

CONFIGURACIÓN DE **HADOOP**

— EN MODO STANDALONE —
Y PSEUDO-CLÚSTER



**STANDALONE
MODE**



PSEUDO-CLÚSTER

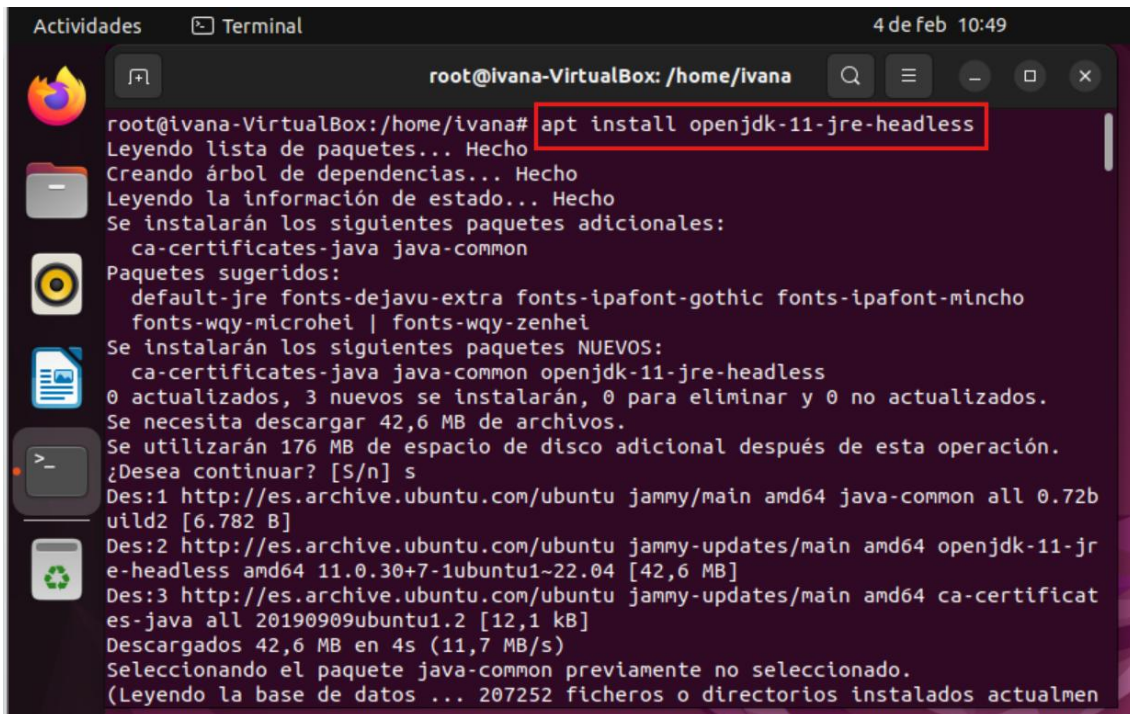


Ivana Sánchez Pérez

CONFIGURACIÓN DE HADOOP EN MODO STANDALONE Y PSEUDO-CLÚSTER

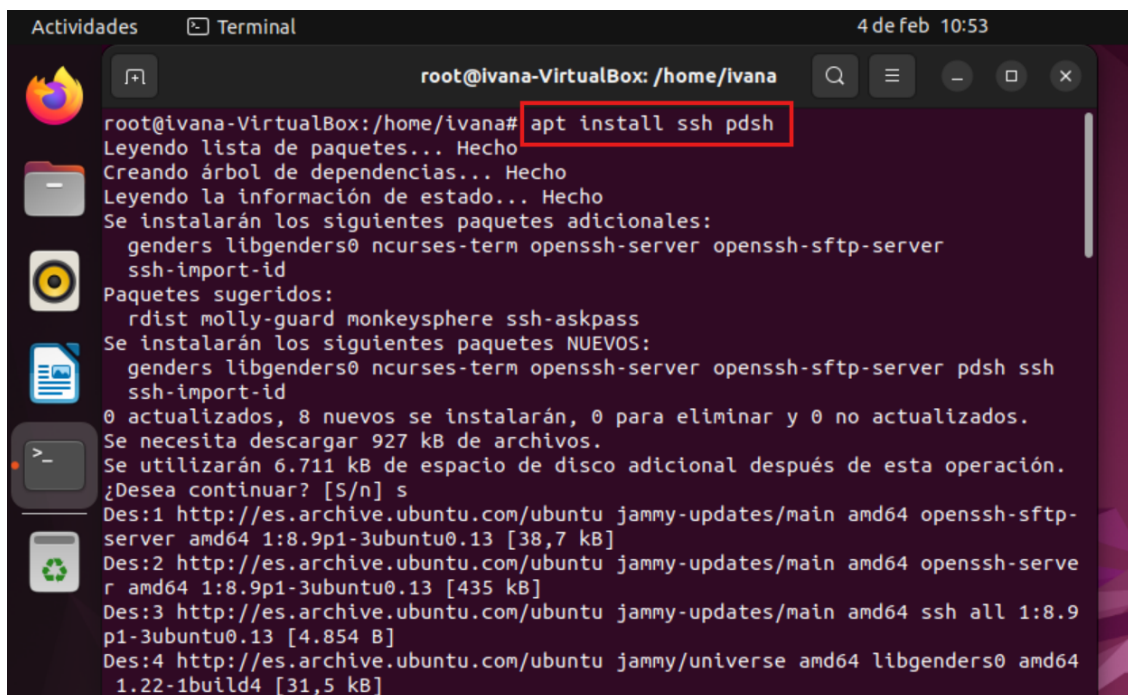
CONFIGURACIÓN DEL ENTORNO (MODO STANDALONE)

1- Instalación de Java 11, SSH y PDSH



A terminal window titled 'root@ivana-VirtualBox: /home/ivana' showing the command 'apt install openjdk-11-jre-headless' being executed. The output shows the package list being read, dependencies being created, and the installation of 'openjdk-11-jre-headless' along with 'ca-certificates-java' and 'java-common'. The terminal also shows the download progress and the selection of the package.

```
root@ivana-VirtualBox: /home/ivana# apt install openjdk-11-jre-headless
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se instalarán los siguientes paquetes adicionales:
  ca-certificates-java java-common
Paquetes sugeridos:
  default-jre fonts-dejavu-extra fonts-ipafont-gothic fonts-ipafont-mincho
  fonts-wqy-microhei | fonts-wqy-zenhei
Se instalarán los siguientes paquetes NUEVOS:
  ca-certificates-java java-common openjdk-11-jre-headless
0 actualizados, 3 nuevos se instalarán, 0 para eliminar y 0 no actualizados.
Se necesita descargar 42,6 MB de archivos.
Se utilizarán 176 MB de espacio de disco adicional después de esta operación.
¿Desea continuar? [S/n] s
Des:1 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 java-common all 0.72b
  uild2 [6.782 B]
Des:2 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openjdk-11-jr
  e-headless amd64 11.0.30+7-1ubuntu1~22.04 [42,6 MB]
Des:3 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ca-certificat
  es-java all 20190909ubuntu1.2 [12,1 kB]
Descargados 42,6 MB en 4s (11,7 MB/s)
Seleccionando el paquete java-common previamente no seleccionado.
(Leyendo la base de datos ... 207252 ficheros o directorios instalados actualmen
```

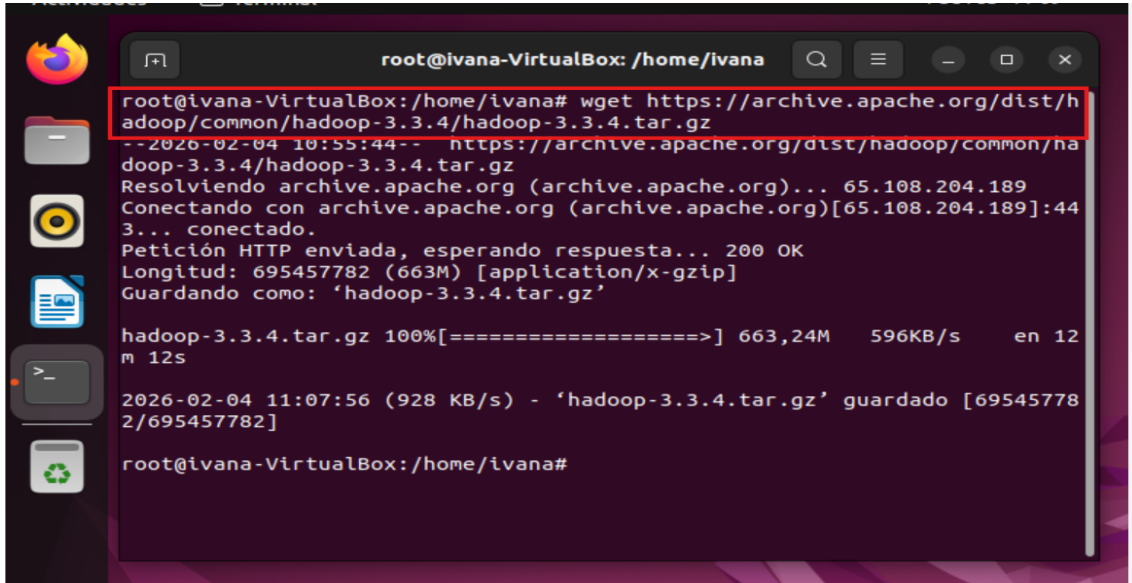


A terminal window titled 'root@ivana-VirtualBox: /home/ivana' showing the command 'apt install ssh pdsh' being executed. The output shows the package list being read, dependencies being created, and the installation of 'ssh' and 'pdsh' along with 'libgenders0', 'ncurses-term', 'openssh-server', and 'openssh-sftp-server'. The terminal also shows the download progress and the selection of the package.

```
root@ivana-VirtualBox: /home/ivana# apt install ssh pdsh
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se instalarán los siguientes paquetes adicionales:
  genders libgenders0 ncurses-term openssh-server openssh-sftp-server
  ssh-import-id
Paquetes sugeridos:
  rdist molly-guard monkeysphere ssh-askpass
Se instalarán los siguientes paquetes NUEVOS:
  genders libgenders0 ncurses-term openssh-server openssh-sftp-server pdsh ssh
  ssh-import-id
0 actualizados, 8 nuevos se instalarán, 0 para eliminar y 0 no actualizados.
Se necesita descargar 927 kB de archivos.
Se utilizarán 6.711 kB de espacio de disco adicional después de esta operación.
¿Desea continuar? [S/n] s
Des:1 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-sftp-
  server amd64 1:8.9p1-3ubuntu0.13 [38,7 kB]
Des:2 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-serve
  r amd64 1:8.9p1-3ubuntu0.13 [435 kB]
Des:3 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ssh all 1:8.9
  p1-3ubuntu0.13 [4.854 B]
Des:4 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 libgenders0 amd64
  1.22-1build4 [31,5 kB]
```

2- Instalación de Hadoop

Descarga y comprensión de la versión 3.3.4 en la ruta
/usr/share/hadoop.

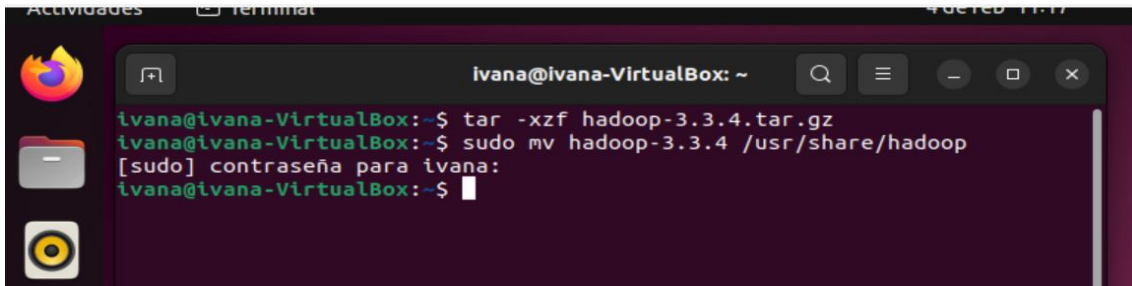


```
root@ivana-VirtualBox: /home/ivana
root@ivana-VirtualBox:/home/ivana# wget https://archive.apache.org/dist/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz
--2026-02-04 10:55:44-- https://archive.apache.org/dist/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz
Resolviendo archive.apache.org (archive.apache.org)... 65.108.204.189
Conectando con archive.apache.org (archive.apache.org)[65.108.204.189]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 695457782 (663M) [application/x-gzip]
Guardando como: 'hadoop-3.3.4.tar.gz'

hadoop-3.3.4.tar.gz 100%[=====] 663,24M 596KB/s en 12m 12s

2026-02-04 11:07:56 (928 KB/s) - 'hadoop-3.3.4.tar.gz' guardado [695457782/695457782]

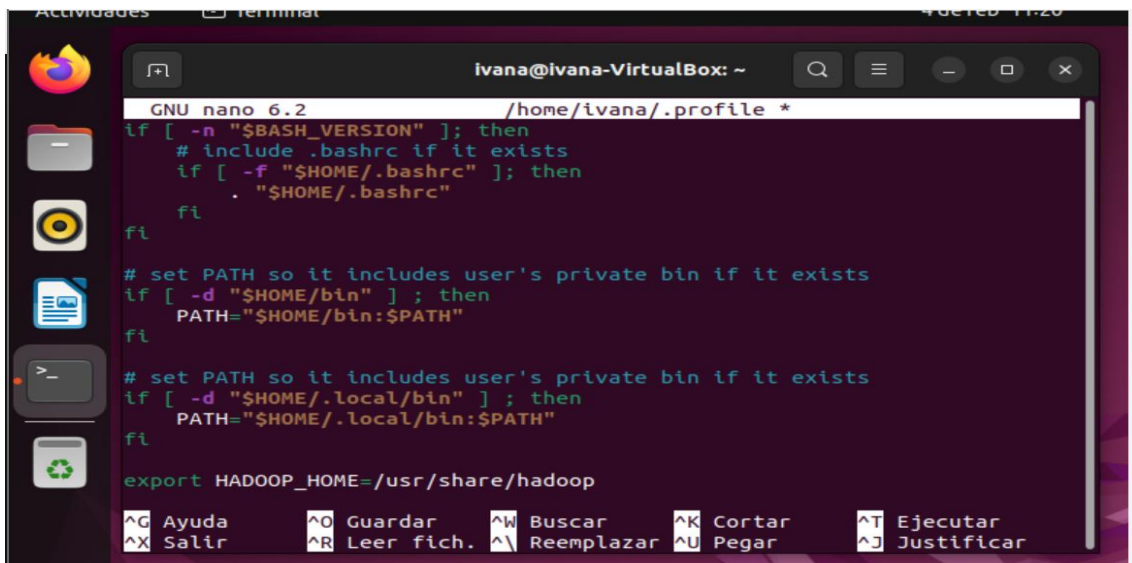
root@ivana-VirtualBox:/home/ivana#
```



```
ivana@ivana-VirtualBox: ~
ivana@ivana-VirtualBox:~$ tar -xzf hadoop-3.3.4.tar.gz
ivana@ivana-VirtualBox:~$ sudo mv hadoop-3.3.4 /usr/share/hadoop
[sudo] contraseña para ivana:
ivana@ivana-VirtualBox:~$
```

3- Variables del entorno

Configuración del archivo **.profile** para definir **HADOOP_HOME** y asegurar que el sistema reconozca los binarios.



```
GNU nano 6.2 /home/ivana/.profile *
if [ -n "$BASH_VERSION" ]; then
    # include .bashrc if it exists
    if [ -f "$HOME/.bashrc" ]; then
        . "$HOME/.bashrc"
    fi
fi

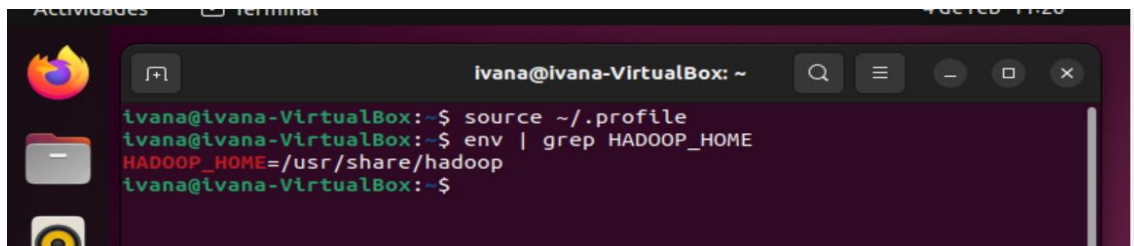
# set PATH so it includes user's private bin if it exists
if [ -d "$HOME/bin" ] ; then
    PATH="$HOME/bin:$PATH"
fi

# set PATH so it includes user's private bin if it exists
if [ -d "$HOME/.local/bin" ] ; then
    PATH="$HOME/.local/bin:$PATH"
fi

export HADOOP_HOME=/usr/share/hadoop

^G Ayuda      ^O Guardar    ^W Buscar     ^K Cortar     ^T Ejecutar
^X Salir      ^R Leer fich. ^P Reemplazar ^J Pegar      ^_ Justificar
```

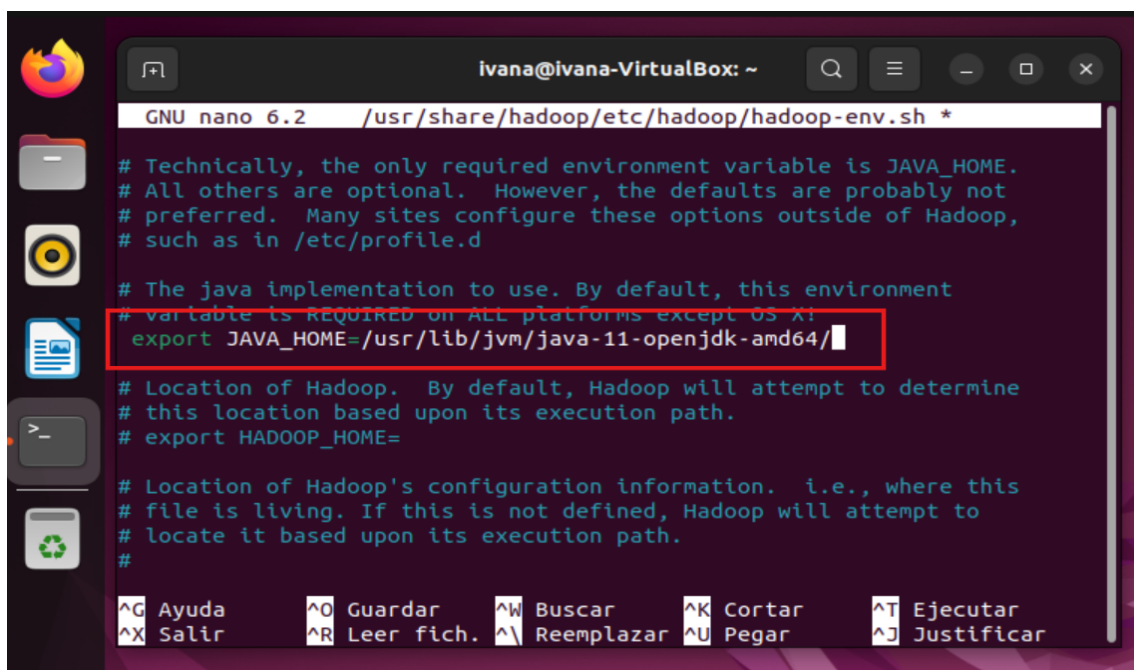
Cargamos los cambios y verificamos



```
ivana@ivana-VirtualBox: ~  
ivana@ivana-VirtualBox:~$ source ~/.profile  
ivana@ivana-VirtualBox:~$ env | grep HADOOP_HOME  
HADOOP_HOME=/usr/share/hadoop  
ivana@ivana-VirtualBox:~$
```

4- Configuración de Java en Hadoop

Hadoop necesita saber dónde está Java para arrancar sus servicios. Para ello definimos la ruta de Java dentro de `hadoop-env.sh`.



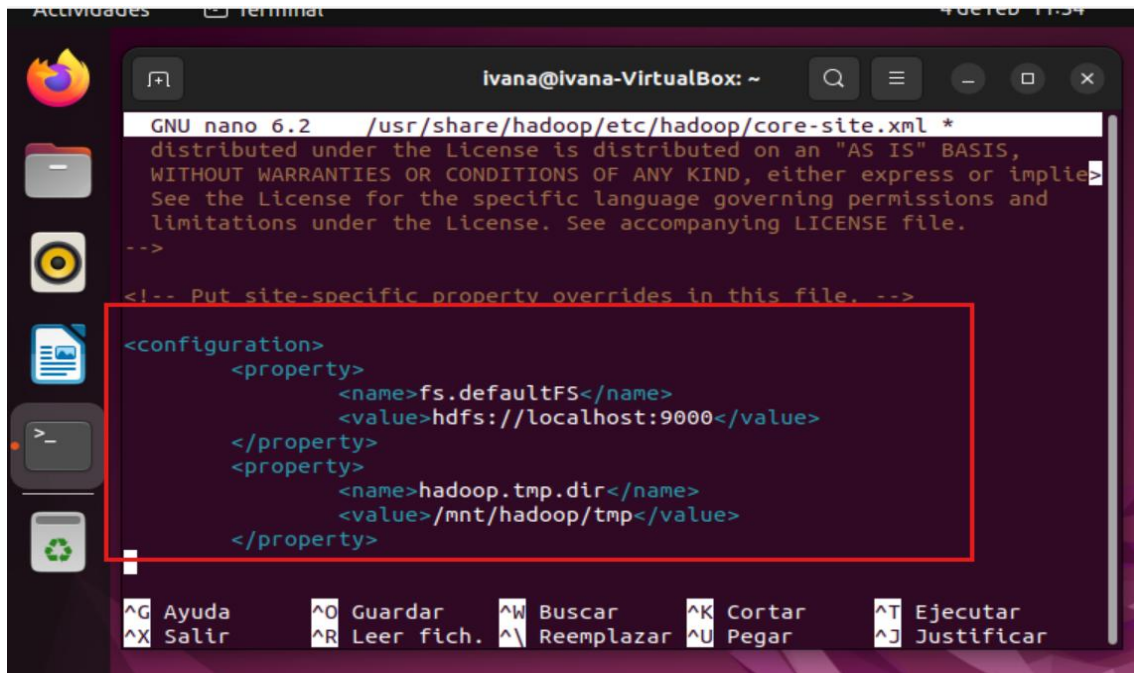
```
GNU nano 6.2 /usr/share/hadoop/etc/hadoop/hadoop-env.sh *  
  
# Technically, the only required environment variable is JAVA_HOME.  
# All others are optional. However, the defaults are probably not  
# preferred. Many sites configure these options outside of Hadoop,  
# such as in /etc/profile.d  
  
# The java implementation to use. By default, this environment  
# variable is REQUIRED on ALL platforms except OS X!  
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=  
  
# Location of Hadoop's configuration information. i.e., where this  
# file is living. If this is not defined, Hadoop will attempt to  
# locate it based upon its execution path.  
#  
  
^G Ayuda      ^O Guardar    ^W Buscar     ^K Cortar     ^T Ejecutar  
^X Salir      ^R Leer fich. ^\ Reemplazar  ^U Pegar      ^J Justificar
```

CONFIGURACIÓN EN MODO PSEUDO-CLÚSTER

Este modo simula un clúster en nuestra propia máquina.

1- Configuración de los archivos XML

Edición de ***core-site.xml*** y ***hdfs-site.xml*** para definir el sistema de archivos por defecto y factor de replicación.

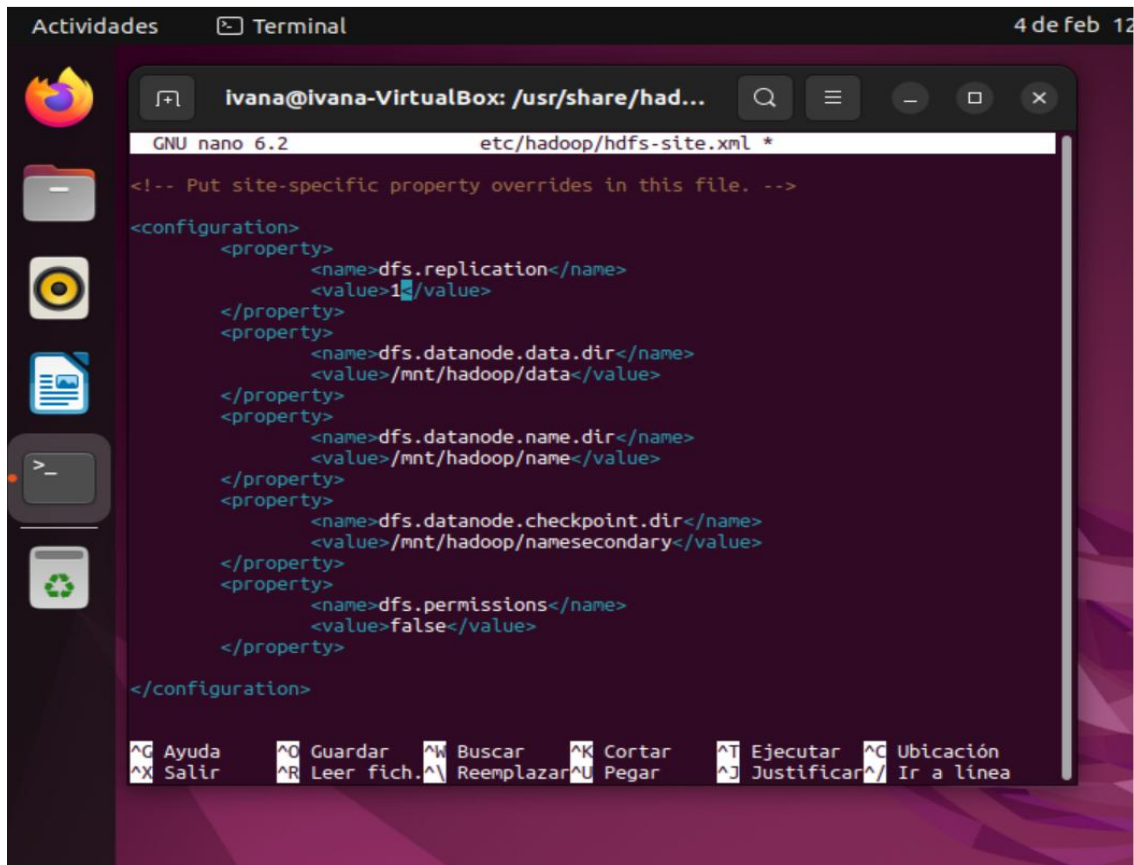


A screenshot of a terminal window titled 'ivana@ivana-VirtualBox: ~'. The terminal shows the GNU nano 6.2 editor editing the file '/usr/share/hadoop/etc/hadoop/core-site.xml'. The file content includes a license notice and a configuration section. The configuration section is highlighted with a red box and contains the following XML code:

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

<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/mnt/hadoop/tmp</value>
  </property>
</configuration>
```

At the bottom of the terminal, there is a row of keyboard shortcuts: ^G Ayuda, ^O Guardar, ^W Buscar, ^K Cortar, ^T Ejecutar, ^X Salir, ^R Leer fich., ^_ Reemplazar, ^U Pegar, ^J Justificar.



A screenshot of a terminal window titled 'ivana@ivana-VirtualBox: /usr/share/had...'. The terminal shows the GNU nano 6.2 editor editing the file 'etc/hadoop/hdfs-site.xml'. The file content includes a license notice and a configuration section. The configuration section contains the following XML code:

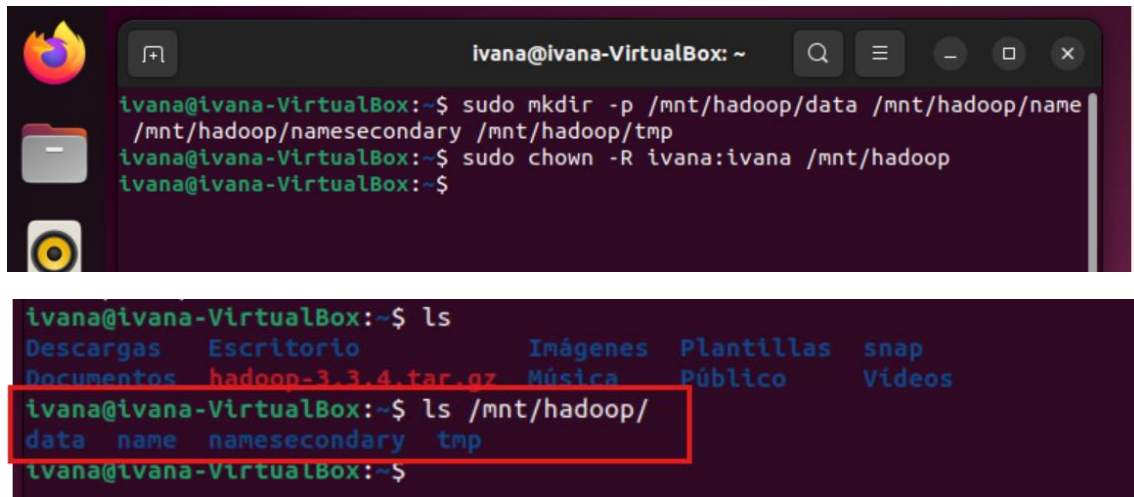
```
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/mnt/hadoop/data</value>
  </property>
  <property>
    <name>dfs.datanode.name.dir</name>
    <value>/mnt/hadoop/name</value>
  </property>
  <property>
    <name>dfs.datanode.checkpoint.dir</name>
    <value>/mnt/hadoop/namesecondary</value>
  </property>
  <property>
    <name>dfs.permissions</name>
    <value>>false</value>
  </property>
</configuration>
```

At the bottom of the terminal, there is a row of keyboard shortcuts: ^G Ayuda, ^O Guardar, ^W Buscar, ^K Cortar, ^T Ejecutar, ^C Ubicación, ^X Salir, ^R Leer fich., ^_ Reemplazar, ^U Pegar, ^J Justificar, ^_ Ir a línea.

2- Preparación de directorios

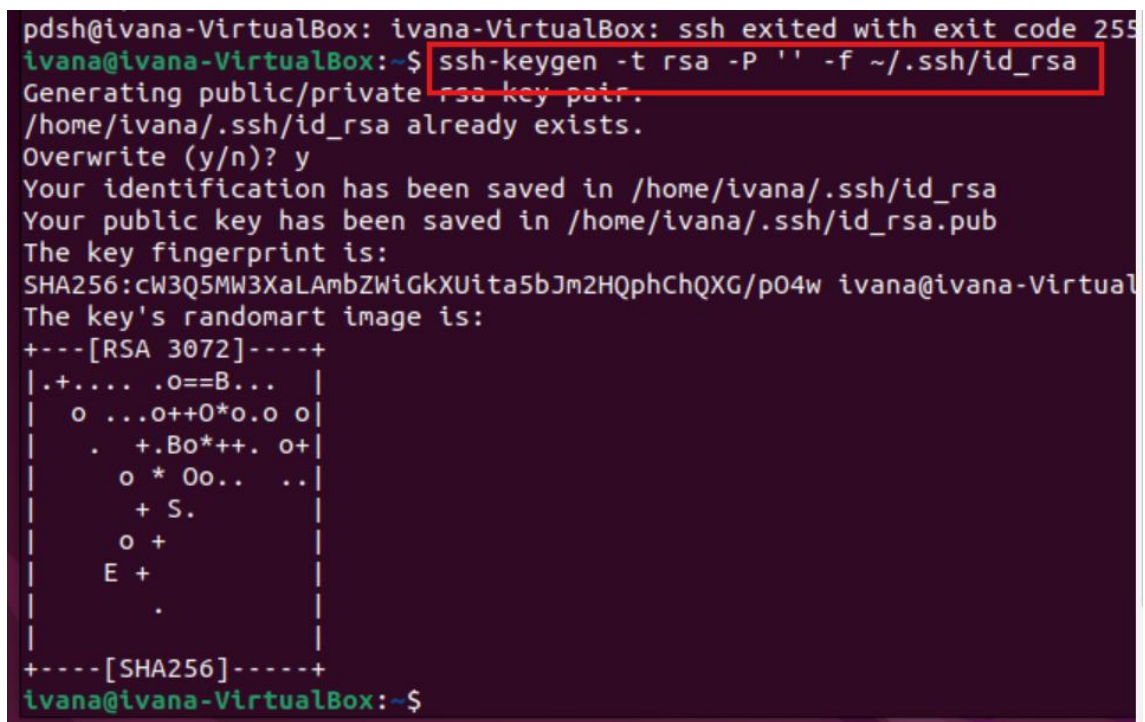
Creación de las carpetas físicas para datos, nombres y temporales en **/mnt/hadoop**, asignando permisos al usuario actual.



```
ivana@ivana-VirtualBox: ~  
ivana@ivana-VirtualBox:~$ sudo mkdir -p /mnt/hadoop/data /mnt/hadoop/name  
/mnt/hadoop/namesecondary /mnt/hadoop/tmp  
ivana@ivana-VirtualBox:~$ sudo chown -R ivana:ivana /mnt/hadoop  
ivana@ivana-VirtualBox:~$  
  
ivana@ivana-VirtualBox:~$ ls  
Descargas  Escritorio  Imágenes  Plantillas  snap  
Documentos hadoop-3.3.4.tar.gz Música  Público  Videos  
ivana@ivana-VirtualBox:~$ ls /mnt/hadoop/  
data  name  namesecondary  tmp  
ivana@ivana-VirtualBox:~$
```

3- Acceso a SSH sin clave

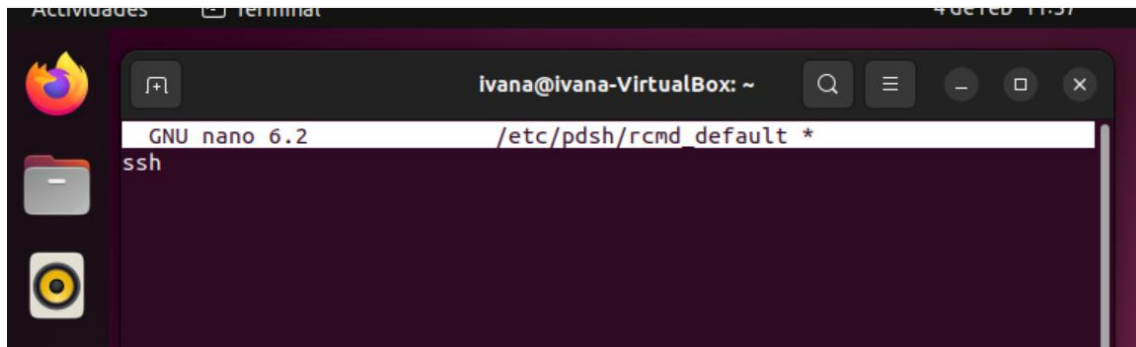
Generación de llaves RSA para permitir que Hadoop se comunice internamente consigo mismo, sin solicitar contraseñas.



```
pdsh@ivana-VirtualBox: ivana-VirtualBox: ssh exited with exit code 255  
ivana@ivana-VirtualBox:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa  
Generating public/private rsa key pair.  
/home/ivana/.ssh/id_rsa already exists.  
Overwrite (y/n)? y  
Your identification has been saved in /home/ivana/.ssh/id_rsa  
Your public key has been saved in /home/ivana/.ssh/id_rsa.pub  
The key fingerprint is:  
SHA256:cW3Q5MW3XaLambZWlGkXUita5bJm2HQphChQXG/p04w ivana@ivana-VirtualBox  
The key's randomart image is:  
+---[RSA 3072]-----+  
|. +.... .o==B... |  
| o ...o++0*o.o o |  
| . +.Bo*++. o+ |  
| o * 0o.. .. |  
| + S. |  
| o + |  
| E + |  
| . |  
+-----[SHA256]-----+  
ivana@ivana-VirtualBox:~$
```

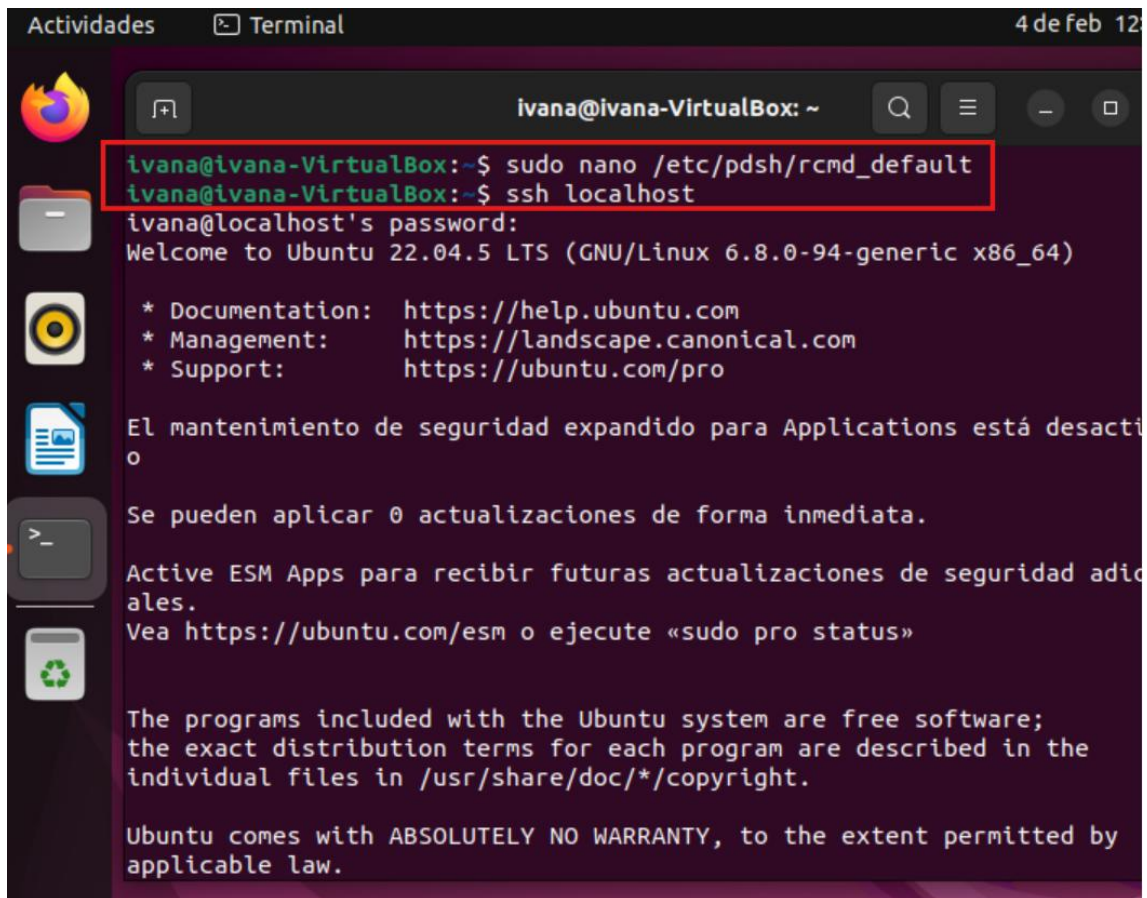
4- Configuramos el Shell remoto por defecto para PDSH

Obligatorio en Ubuntu para que Hadoop pueda arrancar los servicios sin errores. Creamos el archivo de configuración **rcmd_default**



```
ivana@ivana-VirtualBox: ~  
GNU nano 6.2 /etc/pdsh/rcmd_default *  
ssh
```

Verificamos la conexión SSH

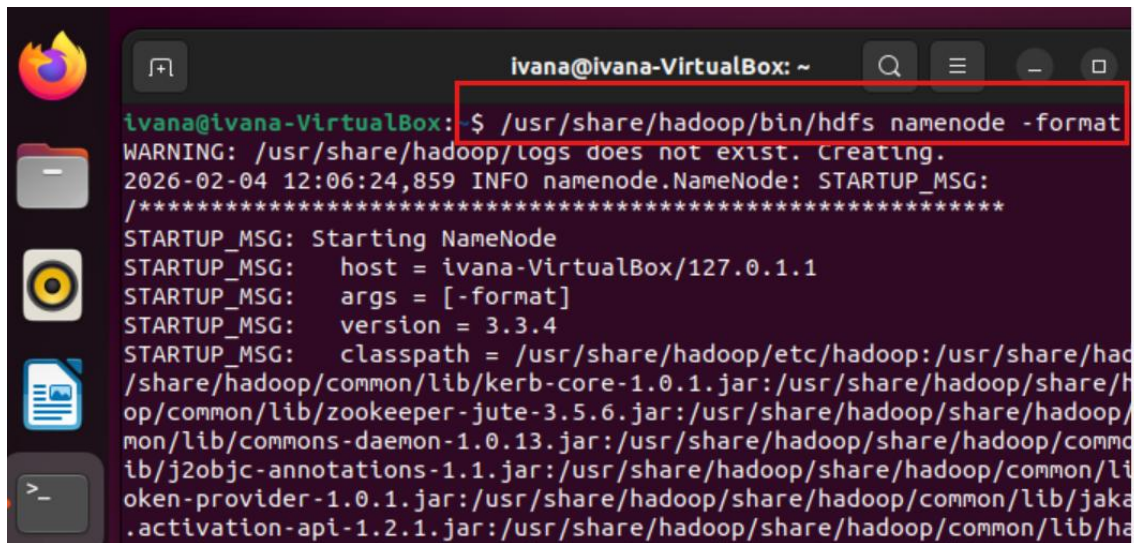


```
ivana@ivana-VirtualBox:~$ sudo nano /etc/pdsh/rcmd_default  
ivana@ivana-VirtualBox:~$ ssh localhost  
ivana@localhost's password:  
Welcome to Ubuntu 22.04.5 LTS (GNU/Linux 6.8.0-94-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/pro  
  
El mantenimiento de seguridad expandido para Applications está desacti  
o  
  
Se pueden aplicar 0 actualizaciones de forma inmediata.  
  
Active ESM Apps para recibir futuras actualizaciones de seguridad adic  
ales.  
Vea https://ubuntu.com/esm o ejecute «sudo pro status»  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.
```

5- Formateo y arranque

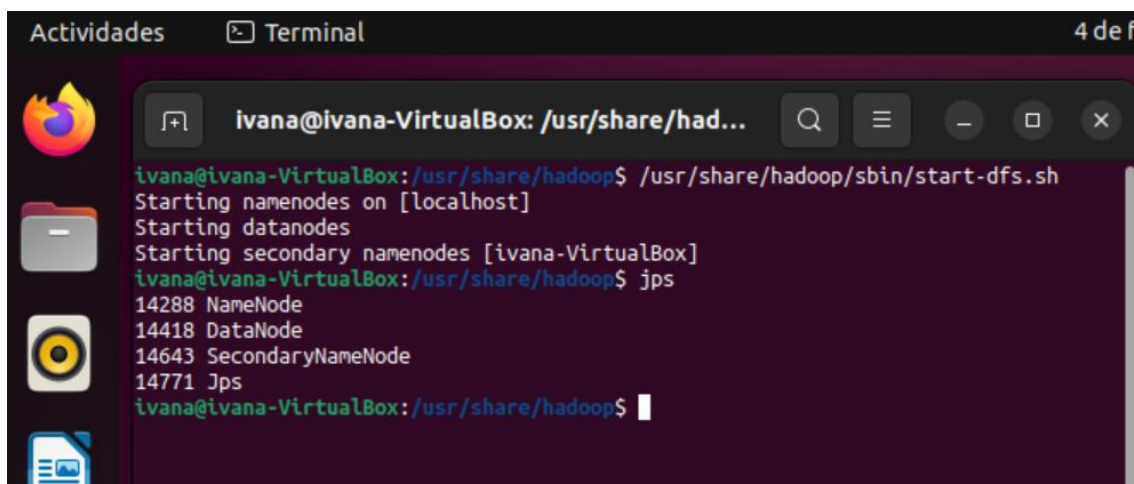
Formateo del NameNode y primer arranque de los servicios de almacenamiento mediante start-dfs.sh

a) Formatear el NameNode



```
ivana@ivana-VirtualBox: ~  
ivana@ivana-VirtualBox:~$ /usr/share/hadoop/bin/hdfs namenode -format  
WARNING: /usr/share/hadoop/logs does not exist. Creating.  
2026-02-04 12:06:24,859 INFO namenode.NameNode: STARTUP_MSG:  
/*****  
STARTUP_MSG: Starting NameNode  
STARTUP_MSG: host = ivana-VirtualBox/127.0.1.1  
STARTUP_MSG: args = [-format]  
STARTUP_MSG: version = 3.3.4  
STARTUP_MSG: classpath = /usr/share/hadoop/etc/hadoop:/usr/share/had  
/share/hadoop/common/lib/kerb-core-1.0.1.jar:/usr/share/hadoop/share/h  
op/common/lib/zookeeper-jute-3.5.6.jar:/usr/share/hadoop/share/hadoop/  
mon/lib/commons-daemon-1.0.13.jar:/usr/share/hadoop/share/hadoop/commo  
ib/j2objc-annotations-1.1.jar:/usr/share/hadoop/share/hadoop/common/li  
oken-provider-1.0.1.jar:/usr/share/hadoop/share/hadoop/common/lib/jaka  
.activation-api-1.2.1.jar:/usr/share/hadoop/share/hadoop/common/lib/ha
```

b) Arrancar los servicios de almacenamiento



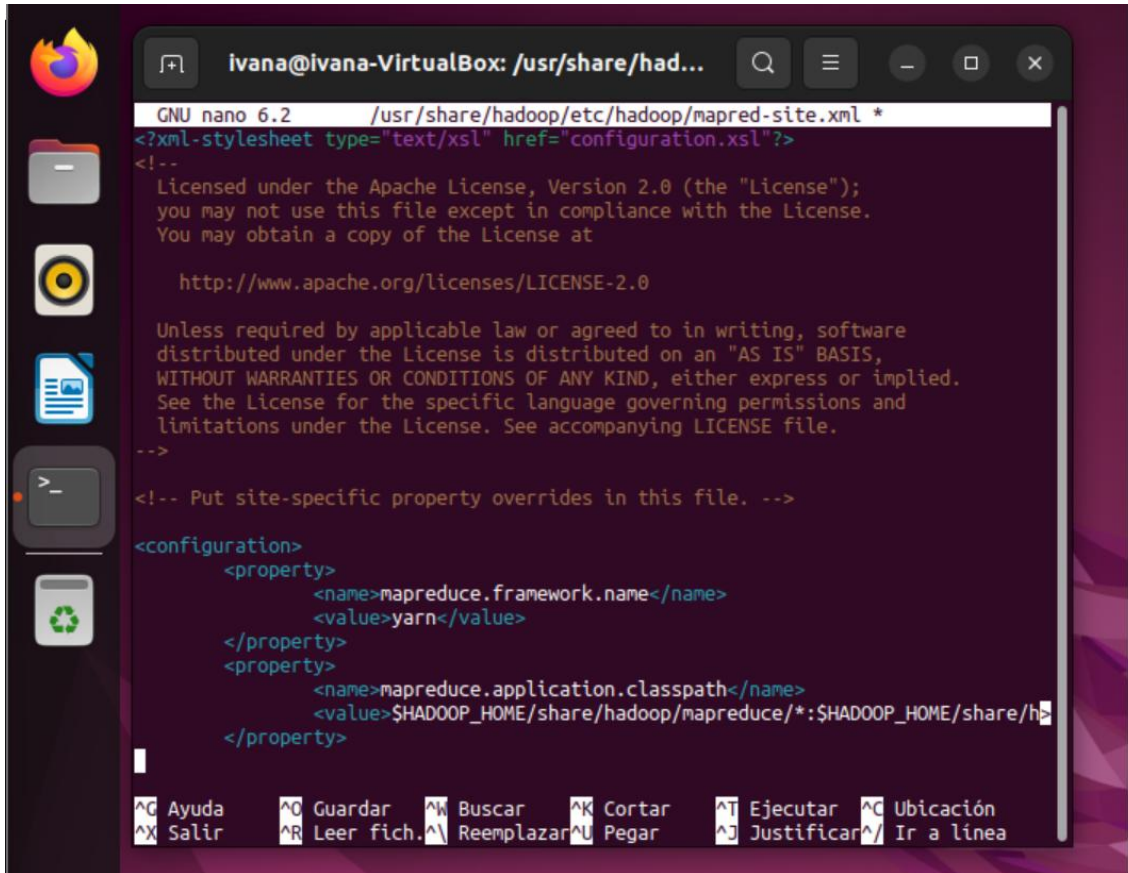
```
ivana@ivana-VirtualBox: /usr/share/had...  
ivana@ivana-VirtualBox:/usr/share/hadoop$ /usr/share/hadoop/sbin/start-dfs.sh  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [ivana-VirtualBox]  
ivana@ivana-VirtualBox:/usr/share/hadoop$ jps  
14288 NameNode  
14418 DataNode  
14643 SecondaryNameNode  
14771 Jps  
ivana@ivana-VirtualBox:/usr/share/hadoop$
```

CONFIGURACIÓN DE YARN Y RESOLUCIÓN DE ERRORES

Necesario para gestionar los recursos del sistema.

1- Activación de YARN

Configuración de mapred.site.xml y arranque del gestor de recursos para ejecutar tareas MapReduce



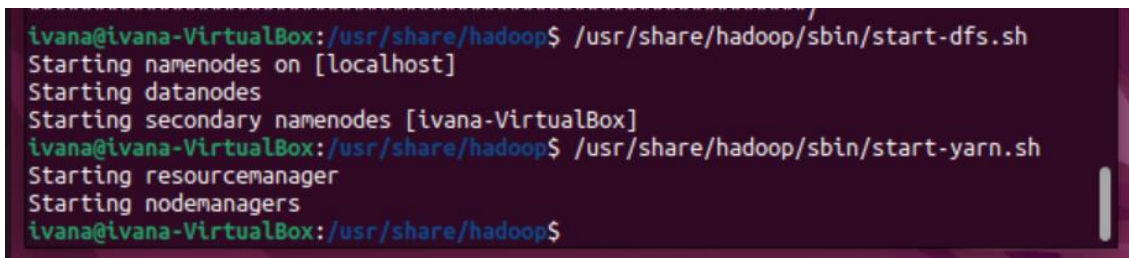
```
ivana@ivana-VirtualBox: /usr/share/had...
GNU nano 6.2 /usr/share/hadoop/etc/hadoop/mapred-site.xml *
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
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>
  <property>
    <name>mapreduce.application.classpath</name>
    <value>$HADOOP_HOME/share/hadoop/mapreduce/*:$HADOOP_HOME/share/h...
```

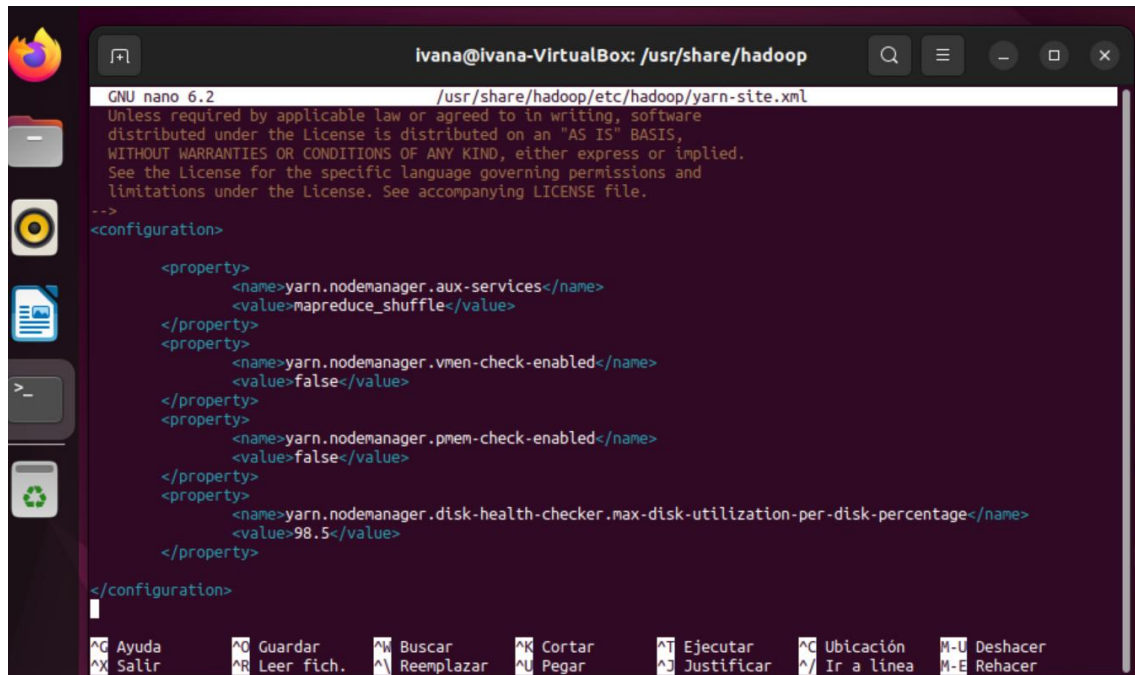


```
ivana@ivana-VirtualBox: /usr/share/hadoop$ /usr/share/hadoop/sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ivana-VirtualBox]
ivana@ivana-VirtualBox: /usr/share/hadoop$ /usr/share/hadoop/sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
ivana@ivana-VirtualBox: /usr/share/hadoop$
```

2- Depuración de errores

Debido al entorno de ejecución en una VM con recursos limitados, la tarea de WordCount fallaba inicialmente al superar los límites de memoria virtual permitidos por el sistema.

Para solucionar este inconveniente se ha procedido a modificar el archivo de configuración yarn-site.xml, para que desactivar la comprobación de la memorias tanto virtual como física y se ajustó al 98,5% para permitir que Hadoop siga trabajando incluso si el disco de la VM está casi lleno

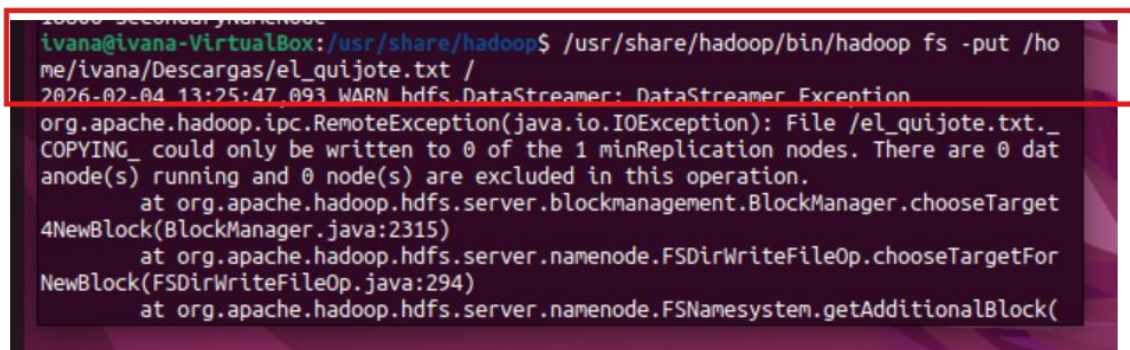


```
GNU nano 6.2 /usr/share/hadoop/etc/hadoop/yarn-site.xml
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.vmem-check-enabled</name>
    <value>>false</value>
  </property>
  <property>
    <name>yarn.nodemanager.pmem-check-enabled</name>
    <value>>false</value>
  </property>
  <property>
    <name>yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage</name>
    <value>98.5</value>
  </property>
</configuration>
```

EJECUCIÓN DE LA TAREA Y RESULTADOS FINALES

1- Carga de Datos

Subida del archivo el_quijote.txt (texto plano) al sistema HDFS



```
ivana@ivana-VirtualBox: /usr/share/hadoop$ /usr/share/hadoop/bin/hadoop fs -put /home/ivana/Descargas/el_quijote.txt /
2026-02-04 13:25:47.093 WARN hdfs.DataStreamer: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /el_quijote.txt._COPYING_ could only be written to 0 of the 1 minReplication nodes. There are 0 datanode(s) running and 0 node(s) are excluded in this operation.
    at org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarget4NewBlock(BlockManager.java:2315)
    at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.chooseTargetForNewBlock(FSDirWriteFileOp.java:294)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(
```

2- Procesamiento WordCount

Ejecución del programa de ejemplo para contar palabras, alcanzando el estado de SUCCEEDED tras las correcciones de memoria.


```

ivana@ivana-VirtualBox: /usr/share/had...
ivana@ivana-VirtualBox:/usr/share/hadoop$ /usr/share/hadoop/bin/hadoop jar /usr/sh
are/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar wordcount /e
l_quijote.txt /el_quijote_resultado
2026-02-04 14:22:14,531 INFO client.DefaultHARMPFailoverProxyProvider: Connecting
to ResourceManager at /0.0.0.0:8032
2026-02-04 14:22:15,240 INFO mapreduce.JobResourceUploader: Disabling Erasure Cod
ing for path: /tmp/hadoop-yarn/staging/ivana/.staging/job_1770211261981_0001
2026-02-04 14:22:16,406 INFO input.FileInputFormat: Total input files to process :
1
2026-02-04 14:22:17,056 INFO mapreduce.JobSubmitter: number of splits:1
2026-02-04 14:22:17,439 INFO mapreduce.JobSubmitter: Submitting tokens for job: jo
b_1770211261981_0001
2026-02-04 14:22:17,439 INFO mapreduce.JobSubmitter: Executing with tokens: []
2026-02-04 14:22:17,696 INFO conf.Configuration: resource-types.xml not found
2026-02-04 14:22:17,697 INFO resource.ResourceUtils: Unable to find 'resource-type
s.xml'.
2026-02-04 14:22:18,458 INFO impl.YarnClientImpl: Submitted application applicatio
n_1770211261981_0001
2026-02-04 14:22:18,572 INFO mapreduce.Job: The url to track the job: http://ivana
-VirtualBox:8088/proxy/application_1770211261981_0001/
2026-02-04 14:22:18,596 INFO mapreduce.Job: Running job: job_1770211261981_0001
2026-02-04 14:22:31,083 INFO mapreduce.Job: Job job_1770211261981_0001 running in
uber mode : false
2026-02-04 14:22:31,085 INFO mapreduce.Job: map 0% reduce 0%
2026-02-04 14:22:46,218 INFO mapreduce.Job: map 100% reduce 0%
2026-02-04 14:22:53,469 INFO mapreduce.Job: map 100% reduce 100%
2026-02-04 14:22:55,504 INFO mapreduce.Job: Job job_1770211261981_0001 completed s
uccessfully
2026-02-04 14:22:55,651 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=347063

```

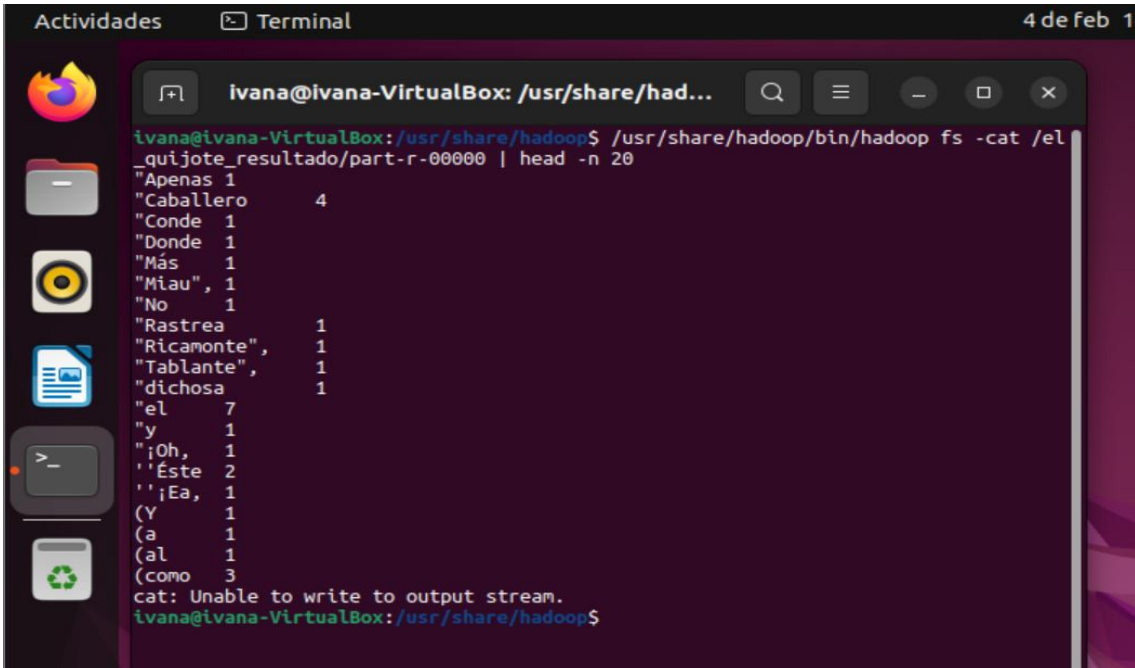
```

Combine input records=187018
Combine output records=22938
Reduce input groups=22938
Reduce shuffle bytes=347063
Reduce input records=22938
Reduce output records=22938
Spilled Records=45876
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=159
CPU time spent (ms)=9080
Physical memory (bytes) snapshot=555212800
Virtual memory (bytes) snapshot=5475381248
Total committed heap usage (bytes)=416284672
Peak Map Physical memory (bytes)=328871936
Peak Map Virtual memory (bytes)=2733268992
Peak Reduce Physical memory (bytes)=226340864
Peak Reduce Virtual memory (bytes)=2742112256
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=1060259
File Output Format Counters
Bytes Written=257233
ivana@ivana-VirtualBox:/usr/share/hadoop$

```

3- Verificaciones

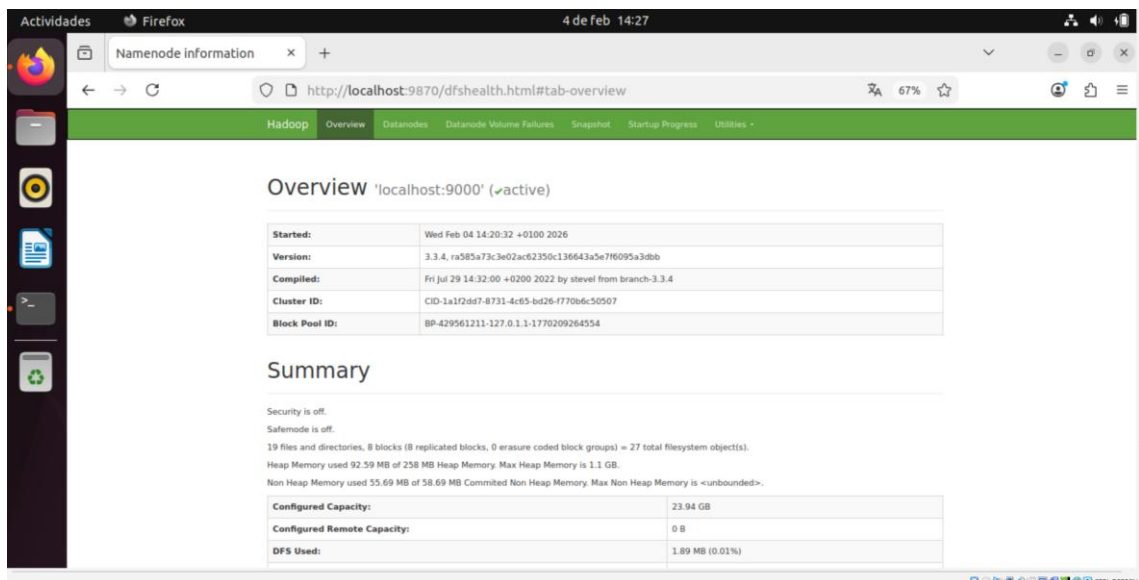
a. **Terminal:** Visualización de las primeras 20 palabras procesadas.



The screenshot shows a terminal window titled "ivana@ivana-VirtualBox: /usr/share/had...". The command executed is `/usr/share/hadoop/bin/hadoop fs -cat /el_quijote_resultado/part-r-00000 | head -n 20`. The output lists words and their counts, such as "Apenas 1", "Caballero 4", "Conde 1", etc. The terminal also shows an error message: "cat: Unable to write to output stream."

```
ivana@ivana-VirtualBox: /usr/share/hadoop$ /usr/share/hadoop/bin/hadoop fs -cat /el_quijote_resultado/part-r-00000 | head -n 20
"Apenas 1
"Caballero 4
"Conde 1
"Donde 1
"Más 1
"Miau", 1
"No 1
"Rastrea 1
"Ricamonte", 1
"Tablante", 1
"dichosa 1
"el 7
"y 1
"¡Oh, 1
"¡Este 2
"¡Ea, 1
(Y 1
(a 1
(al 1
(como 3
cat: Unable to write to output stream.
ivana@ivana-VirtualBox: /usr/share/hadoop$
```

b. **Interfaz Web:** Comprobación en <https://localhost:9870> del estado Active del NameNode y la presencia de los datos en el clúster.



DFS Used:	1.89 MB (0.01%)
Non DFS Used:	15.13 GB
DFS Remaining:	7.57 GB (31.63%)
Block Pool Used:	1.89 MB (0.01%)
DataNodes usages% (Min/Median/Max/stdDev):	0.01% / 0.01% / 0.01% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Wed Feb 04 14:20:32 +0100 2026
Last Checkpoint Time	Wed Feb 04 13:47:45 +0100 2026
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 161

Journal Manager	State
FileJournalManager(root=/mnt/hadoop/tmp/dfs/name)	EditLogFileOutputStream(/mnt/hadoop/tmp/dfs/name/current/edits_inprogress_00000000000000000085)

NameNode Storage

Storage Directory	Type	State
/mnt/hadoop/tmp/dfs/name	IMAGE_AND_EDITS	Active

En la imagen se observan los siguientes puntos clave que demuestran que el sistema está bien configurado:

- **Estado Activo:** El NameNode aparece como 'localhost:9000' (active), lo que indica que el formateo y el arranque fueron exitosos.
- **Seguridad y Modo Seguro:** Ambos parámetros, Security is off y Safemode is off, están desactivados, tal como se pedía en la guía para facilitar las pruebas.
- **Nodos Vivos:** Se muestra Live Nodes: 1, confirmando que el DataNode está conectado y listo para recibir datos.
- **Uso de Disco:** El sistema reconoce la capacidad total y muestra que hay archivos almacenados (1.89 MB de uso de DFS), que corresponden al Quijote y sus resultados.

ÚLTIMO PASO: APAGADO CORRECTO

Para asegurar que no se dan los archivos de datos que acabamos de ver en el navegador, ejecutaremos el comando de apagado antes de cerrar nuestra VM.

