UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICA Y FÍSICA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

CAPTURAS DE PANTALLA - INSTALACION DE HADOOP EN UBUNTU

1.- verificamos la versión de ubuntu

lsb_release -a

uname -a

```
ubuntu@ip-172-31-29-92:~$ lsb_release -a
No LSB modules are available.
Distributor ID: Ubuntu
Description:    Ubuntu 16.04.4 LTS
Release:        16.04
Codename:       xenial
ubuntu@ip-172-31-29-92:~$ uname -a
Linux ip-172-31-29-92 4.4.0-1052-aws #61-Ubuntu SMP Mon Feb 12 23:05:58 UTC 2018
 x86_64 x86_64 x86_64 GNU/Linux
ubuntu@ip-172-31-29-92:~$ []
```

2.- verificamos si está instalado java:

java -version

```
ubuntu@ip-172-31-29-92:~$ java -version
The program 'java' can be found in the following packages:
 * default-jre
 * gcj-5-jre-headless
 * openjdk-8-jre-headless
 * gcj-4.8-jre-headless
 * gcj-4.9-jre-headless
 * openjdk-9-jre-headless
Try: sudo apt install <selected package>
```
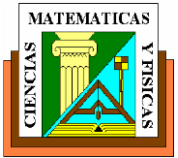
NO se encuentra instalado, procedemos con la instalación.

3.- Actualizamos los paquetes para proceder con la instalación:

sudo apt-get update

Alan Reyes Bacusoy

```
ubuntu@ip-172-31-29-92:~$ sudo apt-get update
Hit:1 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial InRelease
Get:2 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates InRelease [1
02 kB]
Get:3 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-backports InRelease
[102 kB]
Get:4 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/main Sources [868 kB
]
Get:5 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/restricted Sources [
4,808 B]
Get:6 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe Sources [7,
728 kB]
Get:7 http://security.ubuntu.com/ubuntu xenial-security InRelease [102 kB]
Get:8 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/multiverse Sources [
179 kB]
Get:9 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 Packa
ges [7,532 kB]
Get:10 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe Translatio
n-en [4,354 kB]
Get:11 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/multiverse amd64 Pa
ckages [144 kB]
Get:12 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/multiverse Translat
ion-en [106 kB]
Get:13 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main Source
s [304 kB]
Get:14 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/restricted
Sources [2,524 B]
Get:15 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/universe So
urces [200 kB]
Get:16 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/multiverse
Sources [7,944 B]
Get:17 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main amd64
Packages [755 kB]
Get:18 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main Transl
ation-en [313 kB]
Get:19 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/universe am
d64 Packages [620 kB]
Get:20 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/universe Tr
anslation-en [250 kB]
Get:21 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/multiverse
amd64 Packages [16.2 kB]
Get:22 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/multiverse
Translation-en [8,076 B]
Get:23 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-backports/main Sour
ces [3,432 B]
```

Alan Reyes Bacusoy

```
Get:29 http://security.ubuntu.com/ubuntu xenial-security/main Sources [120 kB]
Get:30 http://security.ubuntu.com/ubuntu xenial-security/restricted Sources [2,1
16 B]
Get:31 http://security.ubuntu.com/ubuntu xenial-security/universe Sources [63.5
kB]
Get:32 http://security.ubuntu.com/ubuntu xenial-security/multiverse Sources [1,5
16 B]
Get:33 http://security.ubuntu.com/ubuntu xenial-security/main amd64 Packages [47
3 kB]
Get:34 http://security.ubuntu.com/ubuntu xenial-security/main Translation-en [20
4 kB]
Get:35 http://security.ubuntu.com/ubuntu xenial-security/universe amd64 Packages
 [341 kB]
Get:36 http://security.ubuntu.com/ubuntu xenial-security/universe Translation-en
 [128 kB]
Get:37 http://security.ubuntu.com/ubuntu xenial-security/multiverse amd64 Packag
es [3,208 B]
Get:38 http://security.ubuntu.com/ubuntu xenial-security/multiverse Translation-
en [1,408 B]
Fetched 25.1 MB in 4s (5,630 kB/s)
Reading package lists... Done
```
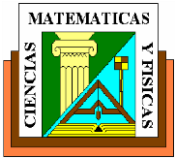
sudo apt-get -y dist-upgrade

```
Setting up xdg-user-dirs (0.15-2ubuntu6.16.04.1) ...
Setting up python3-problem-report (2.20.1-0ubuntu2.16) ...
Setting up python3-apport (2.20.1-0ubuntu2.16) ...
Setting up apport (2.20.1-0ubuntu2.16) ...
Setting up cloud-init (18.2-4-g05926e48-0ubuntu1~16.04.1) ...
Installing new version of config file /etc/cloud/cloud.cfg ...
Leaving 'diversion of /etc/init/ureadahead.conf to /etc/init/ureadahead.conf.dis
abled by cloud-init'
Setting up grub-legacy-ec2 (18.2-4-g05926e48-0ubuntu1~16.04.1) ...
Searching for GRUB installation directory ... found: /boot/grub
Searching for default file ... found: /boot/grub/default
Testing for an existing GRUB menu.lst file ... found: /boot/grub/menu.lst
Searching for splash image ... none found, skipping ...
Found kernel: /boot/vmlinuz-4.4.0-1055-aws
Found kernel: /boot/vmlinuz-4.4.0-1052-aws
Found kernel: /boot/vmlinuz-4.4.0-1055-aws
Found kernel: /boot/vmlinuz-4.4.0-1052-aws
Updating /boot/grub/menu.lst ... done

Processing triggers for libc-bin (2.23-0ubuntu10) ...
Processing triggers for initramfs-tools (0.122ubuntu8.11) ...
update-initramfs: Generating /boot/initrd.img-4.4.0-1055-aws
W: mdadm: /etc/mdadm/mdadm.conf defines no arrays.
ubuntu@ip-172-31-23-31:~$ []
```

esperamos unos minutos

4.-Reiniciamos la máquina

Alan Reyes Bacusoy

5.-Instalamos el jdk
sudo apt-get install default-jdk

```
ubuntu@ip-172-31-29-92:~$ sudo apt-get install default-jdk
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  ca-certificates-java default-jdk-headless default-jre default-jre-headless
  fontconfig fontconfig-config fonts-dejavu-core fonts-dejavu-extra
  hicolor-icon-theme java-common libasound2 libasound2-data libasyncns0
  libatk1.0-0 libatk1.0-data libavahi-client3 libavahi-common-data
  libavahi-common3 libcairo2 libcups2 libdatrie1 libdrm-amdgpu1 libdrm-intel1
  libdrm-nouveau2 libdrm-radeon1 libflac8 libfontconfig1 libgdk-pixbuf2.0-0
  libgdk-pixbuf2.0-common libgif7 libgl1-mesa-dri libgl1-mesa-glx libglapi-mesa
  libgraphite2-3 libgtk2.0-0 libgtk2.0-bin libgtk2.0-common libharfbuzz0b
  libice-dev libice6 libjbig0 libjpeg-turbo8 libjpeg8 liblcms2-2 libllvm5.0
  libnspr4 libnss3 libnss3-nssdb libogg0 libpango-1.0-0 libpangocairo-1.0-0
  libpangoft2-1.0-0 libpciaccess0 libpcsclite1 libpixman-1-0
  libpthread-stubs0-dev libpulse0 libsensors4 libsm-dev libsm6 libsndfile1
  libthai-data libthai0 libtiff5 libtxc-dxtn-s2tc0 libvorbis0a libvorbisenc2
  libx11-dev libx11-doc libx11-xcb1 libxau-dev libxcb-dri2-0 libxcb-dri3-0
  libxcb-glx0 libxcb-present0 libxcb-render0 libxcb-shm0 libxcb-sync1
  libxcb1-dev libxcomposite1 libxcursor1 libxdamage1 libxdmcp-dev libxfixes3
  libxi6 libxinerama1 libxrandr2 libxrender1 libxshmfence1 libxt-dev libxt6
  libxtst6 libxxf86vm1 openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre
  openjdk-8-jre-headless x11-common x11proto-core-dev x11proto-input-dev
  x11proto-kb-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
  default-java-plugin libasound2-plugins alsa-utils cups-common librsvg2-common
  gvfs libice-doc liblcms2-utils pcscd pulseaudio lm-sensors libsm-doc
  libxcb-doc libxt-doc openjdk-8-demo openjdk-8-source visualvm
  icedtea-8-plugin libnss-mdns fonts-ipafont-gothic fonts-ipafont-mincho
  fonts-wqy-microhei fonts-wqy-zenhei fonts-indic
The following NEW packages will be installed:
  ca-certificates-java default-jdk default-jdk-headless default-jre
  default-jre-headless fontconfig fontconfig-config fonts-dejavu-core
  fonts-dejavu-extra hicolor-icon-theme java-common libasound2 libasound2-data
  libasyncns0 libatk1.0-0 libatk1.0-data libavahi-client3 libavahi-common-data
  libavahi-common3 libcairo2 libcups2 libdatrie1 libdrm-amdgpu1 libdrm-intel1
  libdrm-nouveau2 libdrm-radeon1 libflac8 libfontconfig1 libgdk-pixbuf2.0-0
  libgdk-pixbuf2.0-common libgif7 libgl1-mesa-dri libgl1-mesa-glx libglapi-mesa
  libgraphite2-3 libgtk2.0-0 libgtk2.0-bin libgtk2.0-common libharfbuzz0b
  libice-dev libice6 libjbig0 libjpeg-turbo8 libjpeg8 liblcms2-2 libllvm5.0
  libnspr4 libnss3 libnss3-nssdb libogg0 libpango-1.0-0 libpangocairo-1.0-0
  libpangoft2-1.0-0 libpciaccess0 libpcsclite1 libpixman-1-0
  libpthread-stubs0-dev libpulse0 libsensors4 libsm-dev libsm6 libsndfile1
  libthai-data libthai0 libtiff5 libtxc-dxtn-s2tc0 libvorbis0a libvorbisenc2
```

nos pedirá confirmar la instalación oprimimos Y y enter

```
After this operation, 387 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

Alan Reyes Bacusoy
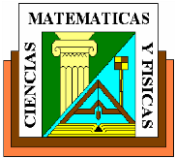
```
usr/bin/jar (jar) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/xjc to provide
usr/bin/xjc (xjc) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/schemagen to pro
vide /usr/bin/schemagen (schemagen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jps to provide
usr/bin/jps (jps) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/extcheck to prov
ide /usr/bin/extcheck (extcheck) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/rmic to provide
/usr/bin/rmic (rmic) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jstatd to provid
e /usr/bin/jstatd (jstatd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jmap to provide
/usr/bin/jmap (jmap) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jhat to provide
/usr/bin/jhat (jhat) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jdb to provide
usr/bin/jdb (jdb) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/serialver to pro
vide /usr/bin/serialver (serialver) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/wsgen to provide
 /usr/bin/wsgen (wsgen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jcmd to provide
/usr/bin/jcmd (jcmd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jarsigner to pro
vide /usr/bin/jarsigner (jarsigner) in auto mode
Setting up default-jdk-headless (2:1.8-56ubuntu2) ...
Setting up openjdk-8-jdk:amd64 (8u162-b12-0ubuntu0.16.04.2) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/appletviewer to
provide /usr/bin/appletviewer (appletviewer) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/jconsole to prov
ide /usr/bin/jconsole (jconsole) in auto mode
Setting up default-jdk (2:1.8-56ubuntu2) ...
Processing triggers for libc-bin (2.23-0ubuntu10) ...
Processing triggers for systemd (229-4ubuntu21.1) ...
Processing triggers for ureadahead (0.100.0-19) ...
Processing triggers for ca-certificates (20170717~16.04.1) ...
Updating certificates in /etc/ssl/certs...
0 added, 0 removed; done.
Running hooks in /etc/ca-certificates/update.d...

done.
done.
```

6.-Comprobamos que version de java se instaló:
java -version

```
ubuntu@ip-172-31-29-92:~$
ubuntu@ip-172-31-29-92:~$ java -version
openjdk version "1.8.0_162"
OpenJDK Runtime Environment (build 1.8.0_162-8u162-b12-0ubuntu0.16.04.2-b12)
OpenJDK 64-Bit Server VM (build 25.162-b12, mixed mode)
ubuntu@ip-172-31-29-92:~$
```

se instalo la version 8 es importante tenerlo en cuenta para la configuración

Alan Reyes Bacusoy
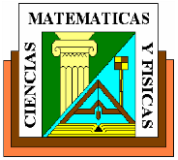
7.- Creamos una carpeta para descargar los paquetes de instalación de hadoop:
la carpeta se llamará serverHadoop
mkdir serverHadoop

```
ubuntu@ip-172-31-23-31:~$ mkdir serverHadoop
ubuntu@ip-172-31-23-31:~$ []
```

8.- Ingresamos a la carpeta creada
cd serverHadoop

```
ubuntu@ip-172-31-23-31:~$ cd serverHadoop
ubuntu@ip-172-31-23-31:~/serverHadoop$ []
```

9.- Verificamos la ruta exacta en la cual se van a descargar los paquetes de instalación de hadoop:
pwd

```
ubuntu@ip-172-31-23-31:~/serverHadoop$ pwd
/home/ubuntu/serverHadoop
ubuntu@ip-172-31-23-31:~/serverHadoop$ []
```

10.- Ingresamos al siguiente link
http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz

 De esta página podemos obtener el link de descarga:

Alan Reyes Bacusoy

11.- para realizar la descarga usamos el siguiente comando:
wget http://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz

12.- Una vez terminada la descarga descomprimimos el paquete:
tar xvzf hadoop-2.7.3.tar.gz

Alan Reyes Bacusoy

```
hadoop-2.7.3/share/hadoop/tools/lib/joda-time-2.9.4.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-azure-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/azure-storage-2.0.0.jar
hadoop-2.7.3/share/hadoop/tools/lib/commons-lang3-3.3.2.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-sls-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/metrics-core-3.0.1.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-ant-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-streaming-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-distcp-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-archives-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-rumen-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-datajoin-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/hadoop-extras-2.7.3.jar
hadoop-2.7.3/share/hadoop/tools/lib/asm-3.2.jar
hadoop-2.7.3/include/
hadoop-2.7.3/include/hdfs.h
hadoop-2.7.3/include/Pipes.hh
hadoop-2.7.3/include/TemplateFactory.hh
hadoop-2.7.3/include/StringUtils.hh
hadoop-2.7.3/include/SerialUtils.hh
hadoop-2.7.3/LICENSE.txt
hadoop-2.7.3/NOTICE.txt
hadoop-2.7.3/README.txt
ubuntu@ip-172-31-23-31:~/serverHadoop$
```

13.- Comenzamos configurando el archivo hadoop-env.sh:
~/serverHadoop/hadoop-2.7.3/etc/hadoop/hadoop-env.sh

```
ubuntu@ip-172-31-23-31:~/serverHadoop$ ~/serverHadoop/hadoop-2.7.3/etc/hadoop/ha
doop-env.sh
-bash: /home/ubuntu/serverHadoop/hadoop-2.7.3/etc/hadoop/hadoop-env.sh: Permissi
on denied
ubuntu@ip-172-31-23-31:~/serverHadoop$
```

Como podemos observar no tenemos permisos, entonces vamos a ingresar a la ruta
específica:
cd hadoop-2.7.3/etc/hadoop

```
ubuntu@ip-172-31-23-31:~/serverHadoop$ cd hadoop-2.7.3/etc/hadoop
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3/etc/hadoop$
```

14.-Editamos el archivo con el comando vi:
vi hadoop-env.sh
Editamos la siguiente línea

```
# The java implementation to use.
export JAVA_HOME=${JAVA_HOME}
```

Presionamos la letra i para empezar a editar el archivo:

Alan Reyes Bacusoy

```
# The java implementation to use.
export JAVA_HOME=${JAVA_HOME}

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol.  Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=${JSVC_HOME}

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}

# Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
for f in $HADOOP_HOME/contrib/capacity-scheduler/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  else
    export HADOOP_CLASSPATH=$f
-- INSERT --                                        25,29              22%
```

Debe quedar así:

```
# The java implementation to use.
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

Para terminar la edición presionamos la tecla ESC
Para guardar el archivo :wq y ENTER

```
# The java implementation to use.
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol.  Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=${JSVC_HOME}

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}

# Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
for f in $HADOOP_HOME/contrib/capacity-scheduler/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  else
    export HADOOP_CLASSPATH=$f
  fi
done
:wq
```

En caso de presentar inconvenientes puede consultar la siguinte gui de como usar el editor
vi http://www.unirioja.es/cu/enriquez/docencia/Quimica/vi.pdf

15.- Editamos el archivo core-site.xm, editamos las líneas de configuración:
vi core-site.xml
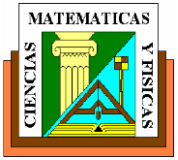
```
<configuration>
</configuration>
~
```

Alan Reyes Bacusoy

```xml
<configuration>
<property>
 <name>hadoop.tmp.dir</name>
 <value>/tmp/hadoop/data</value>
 <description>Location for HDFS.</description>
</property>
<property>
 <name>fs.default.name</name>
 <value>hdfs://localhost:54310</value>
 <description>The name of the default file system. A URI whose
 scheme and authority determine the FileSystem implementation. </description>
</property>
</configuration>
```

Debe quedar así:

```xml
<configuration>
<property>
 <name>hadoop.tmp.dir</name>
 <value>/tmp/hadoop/data</value>
 <description>Location for HDFS.</description>
</property>
<property>
 <name>fs.default.name</name>
 <value>hdfs://localhost:54310</value>
 <description>The name of the default file system. A URI whose
 scheme and authority determine the FileSystem implementation. </description>
</property>
</configuration>
```

Guardamos y salimos

16.- Salimos de la ruta con el comando cd:

```
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3/etc/hadoop$ cd
ubuntu@ip-172-31-23-31:~$ 
```

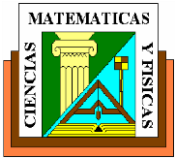17.-Creamos la ruta de la carpeta donde se guardará el directorio de datos:
sudo mkdir -p /usr/local/hadoop/hdfs/data

```
ubuntu@ip-172-31-23-31:~$ sudo mkdir -p /usr/local/hadoop/hdfs/data
ubuntu@ip-172-31-23-31:~$ 
```

sudo chown -R ubuntu:ubuntu /usr/local/hadoop/hdfs/data

```
ubuntu@ip-172-31-23-31:~$ sudo chown -R ubuntu:ubuntu /usr/local/hadoop/hdfs/data
ubuntu@ip-172-31-23-31:~$ 
```

Alan Reyes Bacusoy

18.- Ingresamos a la siguiente ruta:

cd /home/ubuntu/serverHadoop/hadoop-2.7.3/etc/hadoop

```
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3/etc/hadoop$ []
```

Editamos el archivo hdfs-site.xml:

vi hdfs-site.xml

```
<configuration>
</configuration>
~
```

```
<configuration>
<property>
 <name>dfs.replication</name>
 <value>1</value>
 <description>Default number of block replications.</description>
</property>
</configuration>
```

Debe quedar de la siguiente manera

```
<configuration>
<property>
 <name>dfs.replication</name>
 <value>1</value>
 <description>Default number of block replications.</description>
</property>
</configuration>
[]
```

19.- Realizamos una copia de los archivos mediante el siguiente comando:

cp mapred-site.xml.template mapred-site.xml

```
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3/etc/hadoop$ cp mapred-site.xm
l.template mapred-site.xml
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3/etc/hadoop$ []
```

20.- Modificamos el archivo mapred-site.xml:

vi mapred-site.xml

```
<configuration>
[]
</configuration>
```

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:54311</value>
<description>The host and port that the MapReduce job tracker runs at.
</description>
</property>
```

Alan Reyes Bacusoy

</configuration>

Debe quedar así:

```
<configuration>
 <property>
  <name>mapred.job.tracker</name>
   <value>localhost:54311</value>
   <description>The host and port that the MapReduce job tracker runs at.
   </description>
 </property>
</configuration>
```

21.-Modificamos el archivo yarn-site.xml
vi yarn-site.xml

```
<configuration>

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

</configuration>
```

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

Debe quedar de la siguiente manera :

```
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
```

22.- Configuramos el ssh para que no tenga contraseña
Salimos de la ruta mediante cd

Alan Reyes Bacusoy

```
ubuntu@ip-172-31-23-31:~$ []
```

ssh-keygen

```
ubuntu@ip-172-31-23-31:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa): []
```

Solo presionamos ENTER para que coja la configuración por defecto y no tenga contraseña:

```
ubuntu@ip-172-31-23-31:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa):
/home/ubuntu/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/ubuntu/.ssh/id_rsa.
Your public key has been saved in /home/ubuntu/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:0RjcaGQGlFy/XHZg7kTRbHHeYKYS14o6ActoyK75+gw ubuntu@ip-172-31-23-31
The key's randomart image is:
+---[RSA 2048]----+
|      oo**o. *==..|
|    . oo+=.*.=++.|
|.. o o .o.+.*.. o|
|..o o . .o.B .   |
|..      oS o .   |
| .     o         |
|E.         .     |
|oo               |
|.++              |
+----[SHA256]-----+
ubuntu@ip-172-31-23-31:~$ []
```

23.-Ejecutamos el siguiente comando:
ssh localhost

```
root@ip-172-31-23-31:~# ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:geW829sUvUZGWuZ6H/b4h3dZZk5ts4jAVVb1E+uN2wc.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Permission denied (publickey).
root@ip-172-31-23-31:~# []
```

24.-Configuramos las variables de entorno:
vi /root/.bashrc

```
root@ip-172-31-23-31:~# vi /root/.bashrc
root@ip-172-31-23-31:~# []
```

Nos vamos al final del archivo y copiamos lo siguiente:

Alan Reyes Bacusoy

```
export HADOOP_HOME=/home/ubuntu/serverHadoop/hadoop-2.7.3
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"
lzohead () {
 hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
export PATH=$PATH:$HADOOP_HOME/bin
```

Debe quedar así:

25.-Cerramos el putty y nos volvemos a conectar para comprobar la configuración mediante los siguientes comandos:

```
sudo su
echo $JAVA_HOME
echo $HADOOP_HOME
```


26.-Antes de iniciar el servicio nos dirigimos a la siguiente ruta:

```
cd /home/ubuntu/serverHadoop/hadoop-2.7.3/bin
```

Ejecutamos el siguiente formato:

```
hadoop namenode -format
```

Alan Reyes Bacusoy

```
minutes = 1,5,25
18/04/27 20:06:12 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
18/04/27 20:06:12 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total
 heap and retry cache entry expiry time is 600000 millis
18/04/27 20:06:12 INFO util.GSet: Computing capacity for map NameNodeRetryCache
18/04/27 20:06:12 INFO util.GSet: VM type        = 64-bit
18/04/27 20:06:12 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 29
7.0 KB
18/04/27 20:06:12 INFO util.GSet: capacity       = 2^15 = 32768 entries
18/04/27 20:06:12 INFO namenode.FSImage: Allocated new BlockPoolId: BP-143043306
6-172.31.23.31-1524859572283
18/04/27 20:06:12 INFO common.Storage: Storage directory /tmp/hadoop/data/dfs/na
me has been successfully formatted.
18/04/27 20:06:12 INFO namenode.FSImageFormatProtobuf: Saving image file /tmp/ha
doop/data/dfs/name/current/fsimage.ckpt_0000000000000000000 using no compression
18/04/27 20:06:12 INFO namenode.FSImageFormatProtobuf: Image file /tmp/hadoop/da
ta/dfs/name/current/fsimage.ckpt_0000000000000000000 of size 350 bytes saved in
0 seconds.
18/04/27 20:06:12 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
18/04/27 20:06:12 INFO util.ExitUtil: Exiting with status 0
18/04/27 20:06:12 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-23-31.us-west-2.compute.intern
al/172.31.23.31
************************************************************/
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3#
```

27.- Ahora procederemos a iniciar el cluster:
cd ..

sbin/start-dfs.sh

```
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3$ sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: Permission denied (publickey).
localhost: Permission denied (publickey).
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:geW829sUvUZGWuZ6H/b4h3dZZk5ts4jAVVb1E+uN2wc.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts
.
0.0.0.0: Permission denied (publickey).
```

Como podemos observar en mi caso no tiene permisos de coneccion y esto se debe a la
configuración ssh , si es necesario debemos continuar con la siguiente configuración (de los
contrario vamos al paso 28):
ssh-keygen -t rsa -P ''
ENTER
cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys

Alan Reyes Bacusoy

```
root@ip-172-31-23-31:/home/ubuntu# ssh-keygen -t rsa -P ''
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:SvX1tYqaf7C0osMD1D1H16ONJ+xeEBomTnAueBoFBs0 root@ip-172-31-23-31
The key's randomart image is:
+---[RSA 2048]----+
|o=.o..        .  |
|. E oo o o +     |
| o oo.+ B * o  .|
|  + .o * O + . ..|
| .  . . S =   .. |
|     o . . +. .  |
|      + . o.+.   |
|       + ooo .   |
|        .+oo..   |
+----[SHA256]-----+
root@ip-172-31-23-31:/home/ubuntu# cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/autho
rized_keys
root@ip-172-31-23-31:/home/ubuntu# sudo vim /etc/ssh/sshd_config
root@ip-172-31-23-31:/home/ubuntu#
```

sudo vim /etc/ssh/sshd_config
Modificamos lo siguiente:

```
# Change to no to disable tunnelled clear text passwords
PasswordAuthentication no
```

Debe quedar asi:

```
# Change to no to disable tunnelled clear text passwords
PasswordAuthentication yes
```

sudo service ssh restart

```
root@ip-172-31-23-31:/home/ubuntu# sudo service ssh restart
root@ip-172-31-23-31:/home/ubuntu#
```

Comprobamos:
ssh localhost

Alan Reyes Bacusoy

```
root@ip-172-31-23-31:/home/ubuntu# ssh localhost
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.4.0-1055-aws x86_64)

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

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

0 packages can be updated.
0 updates are security updates.


Last login: Fri Apr 27 19:34:26 2018 from 200.1.161.253
root@ip-172-31-23-31:~#
```

exit

```
root@ip-172-31-23-31:~# exit
logout
Connection to localhost closed.
root@ip-172-31-23-31:/home/ubuntu#
```

Volvemos a iniciar los servicios de hadoop

28.-Ingresamos a la siguiente ruta:
cd /home/ubuntu/serverHadoop/hadoop-2.7.3
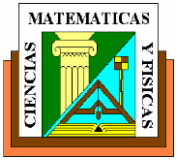ENTER
sudo su
ENTER
sbin/start-dfs.sh

```
ubuntu@ip-172-31-23-31:~/serverHadoop/hadoop-2.7.3$ sudo su
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3# sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/ubuntu/serverHadoop/hadoop-2.7.3/
logs/hadoop-root-namenode-ip-172-31-23-31.out
localhost: starting datanode, logging to /home/ubuntu/serverHadoop/hadoop-2.7.3/
logs/hadoop-root-datanode-ip-172-31-23-31.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:geW829sUvUZGWuZ6H/b4h3dZZk5ts4jAVVb1E+uN2wc.
Are you sure you want to continue connecting (yes/no)? yes
```

sbin/start-yarn.sh

```
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3# sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/serverHadoop/hadoop-2.7.3/logs
/yarn-root-resourcemanager-ip-172-31-23-31.out
localhost: starting nodemanager, logging to /home/ubuntu/serverHadoop/hadoop-2.7
.3/logs/yarn-root-nodemanager-ip-172-31-23-31.out
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3#
```

Alan Reyes Bacusoy

29.- Finalmente jps para ver todos los procesos de java que se están ejecutando:

```
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3# jps
11424 NodeManager
10022 ResourceManager
10969 DataNode
10825 NameNode
11147 SecondaryNameNode
11550 Jps
root@ip-172-31-23-31:/home/ubuntu/serverHadoop/hadoop-2.7.3#
```

30.- Como podemos ver ya se encuentra corriendo la configuración de nuestro nodo en Hadoop, ahora podemos empezar a trabajar sobre el.

Bibliografía

http://www.cs.cityu.edu.hk/~heleicui2/doc/Setup-Hadoop-2.7.3-(single-node)-on-AWS-EC2-Ubuntu-AMI.pdf
https://sagarruchandani.wordpress.com/2015/08/01/hadoop-setting-up-hadoop-2-6-0-single-node-on-aws-ec2-ubuntu-ami/
http://hadoopexam.com/singlenode.pdf
https://aws.amazon.com/es/config/?sc_channel=PS&sc_campaign=acquisition_LATAM&sc_publisher=google&sc_medium=english_config_b&sc_content=aws_config_bmm&sc_detail=%2Baws%20%2Bconfig&sc_category=config&sc_segment=222159572858&sc_matchtype=b&sc_country=LATAM&s_kwcid=AL!4422!3!222159572858!b!!g!!%2Baws%20%2Bconfig&ef_id=WrlJ6QAAALy0pSua:20180427193152:s
https://dzone.com/articles/how-set-multi-node-hadoop
https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/managing-users.html
https://www.novixys.com/blog/setup-apache-hadoop-cluster-aws-ec2/#AWS_EC2_Startup
http://telesalud.udea.edu.co/blog/wp-content/uploads/2015/09/Instalaci%C3%B3n-Hadoop-2.pdf
http://knowdimension.com/blog/administracion-de-datos/instalando-hadoop-2-7-en-un-sistema-ubuntu/

Alan Reyes Bacusoy