

GUÍA DE USUARIO PARA INSTALAR HADOOP EN LINUX (FEDORA)



ESTUDIANTES:
BRAYAN DAVID AHUMADA CASTAÑEDA
JUANA VALENTINA MENDOZA SANTAMARÍA
LUIS FELIPE NÁRVAEZ GÓMEZ

PRESENTADO A:
INGENIERO JULIÁN ALBERTO MONSALVE PULIDO

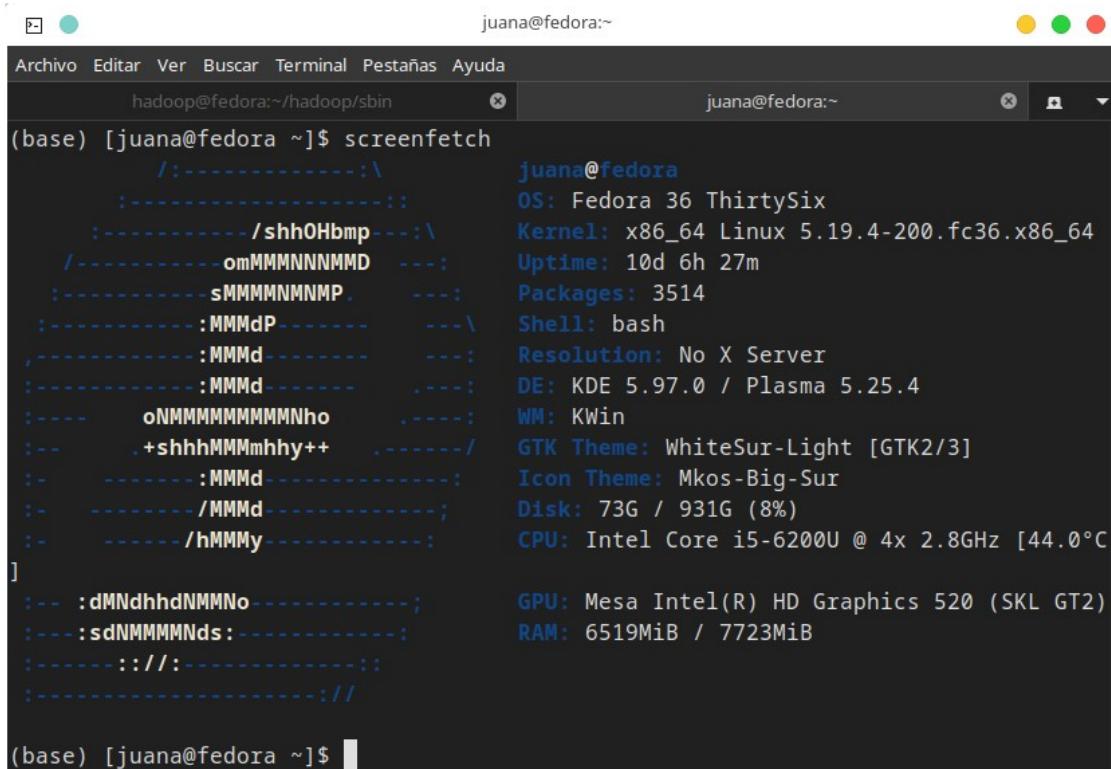
ASIGNATURA BIG DATA
CARRERA DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD SANTO TOMÁS
SEPTIEMBRE 12 DEL 2022
TUNJA, BOYACÁ



Guía de usuario

A continuación, se presenta una guía de usuario para realizar la instalación de Hadoop en la distribución Fedora para Linux. El proceso de instalación está basado del vídeo: Hadoop Installation| Fedora | Hadoop 3.2.2 | Easy steps |.

Información del sistema:



```
juana@fedora:~
juana@fedora:~
```

(base) [juana@fedora ~]\$ screenfetch

```
          :-----:\
          :-----:::
          :-----/shh0Hbmp---:\ \
          :-----omMMNNNNMMD   ---:
          :-----sMMMMNMNMMP,   ---:
          :-----:MMMdP-----\ \
          ,:-----:MMMd-----:
          :-----:MMMd-----:
          :-----oNMMMMMMMMMNho  .----:
          .+shhhMMMmhhy++  .----/
          :-----:MMMd-----:
          :-----/MMMd-----;
          :-----/hMMMy-----:
          ]
          :dMNdhhdNMNNo-----;
          :sdNMMMMNdS-----:
          :-----::/:-----:::
          :-----:://-----:::

(base) [juana@fedora ~]$
```

juana@fedora
OS: Fedora 36 ThirtySix
Kernel: x86_64 Linux 5.19.4-200.fc36.x86_64
Uptime: 10d 6h 27m
Packages: 3514
Shell: bash
Resolution: No X Server
DE: KDE 5.97.0 / Plasma 5.25.4
WM: KWin
GTK Theme: WhiteSur-Light [GTK2/3]
Icon Theme: Mkos-Big-Sur
Disk: 73G / 931G (8%)
CPU: Intel Core i5-6200U @ 4x 2.8GHz [44.0°C]
GPU: Mesa Intel(R) HD Graphics 520 (SKL GT2)
RAM: 6519MiB / 7723MiB

Nota: el comando screenfetch permite mostrar la información general del sistema operativo.



Desarrollo:

1. Acceder al Super Usuario.

→ **su**

```
juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Ayuda
(base) [juana@fedora ~]$ su
Contraseña:
[root@fedora juana]#
```

2. Verificar la instalación de Java.

→ **dnf install -y java**

```
juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Ayuda
[root@fedora juana]# dnf install -y java
Última comprobación de caducidad de metadatos hecha hace 4:28:14, el sáb 10 sep
2022 13:19:04.
El paquete java-17-openjdk-1:17.0.4.0.8-1.fc36.x86_64 ya está instalado.
Dependencias resueltas.
Nada por hacer.
¡Listo!
[root@fedora juana]#
```



3. A continuación, se crea el usuario para Hadoop.

→ **useradd hadoop**

```
juana@fedora:~$ useradd hadoop
[root@fedora juana]#
```

4. Se adiciona una contraseña al usuario creado para Hadoop.

→ **passwd hadoop**

```
juana@fedora:~$ passwd hadoop
[root@fedora juana]# passwd hadoop
Cambiando la contraseña del usuario hadoop.
Nueva contraseña:
CONTRASEÑA INCORRECTA: La contraseña no supera la verificación de diccionario - está basada en una palabra del diccionario
Vuelva a escribir la nueva contraseña:
passwd: todos los tokens de autenticación se actualizaron exitosamente.
[root@fedora juana]#
```

Nota: sale una advertencia, porque recomiendan que sea una contraseña fuerte.



5. Luego, se adicionan privilegios al usuario Hadoop. Para ello se matricula el usuario con permisos de sudo.

→ **visudo**

```
juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Ayuda
[root@fedora juana]# visudo
[root@fedora juana]#
```

```
juana@fedora:/home/juana — visudo
Archivo Editar Ver Buscar Terminal Ayuda
GNU nano 6.0 /etc/sudoers.tmp
## Sudoers allows particular users to run various commands as
## the root user, without needing the root password.
##
## Examples are provided at the bottom of the file for collections
## of related commands, which can then be delegated out to particular
## users or groups.
##
## This file must be edited with the 'visudo' command.

## Host Aliases
## Groups of machines. You may prefer to use hostnames (perhaps using
## wildcards for entire domains) or IP addresses instead.
# Host_Alias      FILESERVERS = fs1, fs2
# Host_Alias      MAILSERVERS = smtp, smtp2

## User Aliases
## These aren't often necessary, as you can use regular groups

^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
```

Nota: el comando visudo abre un editor de texto igual al normal, pero valida la sintaxis del archivo al guardarlo. Esto evita que los errores de configuración bloquen las operaciones sudo, que pueden ser la única forma de obtener privilegios root.



5. Posteriormente, se desplaza hasta el apartado con la lista de usuarios. Al usuario Hadoop se le van a asignar todos los permisos. Se agrega la línea:

→ **hadoop ALL=(ALL) ALL**

```
juana@fedora:/home/juana — visudo
Archivo Editar Ver Buscar Terminal Ayuda
GNU nano 6.0          /etc/sudoers.tmp          Modificado
root    ALL=(ALL)      ALL
hadoop  ALL=(ALL)      ALL
## Allows members of the 'sys' group to run networking, software,
## service management apps and more.
# %sys ALL = NETWORKING, SOFTWARE, SERVICES, STORAGE, DELEGATING, PROCESSES, LOGGING
## Allows people in group wheel to run all commands
%wheel  ALL=(ALL)      ALL
## Same thing without a password
# %wheel      ALL=(ALL)      NOPASSWD: ALL
## Allows members of the users group to mount and umount the
## cdrom as root
# %users  ALL=/sbin/mount /mnt/cdrom, /sbin/umount /mnt/cdrom
## Allows members of the users group to shutdown this system
^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
```

6. Se presiona las teclas **ctrl + o** para guardar, **enter** y **ctrl + x** para salir de Visudo.



7. Para confirmar la creación correcta del usuario Hadoop, se ingresa al shell como usuario Hadoop Hadoop. Posteriormente, se pregunta que usuario soy para confirmar que sea el usuario correspondiente.

→ **su - hadoop**

→ **whoami**

```
hadoop@fedora:~$ su - hadoop
Contraseña:
[hadoop@fedora ~]$ ls
[hadoop@fedora ~]$ pwd
/home/hadoop
[hadoop@fedora ~]$ whoami
hadoop
```

8. Luego, se genera una llave RSA para la autenticación. Como se puede observar en la terminal, se crean una llave privada y una llave pública.

→ **ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa**

```
hadoop@fedora:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hadoop/.ssh'.
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:VEg5litizrOkV0WJ1C8BKMNjacmHT+E6jCD+bKx0v0Q hadoop@fedora
The key's randomart image is:
+---[RSA 3072]---+
|o +oo ..+.
|oXoo. o *.
|=o=. . =.o
|* ..o =.o
|.= + + oS
|. = = o
| + B +
|+ = o
|.E .
+---[SHA256]---+
```

Nota: SSH (Secure Shell), es un protocolo que se utiliza para iniciar sesión de forma segura en sistemas remotos. Es la forma más común de acceder a servidores Linux remotos.



9. Para verificar que se hayan creado las llaves, se lista el contenido de ssh.

→ **ls .ssh/**

```
hadoop@fedora:~$ ls .ssh/
id_rsa  id_rsa.pub
[hadoop@fedora ~]$
```

10. A continuación, se copia la llave pública como una llave autorizada; esto para evitar modificar el contenido de la llave original (pública). Además, en el siguiente comando, se muestra la nueva llave creada.

→ **cp -r ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys**

→ **ls -l .ssh/**

```
hadoop@fedora:~$ cp -r ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
[hadoop@fedora ~]$ ls -l ~/.ssh
total 12
-rw-r--r--. 1 hadoop hadoop 567 sep 10 18:07 authorized_keys
-rw-----. 1 hadoop hadoop 2602 sep 10 18:03 id_rsa
-rw-r--r--. 1 hadoop hadoop 567 sep 10 18:03 id_rsa.pub
[hadoop@fedora ~]$
```



11. Ahora se otorgan los permisos de lectura y escritura, pero solo para el usuario Hadoop. A continuación, se observa que para la llave autorizada solo el usuario Hadoop tiene permisos de lectura y escritura.

→ **chmod 0600 ~/ssh/authorized_keys**

→ **ls -l .ssh/**

```
hadoop@fedora:~$ chmod 0600 ~/ssh/authorized_keys
[hadoop@fedora ~]$ ls -l
total 0
[hadoop@fedora ~]$ ls -l ~/ssh
total 12
-rw----- 1 hadoop hadoop 567 sep 10 18:07 authorized_keys
-rw----- 1 hadoop hadoop 2602 sep 10 18:03 id_rsa
-rw-r--r-- 1 hadoop hadoop 567 sep 10 18:03 id_rsa.pub
[hadoop@fedora ~]$
```

En la siguiente tabla se explica como Linux maneja los permisos. Las rayas que se ven en la parte izquierda representan lo siguiente:

Tipo de archivo	Usuario			Grupo			Others		
	read	write	execute	read	write	execute	read	write	execute
-									
0				6			0		

¿Por qué 6?

read		write		execute
2 ²		2 ¹		2 ⁰
4		2		0 (No se da este permiso)
		4 + 4 = 6		

El tipo de archivo en Linux puede ser:

Representación de tipo de archivo

-
b
c
d
l
n
p
s

Tipo de archivo

Archivo regular
Bloquear archivo especial
Archivo especial de caracteres
Directorio
Enlace simbólico
Archivo de red
FIFO
Socket



12. Luego, se habilita el protocolo de ssh (Secure Shell) como usuario root. En otra ventana de la terminal, se verifica el estado de ssh:

→ **systemctl status sshd.service**

The screenshot shows two terminal windows side-by-side. The left window is run by a regular user named juana, and the right window is run by the root user. Both windows are titled 'root@fedora:~'. In the left window, the command 'systemctl status sshd.service' is run, showing that the service is inactive (dead). In the right window, the command 'systemctl status sshd.service' is run again, but this time it shows that the service is active (running).

```
root@fedora:~  
[base] [juana@fedora ~]$ su -  
Contraseña:  
[root@fedora ~]# systemctl status sshd.service  
● sshd.service - OpenSSH server daemon  
  Loaded: loaded (/usr/lib/systemd/system/sshd.service; disabled; vendor pre>  
  Active: inactive (dead)  
    Docs: man:sshd(8)  
          man:sshd_config(5)  
lines 1-5/5 (END)...skipping...  
● sshd.service - OpenSSH server daemon  
  Loaded: loaded (/usr/lib/systemd/system/sshd.service; disabled; vendor pre>  
  Active: inactive (dead)  
    Docs: man:sshd(8)  
          man:sshd_config(5)
```

13. Ahora, se lanza el protocolo ssh:

→ **systemctl start sshd.service**

The screenshot shows two terminal windows side-by-side. The left window is run by a regular user named juana, and the right window is run by the root user. Both windows are titled 'root@fedora:~'. In the left window, the command 'systemctl start sshd.service' is run. In the right window, the command 'systemctl status sshd.service' is run, showing that the service is now active (running).

```
root@fedora:~  
[root@fedora ~]# systemctl start sshd.service  
[root@fedora ~]# systemctl status sshd.service  
● sshd.service - OpenSSH server daemon  
  Loaded: loaded (/usr/lib/systemd/system/sshd.service; disabled; vendor pre>  
  Active: active (running) since Sat 2022-09-10 18:18:13 -05; 6s ago  
    Docs: man:sshd(8)  
          man:sshd_config(5)  
  Main PID: 571429 (sshd)  
     Tasks: 1 (limit: 9218)  
    Memory: 2.3M  
      CPU: 18ms  
     CGroup: /system.slice/sshd.service  
lines 1-10...skipping...  
● sshd.service - OpenSSH server daemon  
  Loaded: loaded (/usr/lib/systemd/system/sshd.service; disabled; vendor pre>  
  Active: active (running) since Sat 2022-09-10 18:18:13 -05; 6s ago  
    Docs: man:sshd(8)  
          man:sshd_config(5)
```



14. Ahora, se activa permanentemente el protocolo ssh:

→ **systemctl enable sshd.service**

```
root@fedora:~  
Archivo Editar Ver Buscar Terminal Pestañas Ayuda  
hadoop@fedora:~  
[root@fedora ~]# systemctl enable sshd.service  
Created symlink /etc/systemd/system/multi-user.target.wants/sshd.service → /usr/  
lib/systemd/system/sshd.service.  
[root@fedora ~]#
```

15. A continuación, se ingresa al usuario Hadoop de forma segura, por medio del protocolo ssh.

→ **ssh localhost**

```
hadoop@fedora:~  
Archivo Editar Ver Buscar Terminal Pestañas Ayuda  
hadoop@fedora:~  
root@fedora:~  
[hadoop@fedora ~]$ ssh localhost  
The authenticity of host 'localhost (::1)' can't be established.  
ED25519 key fingerprint is SHA256:1vFOGU3/trx58Ewhu4iWSfKoF7rm0xvm8AoU+80p+Uo.  
This key is not known by any other names  
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes  
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.  
Last login: Sat Sep 10 18:01:41 2022  
[hadoop@fedora ~]$ clear
```

Nota: SSH (Secure Shell), es un protocolo que se utiliza para iniciar sesión de forma segura en sistemas remotos. Es la forma más común de acceder a servidores Linux remotos.



17. Posteriormente, en la otra consola se descarga el archivo de Hadoop.

→ **wget <https://archive.apache.org/dist/hadoop/core/hadoop-3.3.4/hadoop-3.3.4.tar.gz>**

```
[hadoop@fedora ~]$ wget https://archive.apache.org/dist/hadoop/core/hadoop-3.3.4/hadoop-3.3.4.tar.gz
--2022-09-10 18:25:37-- https://archive.apache.org/dist/hadoop/core/hadoop-3.3.4/hadoop-3.3.4.tar.gz
Resolviendo archive.apache.org (archive.apache.org)... 138.201.131.134, 2a01:4f8:172:2ec5::2
Conectando con archive.apache.org (archive.apache.org)[138.201.131.134]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 695457782 (663M) [application/x-gzip]
Grabando a: «hadoop-3.3.4.tar.gz»

hadoop-3.3.4.tar.gz 1%[          ] 10,23M 3,39MB/s eta 3m 12s
```

Nota: tar.gz es un archivo comprimido en Linux.

18. Se descomprime el archivo descargado y se muestra.

→ **tar xvf hadoop-3.3.4.tar.gz**

→ **ls**

```
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_error_sml.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_success_sml.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/expanded.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/external.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_info_sml.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logo_apache.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/bg.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/newwindow.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/h3.jpg
[hadoop@fedora ~]$ ls
hadoop-3.3.4  hadoop-3.3.4.tar.gz
```



19. A continuación, se renombra la carpeta y se muestra.

→ **mv hadoop-3.3.4 hadoop**

```
hadoop@fedora:~$ mv hadoop-3.3.4 hadoop
[hadoop@fedora ~]$ mv hadoop-3.3.4 hadoop
[hadoop@fedora ~]$ ls
```

20. Posteriormente, se verifica las alternativas de Java.

→ **sudo update-alternatives --config java**

```
hadoop@fedora:~$ sudo update-alternatives --config java
[sudo] password for hadoop:
Hay 2 programas que proporcionan 'java'.
  Selección     Comando
  -----
*+ 1           java-17-openjdk.x86_64 (/usr/lib/jvm/java-17-openjdk-17.0.4.0.8-1.fc36.x86_64/bin/java)
  2           java-1.8.0-openjdk.x86_64 (/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre/bin/java)

Presione Intro para mantener la selección actual[+], o escriba el número de la selección: [
```

Nota: se recomienda utilizar la versión de Java seleccionada. Si no se cuenta con la versión, en otra terminal realizar la instalación con el comando:

→ **sudo dnf install java-1.8.0-openjdk_x86_64**



```
juana@fedora:/home/juana — /usr/bin/python3 /usr/bin/dnf install java-1.8.0-openjdk.x86_64
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~ juana@fedora:/home/juana — /usr/bin/python3 /usr/bin/dnf install java-1.8.0-openjdk.x86_64
(base) [juana@fedora ~]$ su
Contraseña:
[root@fedora juana]# dnf install java-1.8.0-openjdk.x86_64
Última comprobación de caducidad de metadatos hecha hace 2:40:05, el sáb 10 sep 2022 19:03:08.
Dependencias resueltas.
=====
Paquete           Arq.      Versión          Repositorio   Tam.
=====
Instalando:
  java-1.8.0-openjdk        x86_64    1:1.8.0.345.b01-1.fc36  updates     273 k
Instalando dependencias:
  java-1.8.0-openjdk-headless x86_64    1:1.8.0.345.b01-1.fc36  updates     33 M

Resumen de la transacción
=====
Instalar 2 Paquetes

Tamaño total de la descarga: 33 M
Tamaño instalado: 119 M
```

Nota: *dnf* es el gestor de paquetes para la distribución Fedora. Posteriormente, correr el comando de alternativas de Java.

21. Copiar la ruta de Java 1.8.8 y pulsar **enter** para salir.

/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre/

```
hadoop@fedora:~ juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~ juana@fedora:/home/juana
[hadoop@fedora ~]$ sudo update-alternatives --config java
[sudo] password for hadoop:

Hay 2 programas que proporcionan 'java'.

  Selección  Comando
  -----
*+ 1          java-17-openjdk.x86_64 (/usr/lib/jvm/java-17-openjdk-17.0.4.0.8-1.fc36.x86_64/bin/java)
  2          java-1.8.0-openjdk.x86_64 (/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre/bin/java)

Presione Intro para mantener la selección actual[+], o escriba el número de la selección:
```



22. A continuación, se editar el archivo bashrc. Bashrc, es un archivo shell script que permite configurar el Bash cuando se inicia una Shell bash.

→ **nano ~/.bashrc**

nota: Nano es un editor de texto de línea de comandos simple, ligero, de código abierto desarrollado como parte del Proyecto GNU. Para instalar nano en otra ventana, se digita el comando: sudo dnf install nano.

23. En la última línea del archivo, se pega el siguiente código. Donde la ruta resaltada en negrilla es la anteriormente copiada hasta jre/.

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre/
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
hadoop@fedora:~$ nano .bashrc
hadoop@fedora:~$ /home/hadoop/.bashrc
Modificado

export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre/
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin

^G Ayuda      ^O Guardar      ^W Buscar      ^K Cortar      ^T Ejecutar      ^C Ubicación
```

Nota: reemplazar por su propia ruta de java.



24. Posteriormente, se ejecuta lo anteriormente realizado.

→ **source ~/.bashrc**

```
hadoop@fedora:~$ source ~/.bashrc
```

25. Luego, se edita el archivo de ambiente de Hadoop.

```
hadoop@fedora:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
```

26. En el editor buscar (ctrl + w) export JAVA_HOME e incluir la ruta de JAVA_HOME en el archivo hadoop-env.sh.

→ **nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh**

```
hadoop@fedora:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh Modificado

# 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-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jre

# 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

^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
```



27. Posteriormente, se ubica en la carpeta de hadoop.

→ **cd /home/hadoop/hadoop/etc/hadoop/**

```
hadoop@fedora:~$ cd /home/hadoop/hadoop/etc/hadoop/
```

28. Se abre el archivo core-site.xml.

→ **nano core-site.xml**

```
hadoop@fedora:~/hadoop/etc/hadoop$ nano core-site.xml
```

Nota: El ecosistema de Hadoop está compuesto por los componentes core y los componentes extendidos. Dentro de los componentes core se encuentran: Hadoop Common, HDFS, Map Reduce y Yarn.

29. Se busca la etiqueta html <configuration>.

```
<!--
 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>
</configuration>
```



30. Dentro de la etiqueta <configuration>, se agrega el siguiente código html. Posteriormente, se guarda (ctrl + o), se da enter y se sale (ctrl + x).

```
<configuration>
```

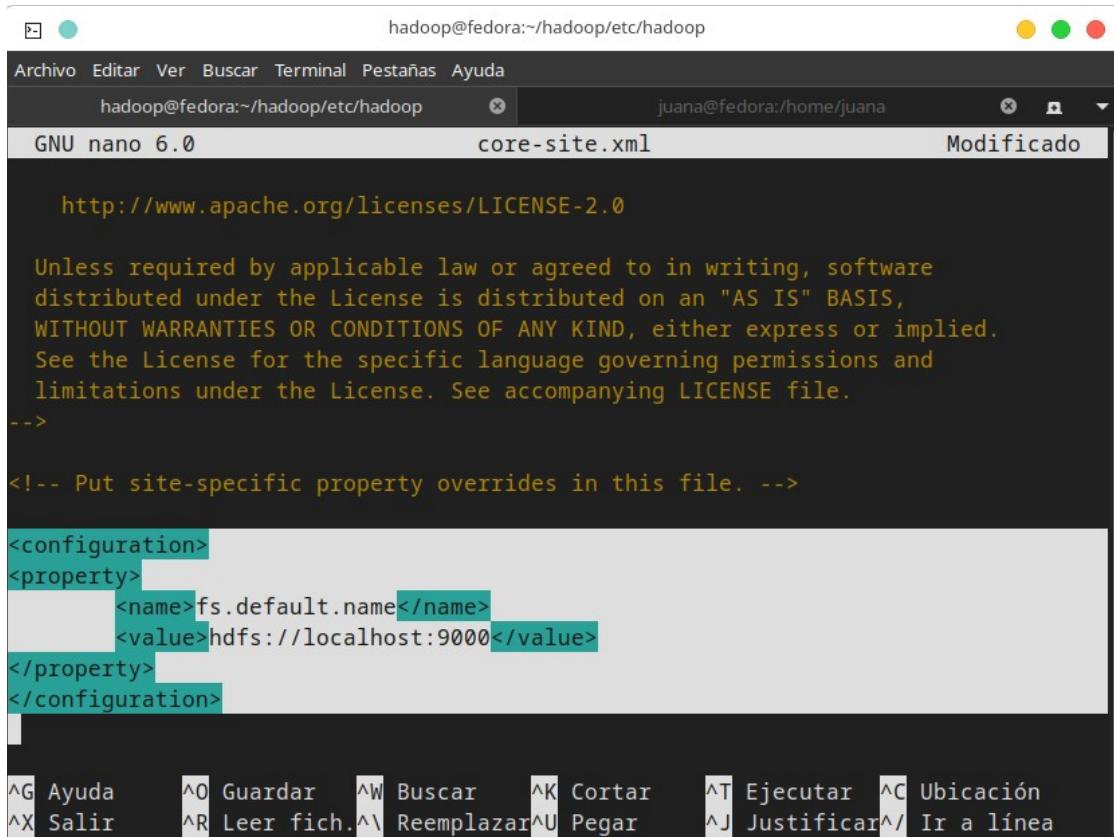
```
<property>
```

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

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

```
</property>
```

```
</configuration>
```



```
hadoop@fedora:~/hadoop/etc/hadoop
```

```
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
```

```
hadoop@fedora:~/hadoop/etc/hadoop juana@fedora:/home/juana
```

```
GNU nano 6.0 core-site.xml Modificado
```

```
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. -->
```

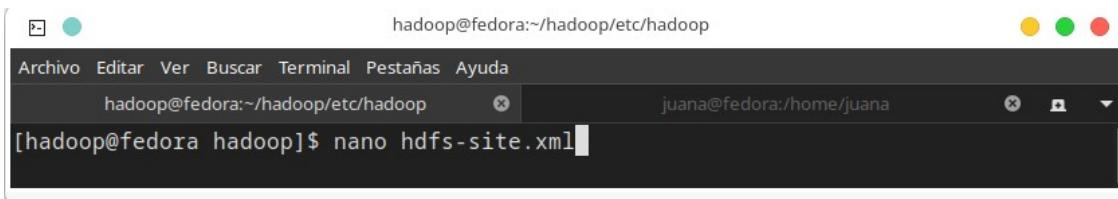
```
<configuration>
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

```
^G Ayuda      ^Q Guardar     ^W Buscar     ^K Cortar     ^T Ejecutar ^C Ubicación
^X Salir      ^R Leer fich. ^\ Reemplazar ^U Pegar      ^J Justificar ^/ Ir a línea
```



31. Se abre el archivo hdfs-site.xml.

→ **nano hdfs-site.xml**



Nota: El HDFS (sistema de archivos distribuidos de Hadoop) es el componente de almacenaje de Hadoop. Se encarga de almacenar los datos recopilados en la fase de ingestión y distribuye los datos a través de múltiples nodos. Un nodo es un sistema (computador) responsable de almacenar y procesar los datos.

32. Dentro de la etiqueta <configuration>, se agrega el siguiente código html. Posteriormente, se guarda (ctrl + o), se da enter y se sale (ctrl + x).

<**configuration**>

<**property**>

 <**namename**>

 <**valuevalue**>

</**property**>

<**property**>

 <**namename**>

 <**valuevalue**>

</**property**>

<**property**>

 <**namename**>

 <**valuevalue**>

</**property**>

</**configuration**>



```
hadoop@fedora:~/hadoop/etc/hadoop juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/etc/hadoop x juana@fedora:/home/juana x
GNU nano 6.0 hdfs-site.xml Modificado
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
    <name>dfs.replication</name>
    <value>1</value>
</property>

<property>
    <name>dfs.name.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
</property>

<property>
    <name>dfs.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
</property>
</configuration>
^G Ayuda      ^O Guardar      ^W Buscar      ^K Cortar      ^T Ejecutar ^C Ubicación
^X Salir      ^R Leer fich. ^A Reemplazar ^U Pegar      ^J Justificar ^/ Ir a línea
```

33. Se abre el archivo mapred-site.xml.

→ **nano mapred-site.xml**

```
hadoop@fedora:~/hadoop/etc/hadoop juana@fedora:/home/juana
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/etc/hadoop x juana@fedora:/home/juana x
[hadoop@fedora hadoop]$ nano mapred-site.xml
```

Nota: MapReduce es la unidad de procesamiento de datos a través de clusters. Map reduce divide los datos en unidades más pequeñas, los cuales son procesados simultáneamente. Como su nombre lo indica, Map reduce está compuesto de dos fases: Map y Reduce. En Map se procesa los datos en pares de llave y valor. Reducer funciona con múltiples funciones de mapa y agrega los pares usando sus claves para producir una salida final.



34. Dentro de la etiqueta <configuration>, se agrega el siguiente código html. Posteriormente, se guarda (ctrl + o), se da enter y se sale (ctrl + x).

<**configuration**>

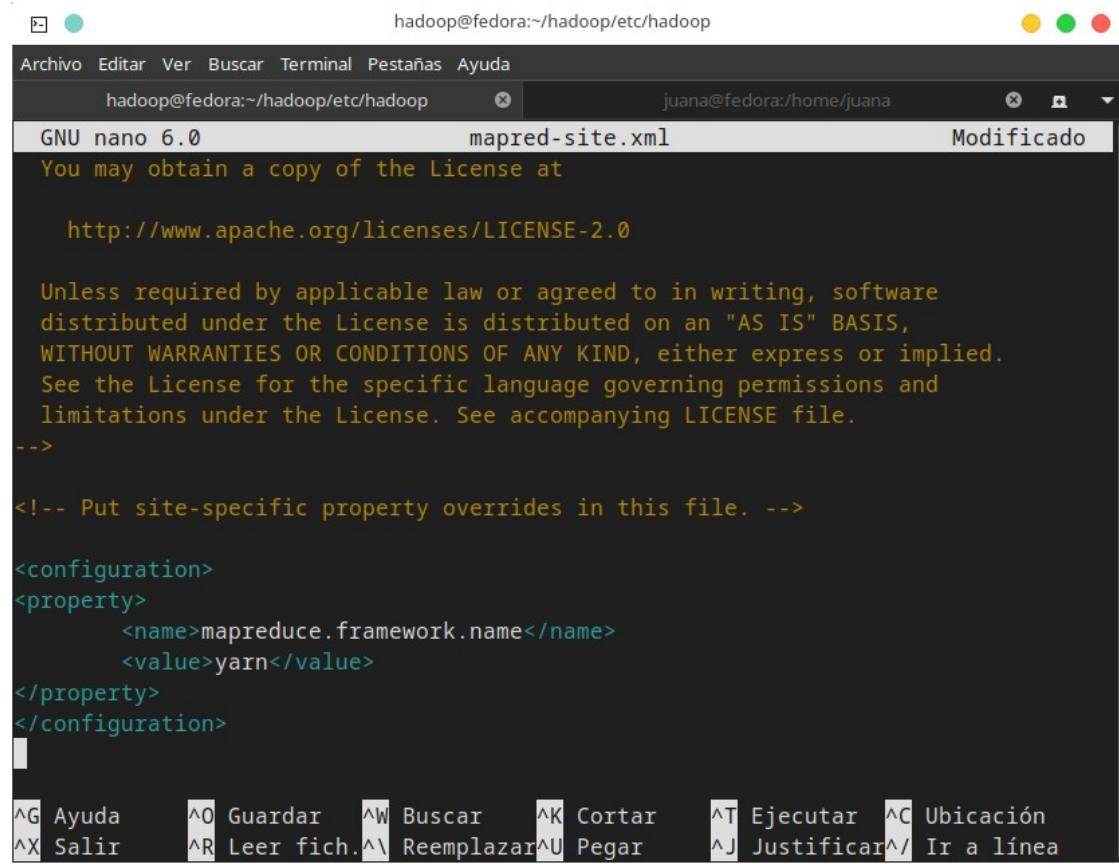
<**property**>

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

```
<value>yarn</value>
```

</**property**>

</**configuration**>



```
hadoop@fedora:~/hadoop/etc/hadoop
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/etc/hadoop      juana@fedora:/home/juana
GNU nano 6.0                         mapred-site.xml           Modificado
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>
</configuration>

^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
```



35. Se abre el archivo yarn-site.xml.

→ **nano yarn-site.xml**

```
hadoop@fedora:~/hadoop/etc/hadoop
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/etc/hadoop juana@fedora:/home/juana
[hadoop@fedora hadoop]$ nano yarn-site.xml
```

Nota: *Yarn Yet Another Resource Negotiator (YARN) es el administrador de recursos en todos los clústeres. Un clúster es una colección de computadores que trabajan juntos al mismo tiempo, para realizar tareas.*

36. Dentro de la etiqueta <configuration>, se agrega el siguiente código html. Posteriormente, se guarda (ctrl + o), se da enter y se sale (ctrl + x).

<configuration>

<property>

```
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
```

</property>

</configuration>

```
hadoop@fedora:~/hadoop/etc/hadoop
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/etc/hadoop juana@fedora:/home/juana
GNU nano 6.0          yarn-site.xml          Modificado
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.
-->
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```



37. A continuación, se da formato al nombre nodo. Esto es para que el sistema de archivos de Hadoop (HDFS) se formatee en la carpeta actual.

→ **hdfs namenode -format**

Nota: Dentro del componente de almacenamiento de Hadoop (HDFS), se encuentra el nodo nombre que se encarga de supervisar las operaciones de apertura, cierre, cambio de nombre de archivo y asignación de bloques de archivo al nodo de datos. Los nodos de datos son responsables de las solicitudes de lectura y escritura del cliente y realizan la creación, replicación y eliminación de bloques de archivos según las instrucciones del nodo de nombre.

```
=0 when meet shutdown.  
2022-09-10 22:20:37,241 INFO namenode.NameNode: SHUTDOWN_MSG:  
*****  
SHUTDOWN_MSG: Shutting down NameNode at fedora/192.168.1.58  
*****  
[hadoop@fedora hadoop]$
```

Nota: Cuando salga el siguiente mensaje, significa que el proceso se realizó de manera exitosa.



38. Se ubica en sbin.

→ **cd /home/hadoop/hadoop/sbin/**

```
hadoop@fedora:~/hadoop/sbin
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/sbin x juana@fedora:/home/juana x
[hadoop@fedora hadoop]$ cd /home/hadoop/hadoop/sbin/
[hadoop@fedora sbin]$
```

39. Se inician los namenodes.

→ **./start-dfs.sh**

```
hadoop@fedora:~/hadoop/sbin
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/sbin x juana@fedora:/home/juana x
[hadoop@fedora hadoop]$ cd /home/hadoop/hadoop/sbin/
[hadoop@fedora sbin]$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [fedora]
fedora: Warning: Permanently added 'fedora' (ED25519) to the list of known hosts.
2022-09-10 22:23:35,902 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@fedora sbin]$
```

40. Se inicia el Yarn.

→ **./start-yarn.sh**

```
hadoop@fedora:~/hadoop/sbin
Archivo Editar Ver Buscar Terminal Pestañas Ayuda
hadoop@fedora:~/hadoop/sbin x juana@fedora:/home/juana x
[hadoop@fedora sbin]$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
[hadoop@fedora sbin]$
```



41. En otra terminal, se revisan los adaptadores de red, para obtener la dirección IP (192.168.1.58).

→ **ifconfig**

```
juana@fedora:~ juana@fedora:~  
Archivo Editar Ver Buscar Terminal Pestañas Ayuda  
hadoop@fedora:~/hadoop/sbin × juana@fedora:~  
(base) [juana@fedora ~]$ ifconfig  
lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536  
    inet 127.0.0.1 netmask 255.0.0.0  
        inet6 ::1 prefixlen 128 scopeid 0x10<host>  
            loop txqueuelen 1000 (Local Loopback)  
            RX packets 3604516 bytes 1028211039 (980.5 MiB)  
            RX errors 0 dropped 0 overruns 0 frame 0  
            TX packets 3604516 bytes 1028211039 (980.5 MiB)  
            TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0  
  
wlp1s0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500  
    inet 192.168.1.58 netmask 255.255.255.0 broadcast 192.168.1.255  
        inet6 fe80::bca3:beda:707f:35 prefixlen 64 scopeid 0x20<link>  
            ether 48:45:20:6a:b1:20 txqueuelen 1000 (Ethernet)  
            RX packets 15152503 bytes 15828875327 (14.7 GiB)  
            RX errors 0 dropped 518 overruns 0 frame 0  
            TX packets 3673364 bytes 937073831 (893.6 MiB)  
            TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```



42. En el navegador se pega la dirección IP con el puerto 9870.

→ <http://192.168.1.58:9870/>

The screenshot shows a web browser window with the following details:

- URL:** http://192.168.1.58:9870/dfshealth.html#tab-overview
- Page Title:** Overview 'localhost:9000' (✓active)
- System Information Table:**

Started:	Sat Sep 10 22:23:27 -0500 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb
Compiled:	Fri Jul 29 07:32:00 -0500 2022 by stevel from branch-3.3.4
Cluster ID:	CID-5f2753d4-7050-419c-8566-f61f53a9fa71
Block Pool ID:	BP-1393222917-192.168.1.58-1662866436983
- Summary Section:**

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

43. Para ver los gestores de recursos, se encuentran en el puerto 8042.

→ <http://192.168.1.58:8042/>

The screenshot shows a web browser window with the following details:

- URL:** http://192.168.1.58:8042/node
- Page Title:** NodeManager information
- Left Sidebar:**
 - ResourceManager
 - NodeManager
 - Node Information
 - List of Applications
 - List of Containers
 - Tools
- Table Data:**

NodeManager information	
Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	8
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Sat Sep 10 22:28:50 COT 2022
NodeHealthReport	
NodeManager started on	Sat Sep 10 22:24:45 COT 2022
NodeManager Version:	3.3.4 from a585a73c3e02ac62350c136643a5e7f6095a3dbb by stevel source checksum 17e8efaf27d922f2de51e5be9e69e9 on 2022-07-29T12:51Z
Hadoop Version:	3.3.4 from a585a73c3e02ac62350c136643a5e7f6095a3dbb by stevel source checksum fb9dd8918a7b8a5b430d61af858f6ec on 2022-07-29T12:32Z



44. Para mostrar la información de las piscinas de bloqueo, se encuentra en el puerto 9864.

→ <http://192.168.1.58:9864/>

The screenshot shows a web browser window with the following details:

- Address Bar:** No seguro | 192.168.1.58:9864/datanode.html
- Header Bar:** Hadoop Overview Utilities ▾
- Main Content:** DataNode on fedora:9866
- Data Table:** Shows cluster information:

Cluster ID:	CID-5f2753d4-7050-419c-8566-f61f53a9fa71
Started:	Sat Sep 10 22:23:31 -0500 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb
- Section:** Block Pools
- Table:** Shows block pool statistics:

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-1393222917-192.168.1.58-1662866436983	RUNNING	0s	8 minutes	0 B (128 MB)
- Section:** Volume Information



Comandos básicos de Hadoop:

Para obtener más información sobre los comandos, se puede acceder a la página:
<https://www.alluxio.io/learn/hdfs/basic-file-operations-commands/>

1. Ayuda general de los comandos que se pueden utilizar con hadoop:

→ **hadoop**

```
hadoop@fedora:~$ 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:

daemonlog      get/set the log level for each daemon

  Client Commands:

archive        create a Hadoop archive
checknative    check native Hadoop and compression libraries availability
classpath       prints the class path needed to get the Hadoop jar and the required libraries
conftest        validate configuration XML files
credential     interact with credential providers
distch         distributed metadata changer
distcp         copy file or directories recursively
dtutil         operations related to delegation tokens
envvars        display computed Hadoop environment variables
fs             run a generic filesystem user client
gridmix        submit a mix of synthetic job, modeling a profiled from production load
jar <jar>       run a jar file. NOTE: please use "yarn jar" to launch YARN applications, not this command.
jnopath        prints the java.library.path
kdiag          Diagnose Kerberos Problems
kerbname       show auth_to_local principal conversion
key            manage keys via the KeyProvider
rumenfolder    scale a rumen input trace
rumentrace     convert logs into a rumen trace
```



2. Con base en la información del anterior comando, se puede verificar la versión de Hadoop:

→ **hadoop version**

```
hadoop@fedora:~$ hadoop version
Hadoop 3.3.4
Source code repository https://github.com/apache/hadoop.git -r a585a73c3e02ac62350c136643a5e7f6095a3dbb
Compiled by stevel on 2022-07-29T12:32Z
Compiled with protoc 3.7.1
From source with checksum fb9dd8918a7b8a5b430d61af858f6ec
This command was run using /home/hadoop/hadoop/share/hadoop/common/hadoop-common-3.3.4.jar
[hadoop@fedora ~]$
```

3. Para iniciar el servicio de Hadoop:

→ **./start-dfs.sh**

```
hadoop@fedora:~/hadoop/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [fedora]
fedora: Warning: Permanently added 'fedora' (ED25519) to the list of known hosts.
2022-09-10 22:23:35,902 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@fedora sbin]$
```



4. Verificar que servicios se están ejecutando actualmente:

→ **jps**

```
hadoop@fedora:~$ jps
590440 NodeManager
589360 NameNode
589554 DataNode
590292 ResourceManager
655159 Jps
589863 SecondaryNameNode
[hadoop@fedora ~]$
```

5. Ayuda general del comando hdfs:

→ **hdfs**

```
hadoop@fedora:~$ hdfs
Usage: hdfs [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]

    OPTIONS is none or any of:

    --buildpaths           attempt to add class files from build tree
    --config dir           Hadoop config directory
    --daemon (start|status|stop)
    --debug                operate on a daemon
    --help                 turn on shell script debug mode
    --hostnames list[,of,host,names]
    --hosts filename        usage information
    --loglevel level       hosts to use in worker mode
    --workers               list of hosts to use in worker mode
                           set the log4j level for this command
                           turn on worker mode

    SUBCOMMAND is one of:

        Admin Commands:

        cacheadmin      configure the HDFS cache
        crypto          configure HDFS encryption zones
        debug           run a Debug Admin to execute HDFS debug commands
        dfsadmin         run a DFS admin client
        dfrouteradmin   manage Router-based federation
        ec              run a HDFS ErasureCoding CLI
        fsck            run a DFS filesystem checking utility
        haadmin          run a DFS HA admin client
        jmxget          get JMX exported values from NameNode or DataNode.
        oev             apply the offline edits viewer to an edits file
        oiv             apply the offline fsimage viewer to an fsimage
        oiv_legacy      apply the offline fsimage viewer to a legacy fsimage
        storagepolicies list/get/set/satisfyStoragePolicy block storage policies
```



```
hadoop@fedora:~
```

Archivo Editar Ver Buscar Terminal Ayuda

Client Commands:

```
classpath      prints the class path needed to get the hadoop jar and the required libraries
dfs           run a filesystem command on the file system
envvars       display computed Hadoop environment variables
fetchht      fetch a delegation token from the NameNode
getconf        get config values from configuration
groups         get the groups which users belong to
lsSnapshottableDir   list all snapshottable dirs owned by the current user
snapshotDiff    diff two snapshots of a directory or diff the current directory contents with a
                snapshot
version        print the version
```

Daemon Commands:

```
balancer      run a cluster balancing utility
datanode       run a DFS datanode
dfsrouter     run the DFS router
diskbalancer  Distributes data evenly among disks on a given node
httpfs         run HttpFS server, the HDFS HTTP Gateway
journalnode   run the DFS journalnode
mover          run a utility to move block replicas across storage types
namenode       run the DFS namenode
nfs3           run an NFS version 3 gateway
portmap        run a portmap service
secondarynamenode  run the DFS secondary namenode
sps            run external storagepolicysatisfier
zkfc           run the ZK Failover Controller daemon
```

SUBCOMMAND may print help when invoked w/o parameters or with -h.

```
[hadoop@fedora ~]$
```



6. Dentro de hdfs se encuentra el comando dfs, que permite realizar operaciones sobre los sistemas de archivos. A continuación, se ve una ayuda general de este comando:

→ **hdfs dfs**

```
hadoop@fedora:~$ hdfs dfs
2022-09-12 00:24:34,659 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
-java classes where applicable
Usage: hadoop fs [generic options]
      [-appendToFile <localsrc> ... <dst>]
      [-cat [-ignoreCrc] <src> ...]
      [-checksum [-v] <src> ...]
      [-chgrp [-R] GROUP PATH...]
      [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
      [-chown [-R] [OWNER]:[GROUP] PATH...]
      [-concat <target path> <src path> <src path> ...]
      [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
      [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
      [-count [-q] [-h] [-v] [-t <storage type>] [-u] [-x] [-e] [-s] <path> ...]
      [-cp [-f] [-p | -p[topax]] [-d] [-t <thread count>] [-q <thread pool queue size>] <src> ... <dst>]
      [-createSnapshot <snapshotDir> [<snapshotName>]]
      [-deleteSnapshot <snapshotDir> <snapshotName>]
      [-df [-h] [<path> ...])
      [-du [-s] [-h] [-v] [-x] <path> ...]
      [-expunge [-immediate] [-fs <path>]]
      [-find <path> ... <expression> ...]
      [-get [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
      [-getfacl [-R] <path>]
      [-getattr [-R] {-n name | -d} {-e en} <path>]
      [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
      [-help [-f] [-h] ...]
      [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...])
      [-mkdir [-p] <path> ...]
      [-moveFromLocal [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
      [-moveToLocal <src> <localdst>]
      [-mv <src> ... <dst>]
      [-put [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
      [-renameSnapshot <snapshotDir> <oldName> <newName>]
      [-rm [-f] [-r|-R] [-skipTrash] [-safely] <src> ...]
      [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
      [-setfacl [-R] {[-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
      [-setfattr {-n name [-v value] | -x name} <path>]
      [-setrep [-R] [-w] <rep> <path> ...]
      [-stat [format] <path> ...]
      [-tail [-f] [-s <sleep interval>] <file>]
      [-test [-defswrz] <path>]
      [-text [-ignoreCrc] <src> ...]
      [-touch [-a] [-m] [-t TIMESTAMP (yyyyMMddHHmmss) ] [-c] <path> ...]
      [-touchz <path> ...]
      [-truncate [-w] <length> <path> ...]
      [-usage [cmd ...]]]

Generic options supported are:
-conf <configuration file>           specify an application configuration file
-D <property=value>                  define a value for a given property
-fs <file:///|hdfs://>namenode:port>  specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local|resourcemanager:port>       specify a ResourceManager
-files <file1,...>                  specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>                  specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>             specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

[hadoop@fedora ~]$
```



7. Crear una carpeta:

→ **hdfs dfs -mkdir /folder_name**

8. Mostrar el contenido desde el Browse Directory:

→ **hdfs dfs -ls /**

9. Subir un archivo al HDFS. El archivo file_name.txt se encuentra de manera local y se adiciona en la carpeta folder_name en Hadoop. Normalmente, este proceso lo realiza herramientas como Hive.

→ **hdfs dfs -put file_name.txt /folder_name**

10. Para mostrar el contenido de un archivo:

→ **hdfs dfs -cat /folder_name/file_name.txt**

11. Copiar un archivo a otra carpeta:

→ **hdfs dfs -cp /folder_name_2/file_name.txt / folder_name_2/change_file_name.txt**

12. Eliminar un archivo:

→ **hdfs dfs -rm /folder_name_2/change_file_name.txt**



Referencias:

Vijay Kant. [Vijay Kant] (2022). Hadoop Installation| Fedora | Hadoop 3.2.2 | Easy steps |.
https://www.youtube.com/watch?v=FwazmbpjKc&ab_channel=VijayKant

Coursera (2022). Introduction to Big Data with Spark and Hadoop.
https://www.coursera.org/learn/introduction-to-big-data-with-spark-hadoop?irclickid=0ovVeC2FVxyNURBSAg1LQ3j8UkDWhiUmdUVpxY0&irgwc=1&utm_medium=partners&utm_source=impact&utm_campaign=2729436&utm_content=b2c

Linuxize. (2020). Ls Command in Linux (List Files and Directories). <https://linuxize.com/post/how-to-list-files-in-linux-using-the-ls-command/>

Auribox Training (2020). MapReduce con Hadoop #3: Contando palabras con Map Reduce y Java.
https://youtu.be/woUzV_liwto?list=PLG1t8jaLbxADG_cmlBYgkGW-TZw5DP3B