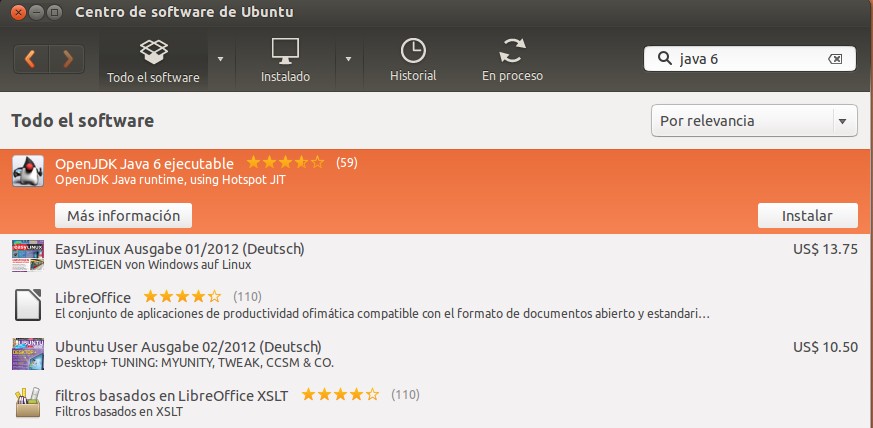
# INSTALACIÓN HADOOP

Hadoop es la herramienta que se utiliza como nodo, donde va a quedar integrado todo el ambiente. Esta herramienta permite correr todas las aplicaciones dentro del mismo cluster, por lo tanto toda la información quedara alojada allí y podrá ser utilizada sin ningún inconveniente.

A continuación se describen los pasos la instalación de Hadoop:

1. Antes de empezar con la instalación de Hadoop se debe tener en la máquina virtual Java, por lo tanto se instala Java 6 desde el centro de Software de Ubuntu, como se muestra en la siguiente figura:

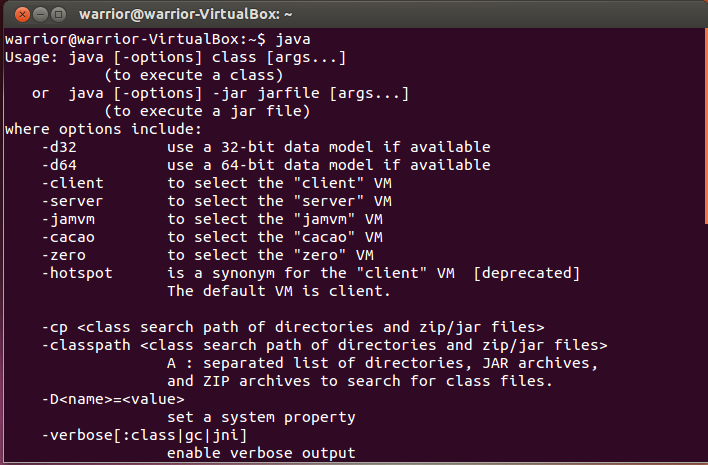
# Figura 14. Instalación Java



Fuente: Autores

1. Verificar que java quedo instalado probando desde la terminal con el comando java. Debe arrojar la información que se muestra en la siguiente figura:

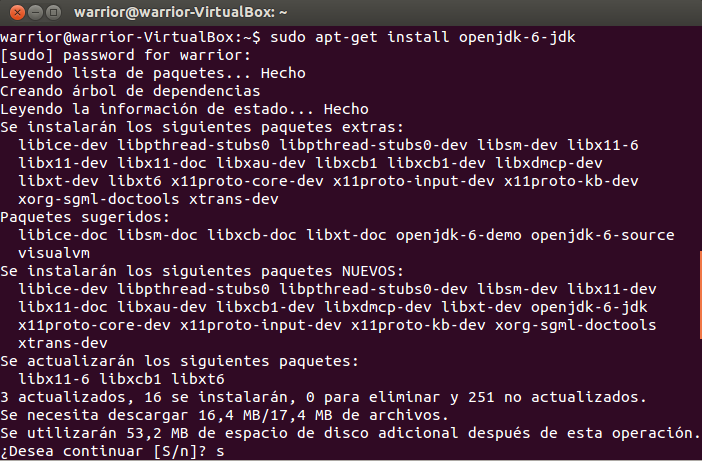
# Figura 15. Validación Java



Fuente: Autores

1. Instalar OpenJdk6 desde la terminal, con el comando sudo apt-get install openjdk-6-jdk. Cuando aparezca el mensaje ¿Desea continuar [s/n]?, escribir s como se muestra en la figura N°16, dar enter y seguir con la instalación.

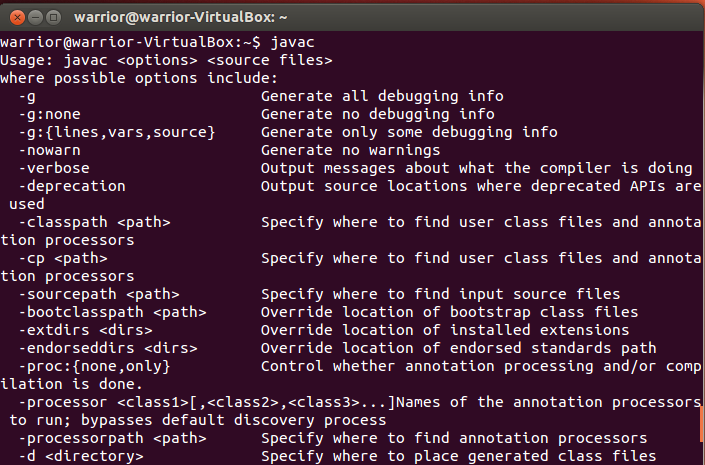
# Figura 16. Instalar OpenJdk



Fuente: Autores

1. Verificar la instalación del OpenJdk a través del comando javac. Debe arrojar la información como muestra la siguiente figura:

# Figura 17. Verificación OpenJdk



Fuente: Autores

1. Descargar la última versión de Hadoop en la página de Apache [http://apache.mirrors.tds.net/hadoop/common/](http://hadoop.apache.org/releases.html), escoger la carpeta stable, como se muestra en la figura N°18 y allí escoger el archivo con extensión .tar.gz.

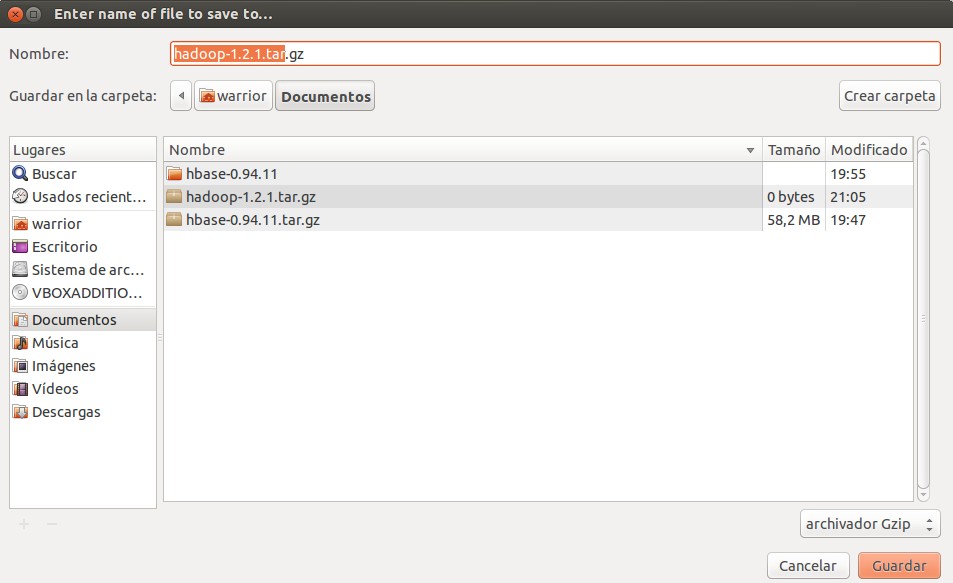
# Figura 18. Descargar Hadoop



Fuente: Autores

1. Guardar el archivo en la carpeta Documentos, como se ilustra en la siguiente imagen:

# Figura 19. Guardar archivo Hadoop



Fuente: Autores

1. Descomprimir el archivo desde la terminal a través del comando: $ tar xzf hadoop-1.2.1.tar.gz.Luego mover el archivo descomprimido a una carpeta de

nombre hadoop a través del comando mv, como se muestra en la siguiente figura:

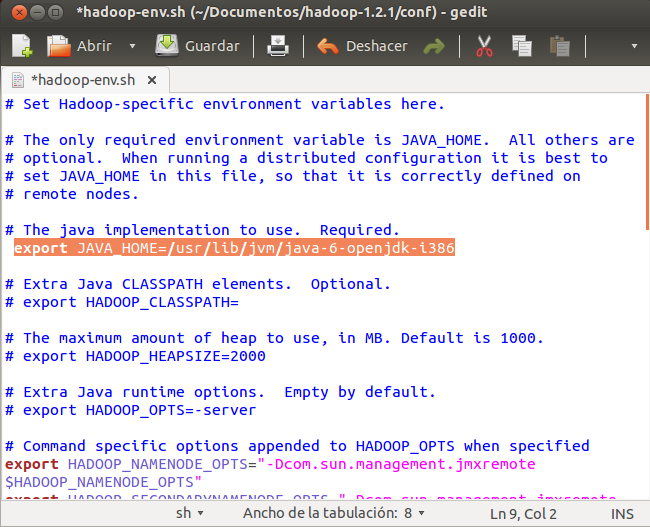
# Figura 20. Descomprimir archivo Hadoop



Fuente: Autores

1. Abrir el archivo hadoop-env.sh que se encuentra en la carpeta conf con el Editor de texto. Quitar el comentario en la dirección del JAVA HOME, como se muestra en la figura N° 21 y cambiar la dirección que se encuentra allí por la que se encuentra instalado el java, que por defecto es usr/lib/jvm/java-6- openjdk-i386. Guardar la configuración.

# Figura 21. Configurar Java Home



Fuente: Autores

A continuación se realiza la modificación de los archivos para ejecutar hadoop en un modo pseudo-distribuido, esto permitirá que las demás herramientas se puedan integrar correctamente.

1. Cambiar la configuración del archivo core-site.xml que se encuentra en la carpeta conf, como se muestra en la figura N°22, por la siguiente información:

<configuration>

<property>

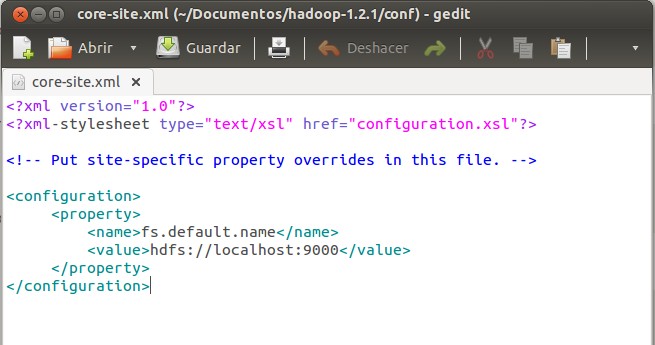
<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

# Figura 22. Configurar el archivo core



Fuente: Autores

1. Cambiar la configuración del archivo hdfs-site.xml que se encuentra en la carpeta conf, como se muestra en la figura N°23, por la siguiente información:

<configuration>

<property>

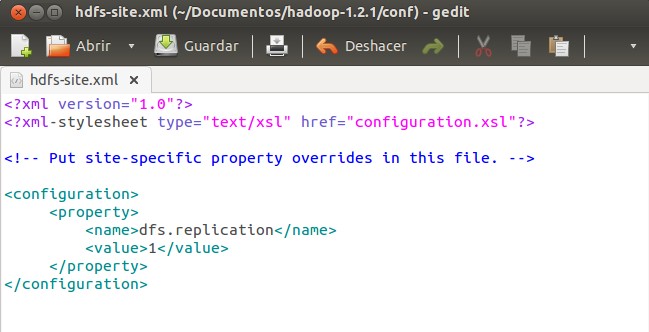
<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

# Figura 23. Configuración archivo hdfs



Fuente: Autores

1. Cambiar la configuración del archivo mapred-site.xml que se encuentra en la carpeta conf, como se muestra en la figura N°24, por la siguiente información:

<configuration>

<property>

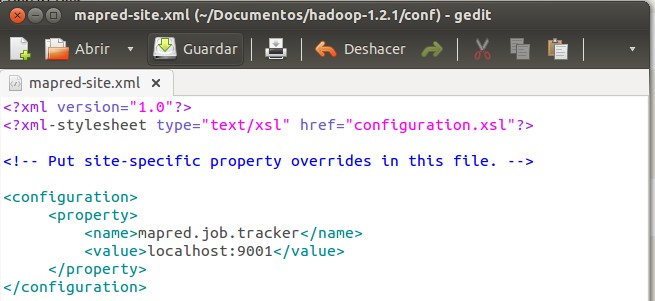
<name>mapred.job.tracker</name>

<value>localhost:9001</value>

</property>

</configuration>

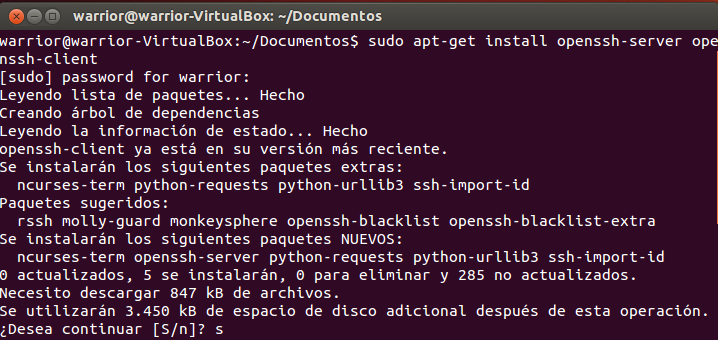
# Figura 24. Configuración archivo mapred



Fuente: Autores

1. Instalar en la terminal el complemento ssh para poder utilizar los scripts de Hadoop a través del siguiente comando: $ sudo apt-get install openssh-server openssh-client. Cuando aparezca el mensaje ¿Desea continuar [s/n]?, como aparece en la figura N°25, se escoge la opción “s” y se sigue la instalación.

# Figura 25. Instalación ssh



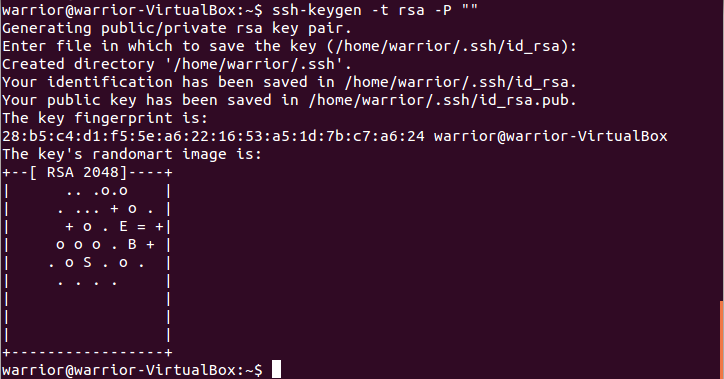
Fuente: Autores

1. Configurar el complemento ssh, como se muestra en la figura N°26, a través de los siguientes comandos:

$ ssh-keygen -t rsa -P "".

$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys.

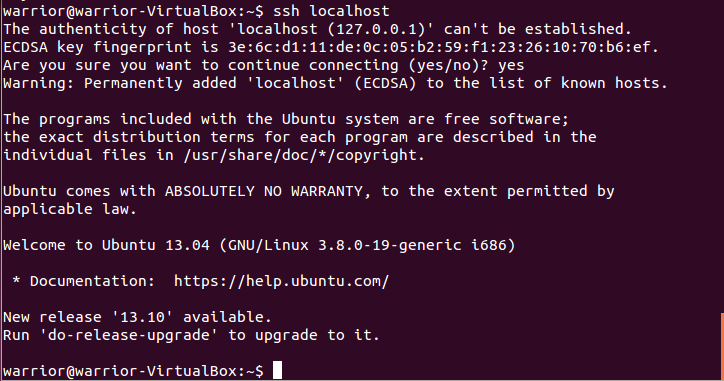
# Figura 26. Configuración ssh



Fuente: Autores

1. Iniciar el script ssh como se muestra en la siguiente figura:

# Figura 27. Inicio script ssh



Fuente: Autores

1. Crear la variable HADOOP\_HOME en el path como se muestra en la figura N°28, a través de los siguientes comandos:

$ export HADOOP\_HOME=/Documentos/hadoop/bin (dirección de descarga).

$ export HADOOP\_IN\_PATH=/Documentos/hadoop/bin.

# Figura 28. Configuración PATH hadoop

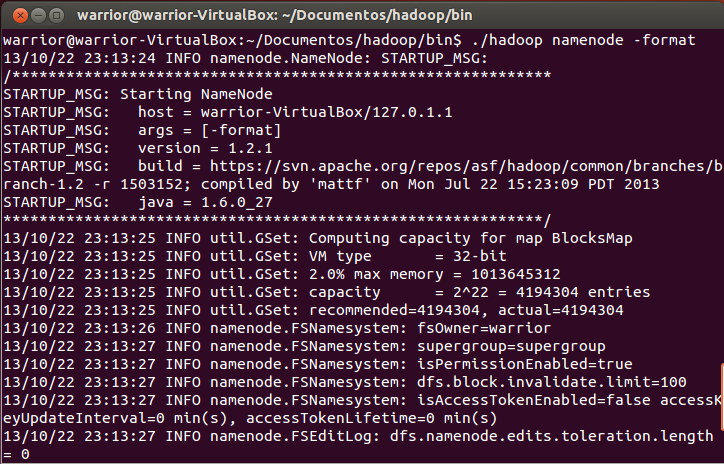


Fuente: Autores

1. Dar formato al namenode, como se muestra en la figura N°29, desde la ruta de la carpeta bin a través del siguiente comando:

$./hadoop namenode –format

# Figura 29. Formato del namenode

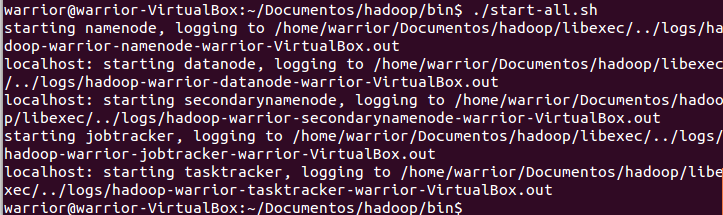


Fuente: Autores

1. Iniciar los daemons de hadoop, ejecutando el archivo start-all.sh que se encuentra en la carpeta bin, a través del siguiente comando: $ ./start-all.sh.

Se debe ver la información que aparece en la siguiente figura:

# Figura 30. Iniciar hadoop



Fuente: Autores

1. Comprobar los servicios que están arriba a través del comando jps, debe mostrar algo como lo siguiente:

14799 NameNode

14977 SecondaryNameNode

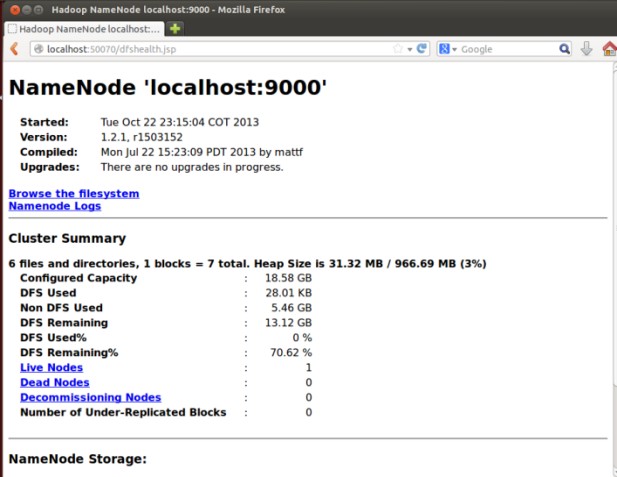
15183 DataNode

15596 JobTracker

15897 TaskTracker

1. Comprobar la conexión a través de http://localhost:50070, arrojara la siguiente imagen.

# Figura 31. Interfaz web namenode



Fuente: Autores

Comprobar la conexión de job tracker a través de http;//localhost:50030, debe mostrar la siguiente imagen