining

Teacher : Dr.harati

Student : Tohid haghghi sis (۸۳۰۵۹۸۰۲۱)

Homework ۱:

How can install Hadoop in your system step by step :

-In the first step we must install VMware in windows system.

<https://soft99.ir/os/virtual-machine/۱۲۳۲-vmware-workstation.html>

-then download one of ubuntu version :

I am download ubuntu ۱۲ iso file

-then install ubuntu into VMware and start ubuntu for start install Hadoop

-We want to install Hadoop in ۳ way :

- Stand Alone
- Same Distributed
- Distributed(for run into more than ۱ system)

-check java is installed in ubuntu

Java -version

-if java do to install you can install it with

Sudo apt-get install openjdk-۸-jdk

-then check ssh is install in your system or not?

Ssh localhost

-you can install openssh

Sudo apt-get install openssh-server

```
Terminal
ubuntu@ubuntu: ~
ubuntu@ubuntu:~$ java -version
openjdk version "1.8.0_151"
OpenJDK Runtime Environment (build 1.8.0_151-8u151-b12-0ubuntu0.17.04.2-b12)
OpenJDK 64-Bit Server VM (build 25.151-b12, mixed mode)
ubuntu@ubuntu:~$ ssh localhost
ubuntu@localhost's password:
Welcome to Ubuntu 17.04 (GNU/Linux 4.10.0-19-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

 * Kata Containers are now fully integrated in Charmed Kubernetes 1.16!
   Yes, charms take the Krazy out of K8s Kata Kluster Konstruktion.

https://ubuntu.com/kubernetes/docs/release-notes

314 packages can be updated.
195 updates are security updates.

Last login: Mon Oct 14 15:34:04 2019 from 127.0.0.1
ubuntu@ubuntu:~$
```

Figure 1check java and ssh is install or not?

-then go to apache Hadoop web site and download Hadoop .

<https://hadoop.apache.org/releases.html>

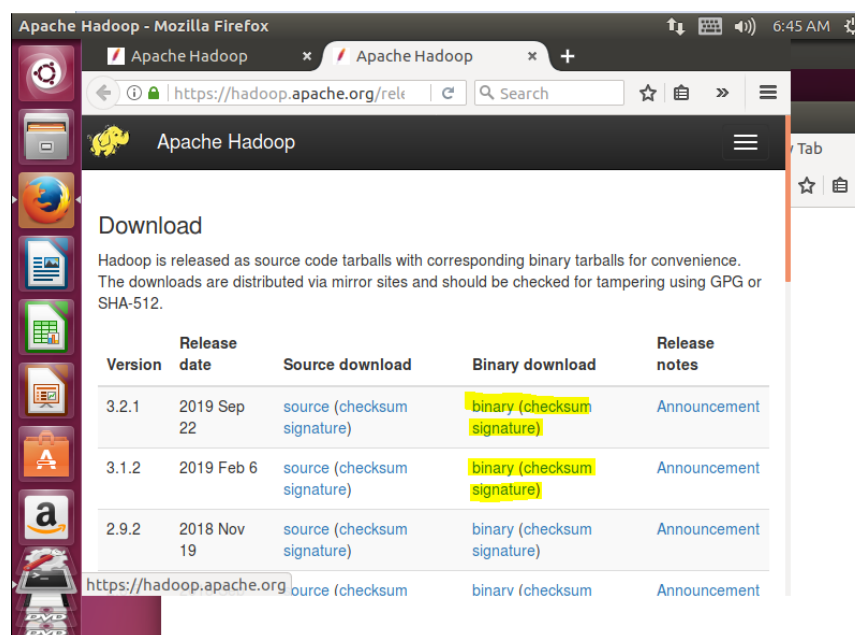


Figure 2download apache Hadoop

-After download file : Hadoop.tar.gz

- you must extract file

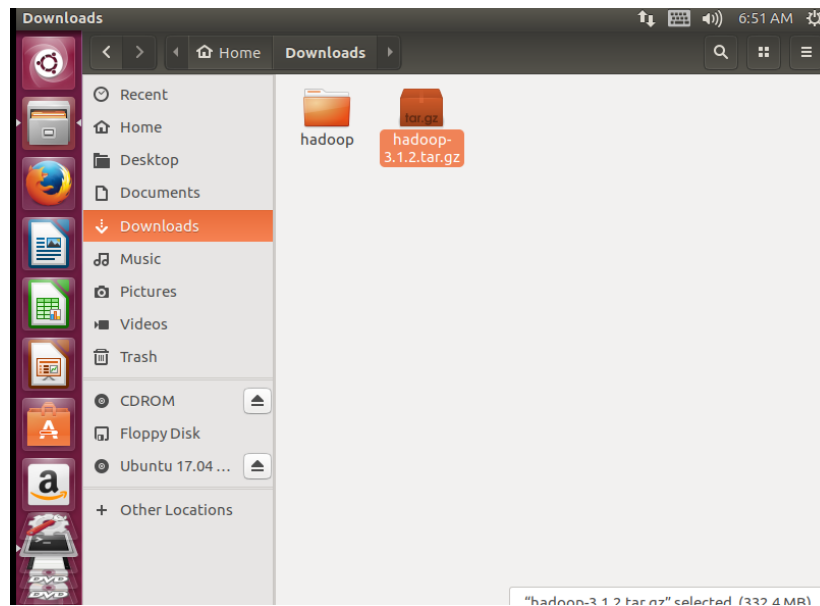


Figure 3extract Hadoop

- We must change some directories in Hadoop etc folder

```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
ubuntu@ubuntu:~/Downloads/hadoop/etc/hadoop$ ls
capacity-scheduler.xml      kms-log4j.properties
configuration.xml           kms-site.xml
container-executor.cfg      log4j.properties
core-site.xml               mapred-env.cmd
hadoop-env.cmd              mapred-env.sh
hadoop-env.sh               mapred-queues.xml.template
hadoop-env.sh.save          mapred-site.xml
hadoop-metrics2.properties  shellprofile.d
hadoop-policy.xml           ssl-client.xml.example
hadoop-user-functions.sh.example  ssl-server.xml.example
hdfs-site.xml               user_ec_policies.xml.template
httpfs-env.sh               workers
httpfs-log4j.properties     yarn-env.cmd
httpfs-signature.secret     yarn-env.sh
httpfs-site.xml             yarnservice-log4j.properties
kms-acls.xml                yarn-site.xml
kms-env.sh

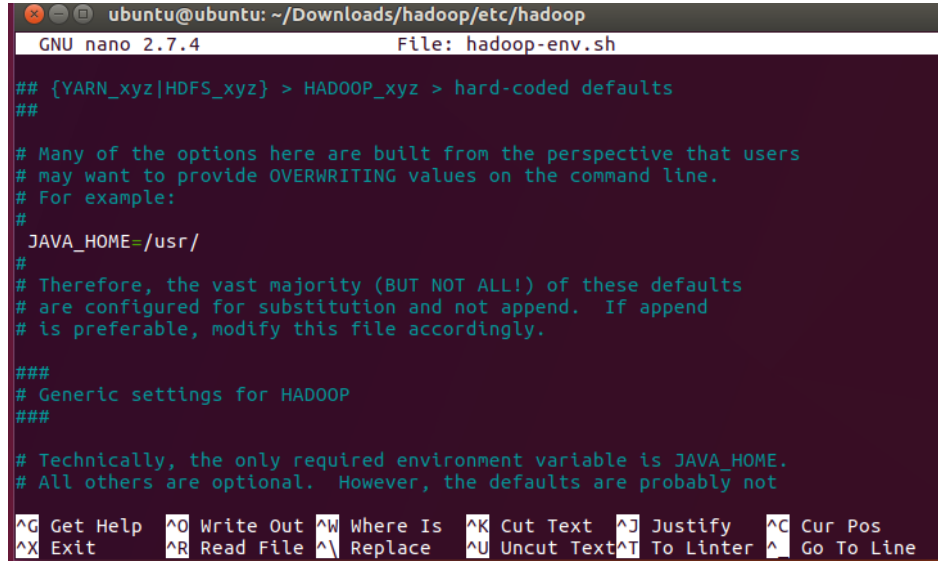
ubuntu@ubuntu:~/Downloads/hadoop/etc/hadoop$ sudo nano hadoop-env.sh
[sudo] password for ubuntu:
Use "fg" to return to nano.

[1]+  Stopped                  sudo nano hadoop-env.sh
ubuntu@ubuntu:~/Downloads/hadoop/etc/hadoop$
```

Figure 4 change this file content

-open text file with nano

Sudo nano Hadoop-env.sh



```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
GNU nano 2.7.4 File: hadoop-env.sh

## {YARN_xyz|HDFS_xyz} > HADOOP_xyz > hard-coded defaults
##

# Many of the options here are built from the perspective that users
# may want to provide OVERWRITING values on the command line.
# For example:
#
# JAVA_HOME=/usr/
#
# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
###

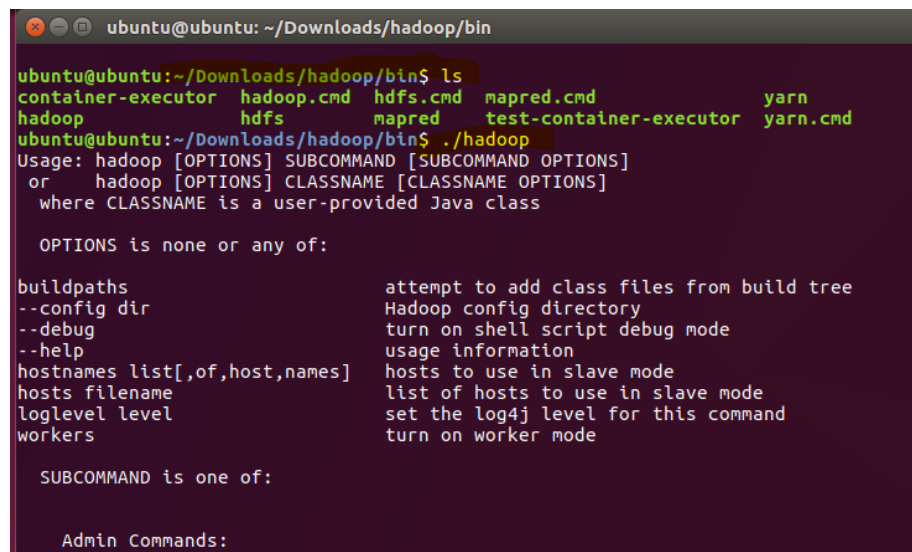
# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^I Replace ^U Uncut Text ^T To Linter ^_ Go To Line
```

Figure 5 change content with nano

-if you want to know your Hadoop in work or not ?

Go to bin folder and ./Hadoop



```
ubuntu@ubuntu: ~/Downloads/hadoop/bin

ubuntu@ubuntu:~/Downloads/hadoop/bin$ ls
container-executor  hadoop.cmd  hdfs.cmd  mapred.cmd  yarn
hadoop             hdfs       mapred    test-container-executor  yarn.cmd
ubuntu@ubuntu:~/Downloads/hadoop/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:

buildpaths          attempt to add class files from build tree
--config dir        Hadoop config directory
--debug             turn on shell script debug mode
--help              usage information
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:
```

Figure 6 run Hadoop

-then make 2 directory for save Hadoop name and data

-and give permission with sudo chmod 777 -R for Hadoop folder

```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
hadoop      hdfs      mapred      test-container-executor  yarn.cmd
ubuntu@ubuntu:~/Downloads/hadoop/bin$ cd ..
ubuntu@ubuntu:~/Downloads/hadoop$ cd ..
ubuntu@ubuntu:~/Downloads$ ls
hadoop  hadoop-3.1.2.tar.gz
ubuntu@ubuntu:~/Downloads$ mkdir hadoopstorage
ubuntu@ubuntu:~/Downloads$ ls
hadoop  hadoop-3.1.2.tar.gz  hadoopstorage
ubuntu@ubuntu:~/Downloads$ cd hadoopstorage
ubuntu@ubuntu:~/Downloads/hadoopstorage$ mkdir namedir
ubuntu@ubuntu:~/Downloads/hadoopstorage$ mkdir datadir
ubuntu@ubuntu:~/Downloads/hadoopstorage$ ls
datadir  namedir
ubuntu@ubuntu:~/Downloads/hadoopstorage$ cd ..
ubuntu@ubuntu:~/Downloads$ ls
hadoop  hadoop-3.1.2.tar.gz  hadoopstorage
ubuntu@ubuntu:~/Downloads$ sudo chmod 777 -R /home/ubuntu/Downloads/hadoop
[sudo] password for ubuntu:
ubuntu@ubuntu:~/Downloads$ cd hadoop
ubuntu@ubuntu:~/Downloads/hadoop$ ls
bin  include  libexec  NOTICE.txt sbin
etc  lib      LICENSE.txt  README.txt  share
ubuntu@ubuntu:~/Downloads/hadoop$ cd etc
ubuntu@ubuntu:~/Downloads/hadoop/etc$ ls
```

Figure 7 make hadoop data and name directory

-start to change some file content in Hadoop/etc

```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
GNU nano 2.7.4 File: core-site.xml

distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
<property>
  <name>dfs.permissions</name>
  <value>>false</value>
</property>
</configuration>

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^L Replace ^U Uncut Text ^T To Spell ^_ Go To Line
```

Figure 8 core-site.xml change

```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
GNU nano 2.7.4 File: mapred-site.xml Modified

distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
</configuration>

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell ^_ Go To Line
```

Figure 9 mapred-site.xml changes

```
ubuntu@ubuntu: ~/Downloads/hadoop/etc/hadoop
GNU nano 2.7.4 File: yarn-site.xml

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
See the License for the specific language governing permissions
limitations under the License. See accompanying LICENSE file.
-->

<configuration>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>localhost</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
</configuration>

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell ^_ Go To Line
```

Figure 10 yarn-site.xml