

# Hadoop: Taller Apache Hive

Autor: Ing. Luis Felipe Narvaez Gomez. E-mail: [luis.narvaez@usantoto.edu.co](mailto:luis.narvaez@usantoto.edu.co). Cod: 2312660. Facultad de Ingeniería de Sistemas.

Este taller esta realizado en Fedora 36 Linux, virtualizado en Windows 10 Home single Language , mediante el uso del software de Virtual Box. Para mas detalles de como instalar Fedora 36 y de como instalar en el Hadoop le recomendamos ver las guías anteriores.

## INICIAR HADOOP

Teniendo previamente instalado y funcionando correctamente hadoop en nuestro SO de Fedora 36. Volvemos nuevamente a iniciar nuestro Hadoop.

1. Abrir el nuevo usuario de hadoop como root. Es posible que nos pida la contraseña de hadoop o la de nuestro usuario principal.

```
sudo su - hadoop
```

2. Situarnos en la raiz de este usuario.

```
cd
```

3. Ir a la carpeta de hadoop y abrir el folder de sbin.

```
cd /home/hadoop/hadoop/sbin/
```

4. Iniciar el namenodes.

```
./start-dfs.sh
```

5. Iniciar el yarn.

```
./start-yarn.sh
```

6. Abrir otra terminal nueva y acceder nuevamente como super usuario de hadoop. Luego preguntar la direccion IPv4

```
[Shift] + [CTRL] + [T] => para abrir una nueva pestaña
[Shift] + [CTRL] + [N] => para abrir una nueva ventana terminal

sudo su - hadoop
cd
ifconfig
```

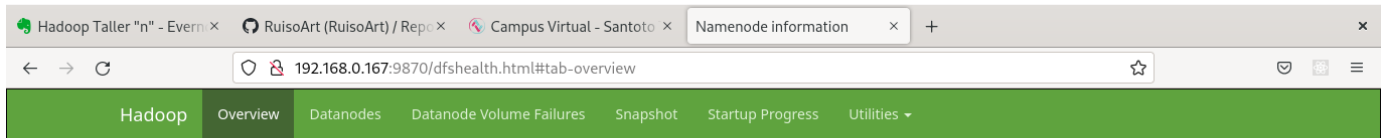
7. Rescatar la Ipv4 que tengamos y guardarla para después.

```
in: ifconfig

out:
...
broadcast 192.168.0.255
...
```

8. Verificar el correcto funcionamiento en su navegador de internet predilecto el OVERVIEW

http://192.168.0.167:9870



## Overview 'localhost:9000' (✓active)

Started:	Tue Sep 20 15:39:29 -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-59aa29a5-3511-485e-8605-6d9c489ecc11
Block Pool ID:	BP-1750418209-192.168.0.167-1663563249094

## Summary

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).

9. Verificar el correcto funcionamiento en su navegador de internet predilecto el Gestor de Recursos.

http://192.168.0.167:8042

A screenshot of a web browser showing the Hadoop NodeManager information page. The browser has multiple tabs open, including 'Hadoop Taller "n" - Evern', 'RuisoArt (RuisoArt) / Repo', 'Campus Virtual - Santoto', and 'NodeManager information'. The address bar shows '192.168.0.167:8042/node'. The page features the Hadoop logo on the left and a sidebar with navigation links: 'ResourceManager', 'NodeManager', 'Node Information', 'List of Applications', 'List of Containers', and 'Tools'. The 'NodeManager' link is selected. The main content area is titled 'NodeManager information' and displays various system metrics and status information in a table-like format.

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=M), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Tue Sep 20 16:03:23 COT 2022
NodeHealthReport	
NodeManager started on	Tue Sep 20 15:43:13 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

10. Verificar el correcto funcionamiento en su navegador de internet predilecto las Piscinas de Bloqueo.

http://192.168.0.167:9864

Hadoop Taller "n" - Evern × RuisoArt (RuisoArt) / Rep × Campus Virtual - Santoto × DataNode Information × +

← → ↻ 192.168.0.167:9864/datanode.html ☆ 📧 🖨 ⋮

Hadoop Overview Utilities ▾

## DataNode on fedora:9866

Cluster ID:	CID-59aa29a5-3511-485e-8605-6d9c489ecc11
Started:	Tue Sep 20 15:39:46 -0500 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb

## Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-1750418209-192.168.0.167-1663563249094	RUNNING	1s	20 minutes	0 B (128 MB)

## Volume Information

Tenga en cuenta que en caso de funcionar con la Plv4 que muestra su computadora, debe recordar que la instalación de Hadoop a veces se liga a la red con la que lo instalo inicialmente por lo que le sugerimos trabar en la misma red de instalación o dirigirse a la guía de instalación de hadoop en Fedora en la sección del paso 37 donde se obtendrá el nombre del nodo al formatear el HDFS.

## JAVA JDK Y JRE

Primero vamos a ver que versiones de JDK tenemos instaladas en nuestra maquina.

```
input:
java -version
```

```
output:
openjdk version "1.8.0_345"
OpenJDK Runtime Environment (build 1.8.0_345-b01)
OpenJDK 64-Bit Server VM (build 25.345-b01, mixed mode)
```

En este caso trataremos de trabajar con el JRE versión 1.8 y no con JDK's mas actuales, esto debido a la construcción de estos últimos la cual interfiere con algunos programas de Apache, por lo que buscamos un interprete de Java que sea genérico y utilizable a nivel de operario y no desarrollador. De esta manera tratamos de garantizar que la compatibilidad y usos de Apache Hive no se vea comprometida y nos signifique problemas a futuro.

## INSTALAR APACHE HIVE

Vamos primer a averiguar la ruta de instalación de nuestro java.

```
input: which java
output: /usr/bin/java
```

Ahora utilizamos READLINK.

input: readlink -f /usr/bin/java

output: /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86\_64/jre/bin/java

Nos quedaremos con la ruta hasta antes del "/bin/java" para mas adelante.

Ahora entremos a super usuario de hadoop de la siguiente manera:

```
sudo su - hadoop
```

Vera un comportamiento similar en consola a lo siguiente:

```
(base) [onruiso@fedora ~]$ sudo su - hadoop
[sudo] password for onruiso:
[hadoop@fedora ~]$
```

Consultemos la versión de hadoop que estamos utilizando.

```
input: hadoop version
```

output:

```
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
```

Ahora descarguemos Apache Hive en su versión mas reciente, este comando que ver a continuación variara dependiendo de la versión en la que se encuentra Hive por lo que se recomienda que visite la pagina oficial de Apache Hive antes de continuar.

```
wget http://archive.apache.org/dist/hive/hive-0.12.0/hive-0.12.0-bin.tar.gz
```

Tendríamos de esta manera lo siguiente:



```
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive-0.12.0-bin.tar.gz
hadoop      hadoopdata           part-r-00000
[hadoop@fedora ~]$
```

Esto descargar el comprimido de Apache Hive que usaremos para instalarle. Ahora vamos a descomprimir este "tar.gz" de la siguiente manera:

```
tar xzf hive-0.12.0-bin.tar.gz
```

Luciendo de la siguiente manera:

```
hadoop@fedora:~  
[hadoop@fedora ~]$ ls  
dracula.txt  hadoop-3.3.4.tar.gz  hive-0.12.0-bin.tar.gz  
hadoop      hadoopdata           part-r-00000  
[hadoop@fedora ~]$ tar xzf hive-0.12.0-bin.tar.gz  
[hadoop@fedora ~]$ ls  
dracula.txt  hadoop-3.3.4.tar.gz  hive-0.12.0-bin      part-r-00000  
hadoop      hadoopdata          hive-0.12.0-bin.tar.gz  
[hadoop@fedora ~]$
```

Instalamos ahora el editor de texto de GEDIT.

```
sudo dnf install gedit
```

Ahora vamos a entrar al fichero BASHRC para modificar algunas rutas de la siguiente manera:

```
gedit ~/.bashrc
```

En caso de no abrir con gedit podemos utilizar la vieja confiable nano.

```
nano ~/.bashrc
```

Abriendo una terminal similar a la siguiente:

```
GNU nano 6.0 /home/hadoop/.bashrc  
if [ -d ~/.bashrc.d ]; then  
    for rc in ~/.bashrc.d/*; do  
        if [ -f "$rc" ]; then  
            . "$rc"  
        fi  
    done  
fi  
  
unset rc  
  
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  
^X Salir      ^R Leer fich. ^_ Reemplazar  ^U Pegar      ^J Justificar ^_ Ir a línea
```

Las líneas a cambiar son:

```
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 PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin  
  
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
```

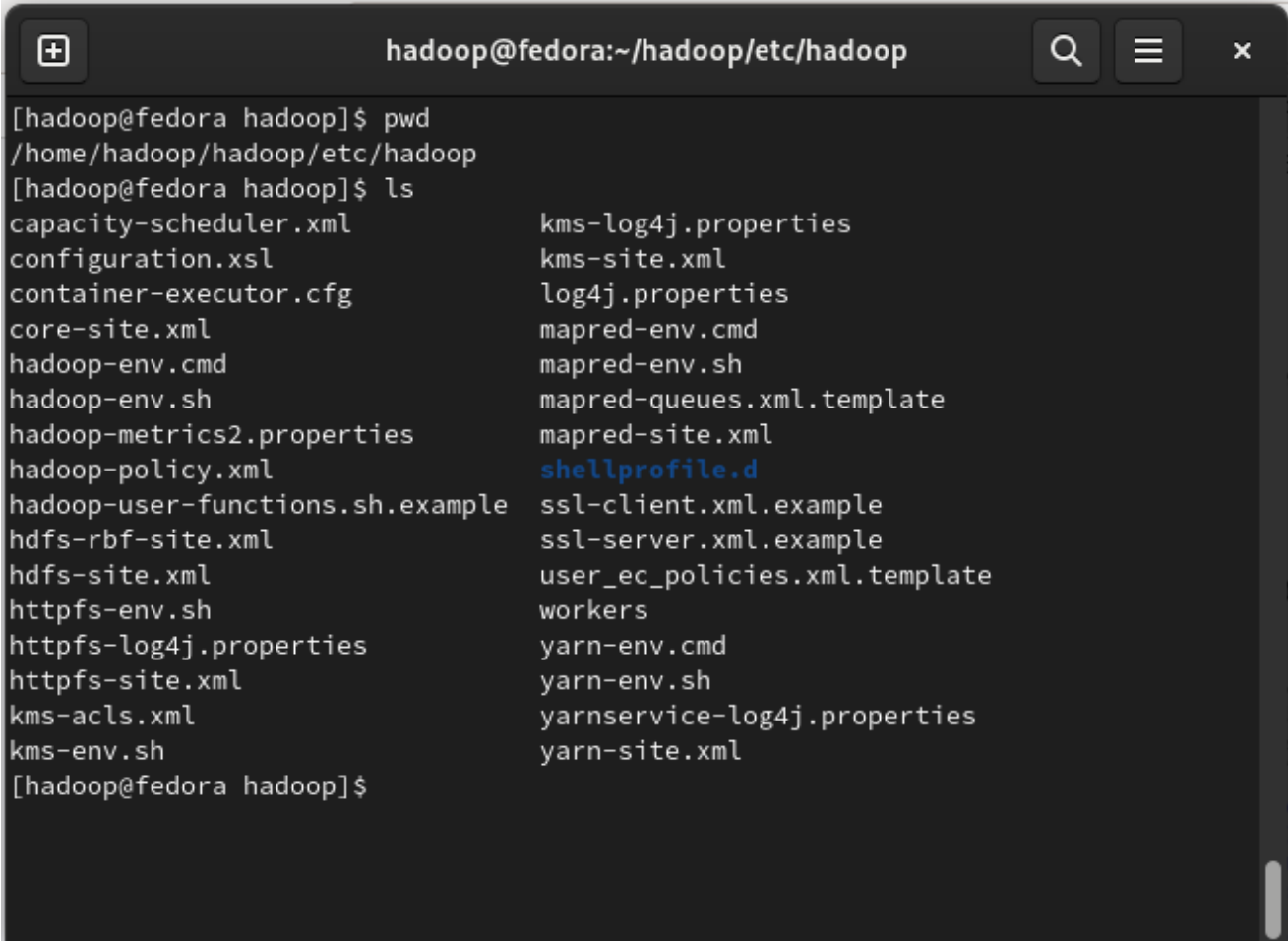
```
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
```

La ruta de JAVA\_HOME corresponde a la hallada con "readlink -f /usr/bin/java", la ruta de hadoop corresponde a la ruta donde se instaló hadoop, en mi caso en el mismo "home" puedo entrar a la carpeta con "cd hadoop" y luego dar "pwd". La tercera es una ruta path que podemos dejar así o dado el caso cambiar a "export PATH=\$PATH:\$HADOOP\_HOME/bin" según como se requiera. Ahora vamos a actualizar la información del archivo en nuestro sistema mediante el uso de la terminal con el comando:

```
source ~/.bashrc
```

Ahora vamos a confirmar los archivos de configuración de Hadoop de la siguiente manera:

1. Primero navegamos hasta el directorio de hadoop.
- 2.



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

3. Ahora abrimos el archivo [hadoop-env.sh](#) y confirmamos la JAVA\_HOME para que luzca de la siguiente manera:

4. 

```
[hadoop@fedora hadoop]$ nano hadoop-env.sh
```

5. 

```
# 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

# Location of Hadoop. By default, Hadoop will attempt to determine
```

6. (NOTA: La anterior imagen toma como enlace al final la parte /jre/ para funcionar correctamente)
7. Confirmemos ahora la versión de hadoop

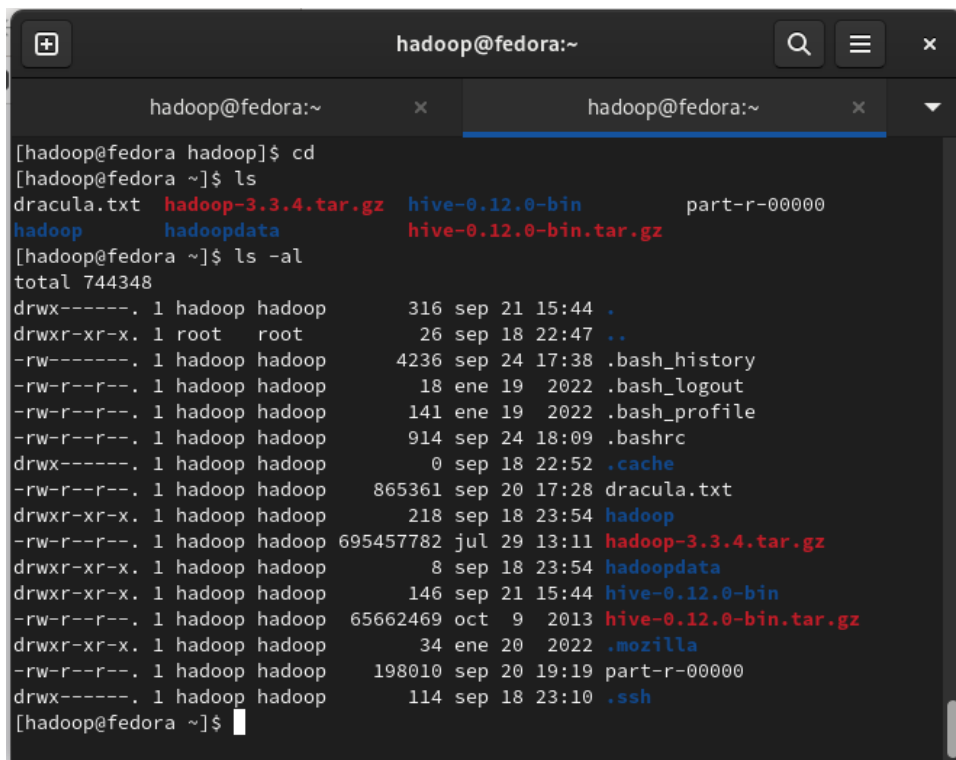
```
hadoop version
```

Teniendo una vista como la siguiente:

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

Ahora, devuelta a nuestra unidad donde descargamos y descomprimos apache hive:

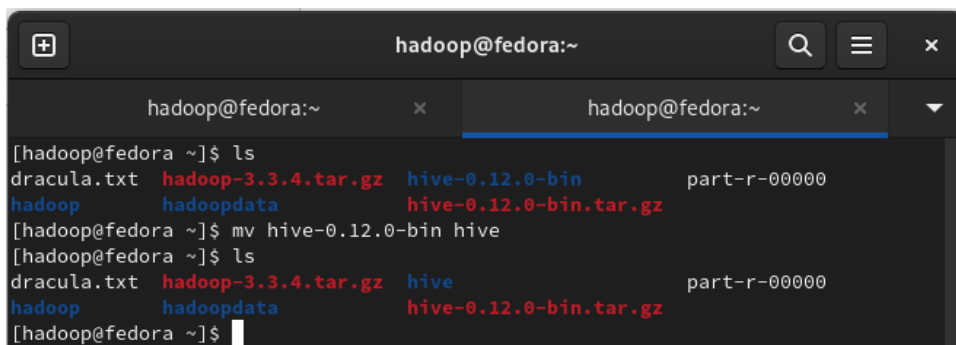
```
cd
ls -al
```

A terminal window titled 'hadoop@fedora:~' with two tabs. The first tab shows the command 'cd' and 'ls', listing files like 'dracula.txt', 'hadoop-3.3.4.tar.gz', 'hive-0.12.0-bin', and 'part-r-00000'. The second tab shows 'ls -al', displaying a detailed directory listing including permissions, owner, size, date, and file names like '.bash\_history', '.cache', 'hadoop', 'hadoop-3.3.4.tar.gz', 'hadoopdata', 'hive-0.12.0-bin', 'hive-0.12.0-bin.tar.gz', 'mozilla', 'part-r-00000', and 'ssh'.

```
[hadoop@fedora hadoop]$ cd
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive-0.12.0-bin  part-r-00000
hadoop      hadoopdata          hive-0.12.0-bin.tar.gz
[hadoop@fedora ~]$ ls -al
total 744348
drwx-----. 1 hadoop hadoop      316 sep 21 15:44 .
drwxr-xr-x.  1 root   root        26 sep 18 22:47 ..
-rw-----. 1 hadoop hadoop    4236 sep 24 17:38 .bash_history
-rw-r--r--. 1 hadoop hadoop      18 ene 19 2022 .bash_logout
-rw-r--r--. 1 hadoop hadoop     141 ene 19 2022 .bash_profile
-rw-r--r--. 1 hadoop hadoop     914 sep 24 18:09 .bashrc
drwx-----. 1 hadoop hadoop        0 sep 18 22:52 .cache
-rw-r--r--. 1 hadoop hadoop  865361 sep 20 17:28 dracula.txt
drwxr-xr-x. 1 hadoop hadoop      218 sep 18 23:54 hadoop
-rw-r--r--. 1 hadoop hadoop 695457782 jul 29 13:11 hadoop-3.3.4.tar.gz
drwxr-xr-x. 1 hadoop hadoop        8 sep 18 23:54 hadoopdata
drwxr-xr-x. 1 hadoop hadoop     146 sep 21 15:44 hive-0.12.0-bin
-rw-r--r--. 1 hadoop hadoop 65662469 oct  9 2013 hive-0.12.0-bin.tar.gz
drwxr-xr-x. 1 hadoop hadoop      34 ene 20 2022 .mozilla
-rw-r--r--. 1 hadoop hadoop  198010 sep 20 19:19 part-r-00000
drwx-----. 1 hadoop hadoop     114 sep 18 23:10 .ssh
[hadoop@fedora ~]$
```

Movemos nuestro descomprimido a una nueva carpeta

```
mv hive-0.12.0-bin hive
```

A terminal window titled 'hadoop@fedora:~' with two tabs. The first tab shows the command 'ls' and the directory listing from the previous screenshot. The second tab shows the command 'mv hive-0.12.0-bin hive' followed by another 'ls' command. The updated directory listing now shows 'hive' instead of 'hive-0.12.0-bin' and 'hive-0.12.0-bin.tar.gz' instead of 'hive-0.12.0-bin.tar.gz'.

```
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive-0.12.0-bin  part-r-00000
hadoop      hadoopdata          hive-0.12.0-bin.tar.gz
[hadoop@fedora ~]$ mv hive-0.12.0-bin hive
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive  part-r-00000
hadoop      hadoopdata          hive-0.12.0-bin.tar.gz
[hadoop@fedora ~]$
```

Damos permisos especiales de usuario a la carpeta recién creada HIVE

```
chown -R (usuario) hive
```

En mi caso seria:

```
chown -R hadoop hive
```

Vamos a agregar las siguientes nuevas líneas de comandos en el BASCHRC

```
export HADOOP_HOME=/home/(usuario) /hadoop
export HADOOP_PREFIX=/home/(usuario)/hadoop
export HIVE_HOME=/home/(usuario)/hive
export PATH=$HIVE_HOME/bin:$PATH
```

En mi caso no utilizaré todas las líneas, esto es debido a que varias ya las tengo como rutas efinidas dentro de mi BASHRC.

Cambiamos de esto:

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jr>
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
```

A esto:

```
export HADOOP_PREFIX=/home/hadoop/hadoop
export HIVE_HOME=/home/hadoop/hive
```

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.345.b01-1.fc36.x86_64/jr>
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

export HADOOP_PREFIX=/home/hadoop/hadoop
export HIVE_HOME=/home/hadoop/hive
```

La siguiente linea de codigo solo es necesaria

```
(export PATH=$HIVE_HOME/bin:$PATH)
```



```
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

export HADOOP_PREFIX=/home/hadoop/hadoop
export HIVE_HOME=/home/hadoop/hive

export PATH=$HIVE_HOME/bin:$PATH
```

Para ver que todo ande bien, podemos comprobar con hadoop versión, en caso de que algo salga mal, tendremos problemas con la lectura del JRE de Java y no podremos utilizar Hadoop, menos Hive. En caso de que esto suceda se recomienda verificar las rutas de Java.

1. Tener seleccionado el Java que se necesita con el comando:

```
sudo update-alternatives --config java
```

Tener en cuenta que la ruta de java va hasta /jre/ y no mas.

2. Copiar la ruta anterior y reemplazarla en JAVA\_HOME en [hadoop-env.sh](#) , la ruta dependerá de donde se tenga instalado hadoop.

```
cd /home/hadoop/hadoop/etc/hadoop
nano hadoop-env.sh
```

3. Y también reemplazar la ruta JAVA\_HOME en Bashrc

```
nano ~/.bashrc
```

4. Confirmar el cambio con

```
source ~/.bashrc
```

Antes de proseguir con cualquier uso de Hive, debemos crear las siguientes carpetas:

```
cd /home/(usuario)/hive
$HADOOP_HOME/bin/hadoop fs -mkdir /tmp
$HADOOP_HOME/bin/hadoop fs -mkdir /user/hive/warehouse
$HADOOP_HOME/bin/hadoop fs -chmod g+w /tmp
$HADOOP_HOME/bin/hadoop fs -chmod g+w /user/hive/warehouse
```

De la siguiente manera:

```
[hadoop@fedora ~]$ pwd
/home/hadoop
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive  part-r-00000
hadoop      hadoopdata           hive-0.12.0-bin.tar.gz
[hadoop@fedora ~]$ cd hive/
[hadoop@fedora hive]$ ls
bin  examples  lib  NOTICE  RELEASE_NOTES.txt
conf hcatalog  LICENSE  README.txt  scripts
[hadoop@fedora hive]$
```

Primer comando:

```
[hadoop@fedora hive]$ $HADOOP_HOME/bin/hadoop fs -mkdir /tmp
2022-09-24 19:09:23,378 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
mkdir: Call From fedora/192.168.0.167 to localhost:9000 failed on connection exc
eption: java.net.ConnectException: Conexión rehusada; For more details see: htt
p://wiki.apache.org/hadoop/ConnectionRefused
[hadoop@fedora hive]$
```

Segundo Comando:

```
[hadoop@fedora hive]$ $HADOOP_HOME/bin/hadoop fs -mkdir /user/hive/warehouse
2022-09-24 19:10:13,989 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
mkdir: Call From fedora/192.168.0.167 to localhost:9000 failed on connection exc
eption: java.net.ConnectException: Conexión rehusada; For more details see: htt
p://wiki.apache.org/hadoop/ConnectionRefused
[hadoop@fedora hive]$
```

Tercer comando:


```
[hadoop@fedora hive]$ $HADOOP_HOME/bin/hadoop fs -chmod g+w /tmp
2022-09-24 19:10:59,908 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
chmod: Call From fedora/192.168.0.167 to localhost:9000 failed on connection exc
eption: java.net.ConnectException: Conexión rehusada; For more details see: htt
p://wiki.apache.org/hadoop/ConnectionRefused
[hadoop@fedora hive]$
```

Cuarto comando:




```
[hadoop@fedora hive]$ $HADOOP_HOME/bin/hadoop fs -chmod g+w /user/hive/warehouse
2022-09-24 19:11:44,213 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
chmod: Call From fedora/192.168.0.167 to localhost:9000 failed on connection exc
eption: java.net.ConnectException: Conexión rehusada; For more details see: htt
p://wiki.apache.org/hadoop/ConnectionRefused
[hadoop@fedora hive]$
```

## EMPEZAR A USAR HIVE

Primero descargaremos el archivo con el que trabajaremos:

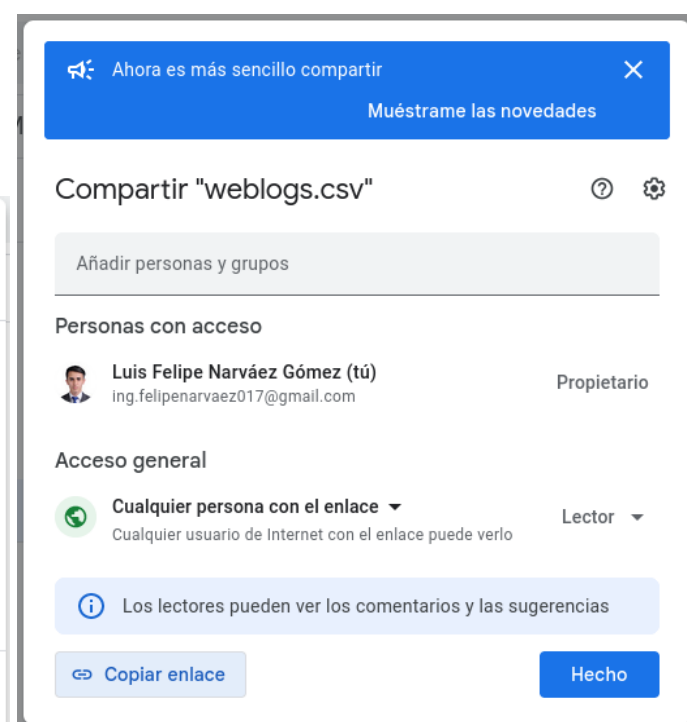
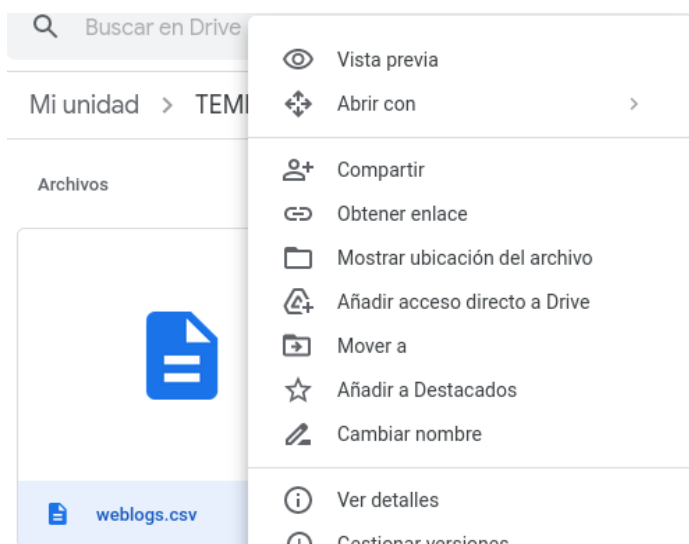
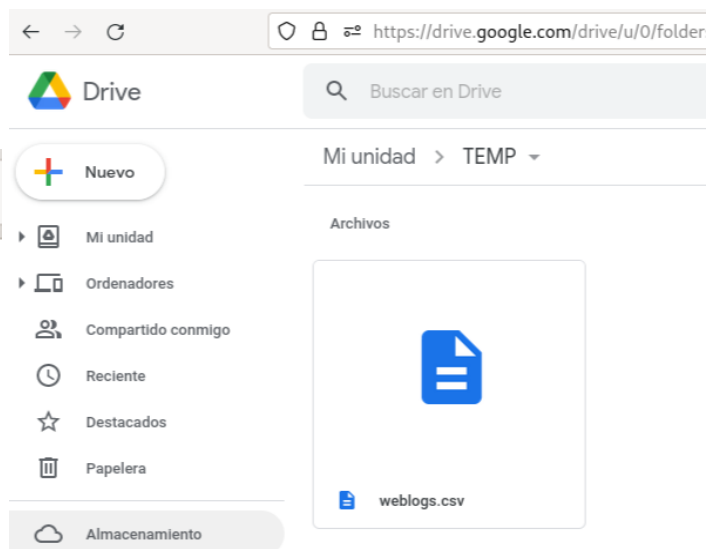
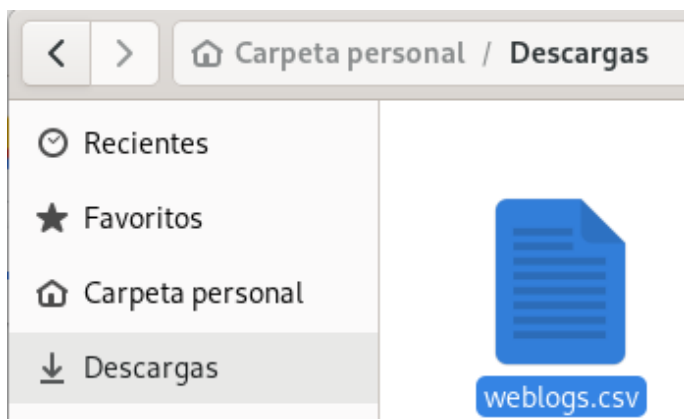
 Apache Hive  
de [DERGHIKZON MARTINEZ MONTERO](#) - lunes, 19 de septiembre de 2022, 09:59

Exposición y Laboratorio de Apache Hive, realizado por Derghikzon Martinez Montero y Maria Paula Aguilar Malaver

 [Apache Hive.pdf](#)  
 [Laboratorio Hive.pdf](#)  
 [weblogs.csv](#)

[Enlace permanente](#) [Responder](#)

En mi cas lo descargue en mi equipo root, pero al no poderlo pasar directamente al usuario Hadoop, resuelvo subirlo a un enlace publico drive y descargarlo en hadoop directamente desde allí con la funcion wget.



Obtendríamos un enlace temporal como este:

<https://drive.google.com/file/d/1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1/view?usp=sharing>

Sin embargo necesitamos el enlace de la forma de descarga, para esto nos quedamos con el id del enlace, este figura después de /d/ y antes de /view, así:

1. <https://drive.google.com/file/d/1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1/view?usp=sharing>
2. <https://drive.google.com/file/d/1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1/view?usp=sharing>
3. 1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1

Lo añadimos a la forma:

1. <https://drive.google.com/uc?export=download&id=>
2. <https://drive.google.com/uc?export=download&id=1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1>

Armando un comando como ete:

```
wget --no-check-certificate 'link' -O archivo_salida
```

```
wget --no-check-certificate 'https://drive.google.com/uc?export=download&id=1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1' -O weblogs.csv
```

Al descargarse tendríamos un archivo de este modo:

```
[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive  part-r-00000
hadoop      hadoopdata           hive-0.12.0-bin.tar.gz
[hadoop@fedora ~]$ wget --no-check-certificate 'https://drive.google.com/uc?export=download&id=1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1' -O weblogs.csv
--2022-09-24 20:44:30-- https://drive.google.com/uc?export=download&id=1jJ7V3-ZzLtDjXlfesKFhRXUrnxEv0FY1
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 38787883 (37M) [text/csv]
Grabando a: «weblogs.csv»

weblogs.csv          100%[=====] 36,99M  9,60MB/s  en 4,1s

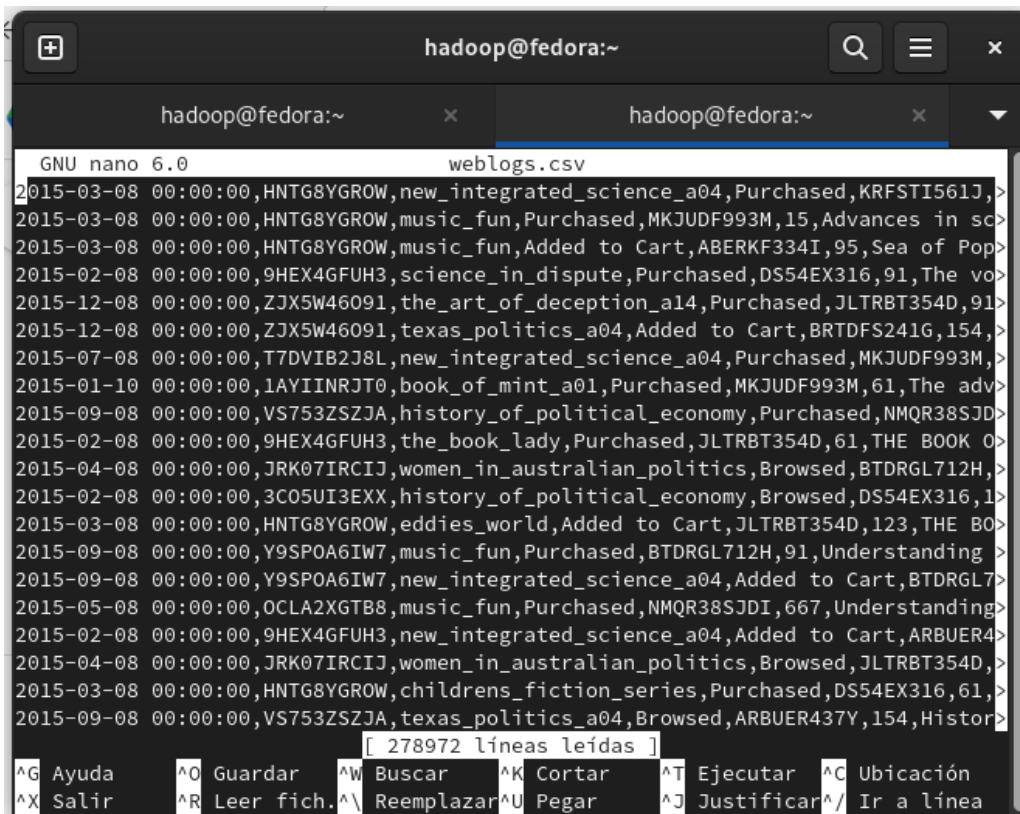
2022-09-24 20:44:37 (8,98 MB/s) - «weblogs.csv» guardado [38787883/38787883]

[hadoop@fedora ~]$ ls
dracula.txt  hadoop-3.3.4.tar.gz  hive  part-r-00000
hadoop      hadoopdata           hive-0.12.0-bin.tar.gz  weblogs.csv
[hadoop@fedora ~]$
```

Vamos a verlo con algun editor para confirmar:

```
nano weblogs.csv
```

Como podemos ver es un archivo CSV separado por comas.



```
GNU nano 6.0 weblogs.csv
2015-03-08 00:00:00,HNTG8YGROW,new_integrated_science_a04,Purchased,KRFSTI561J,>
2015-03-08 00:00:00,HNTG8YGROW,music_fun,Purchased,MKJUDF993M,15,Advances in sc>
2015-03-08 00:00:00,HNTG8YGROW,music_fun,Added to Cart,ABERKF334I,95,Sea of Pop>
2015-02-08 00:00:00,9HEX4GFUH3,science_in_dispute,Purchased,DS54EX316,91,The vo>
2015-12-08 00:00:00,ZJX5W46091,the_art_of_deception_a14,Purchased,JLTRBT354D,91>
2015-12-08 00:00:00,ZJX5W46091,texas_politics_a04,Added to Cart,BRTDFS241G,154,>
2015-07-08 00:00:00,T7DVIB2J8L,new_integrated_science_a04,Purchased,MKJUDF993M,>
2015-01-10 00:00:00,IAYIINRJT0,book_of_mint_a01,Purchased,MKJUDF993M,61,The adv>
2015-09-08 00:00:00,VS753ZSZA,history_of_political_economy,Purchased,NMQR38SJD>
2015-02-08 00:00:00,9HEX4GFUH3,the_book_lady,Purchased,JLTRBT354D,61,THE BOOK O>
2015-04-08 00:00:00,JRK07IRCIJ,women_in_australian_politics,Browsed,BTDRGL712H,>
2015-02-08 00:00:00,3C05UI3EXX,history_of_political_economy,Browsed,DS54EX316,1>
2015-03-08 00:00:00,HNTG8YGROW,eddie's_world,Added to Cart,JLTRBT354D,123,THE BO>
2015-09-08 00:00:00,Y9SPOA6IW7,music_fun,Purchased,BTDRGL712H,91,Understanding >
2015-09-08 00:00:00,Y9SPOA6IW7,new_integrated_science_a04,Added to Cart,BTDRGL7>
2015-05-08 00:00:00,OCLA2XGTB8,music_fun,Purchased,NMQR38SJD,667,Understanding>
2015-02-08 00:00:00,9HEX4GFUH3,new_integrated_science_a04,Added to Cart,ARBUER4>
2015-04-08 00:00:00,JRK07IRCIJ,women_in_australian_politics,Browsed,JLTRBT354D,>
2015-03-08 00:00:00,HNTG8YGROW,childrens_fiction_series,Purchased,DS54EX316,61,>
2015-09-08 00:00:00,VS753ZSZA,texas_politics_a04,Browsed,ARBUER437Y,154,Histor>
[ 278972 líneas leídas ]
^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
```

Vamos a iniciar nuestro servidor de Hadoop:

```
cd /home/hadoop/hadoop/sbin
```

```
./start-all.sh
```

Confirmamos en el navegador:



## DataNode on fedora:9866

Cluster ID:	CID-59aa29a5-3511-485e-8605-6d9c489ecc11
Started:	Sat Sep 24 20:52:12 -0500 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb

## Block Pools

Ahora subiremos nuestro archivo mediante el uso de DHFS, primero confirmamos el espacio en el cluster de hadoop.

```
hdfs dfs -ls /
```

Como observamos no existe una carpeta llamada /data/ así que la creamos:

```
hdfs dfs -mkdir /data
```

Dentro de Data creamos /weblogs/

```
hdfs dfs -mkdir /data/weblogs
```

Dentro de /weblogs/ subimos nuestro archivo

```
hdfs dfs -put weblogs.csv /data/weblogs
```

Tenga en cuenta que es recomendable estar en la misma dirección del archivo cuando se ejecuta este comando.

Y listo, confirmamos su subida a la carpeta del cluster de hadoop.

```
hadoop@fedora:~  
[hadoop@fedora ~]$ ls  
dracula.txt  hadoop-3.3.4.tar.gz  hive  part-r-00000  
hadoop  hadoopdata  hive-0.12.0-bin.tar.gz  weblogs.csv  
[hadoop@fedora ~]$ hdfs dfs -put weblogs.csv /data/weblogs  
2022-09-24 21:04:59,901 WARN util.NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable  
[hadoop@fedora ~]$ hdfs dfs -ls /  
2022-09-24 21:05:46,798 WARN util.NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable  
Found 5 items  
drwxr-xr-x - hadoop supergroup 0 2022-09-24 21:01 /data  
drwxr-xr-x - hadoop supergroup 0 2022-09-20 17:39 /libros  
drwxr-xr-x - hadoop supergroup 0 2022-09-20 18:09 /result_libros  
drwxr-xr-x - hadoop supergroup 0 2022-09-20 19:13 /result_libros2  
drwx----- - hadoop supergroup 0 2022-09-20 18:04 /tmp  
[hadoop@fedora ~]$ hdfs dfs -ls /data  
2022-09-24 21:05:59,910 WARN util.NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable  
Found 1 items  
drwxr-xr-x - hadoop supergroup 0 2022-09-24 21:05 /data/weblogs  
[hadoop@fedora ~]$ hdfs dfs -ls /data/weblogs  
2022-09-24 21:06:20,039 WARN util.NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable  
Found 1 items  
-rw-r--r-- 1 hadoop supergroup 38787883 2022-09-24 21:05 /data/weblogs/weblogs.csv  
[hadoop@fedora ~]$
```

Podemos hasta ver su contenido dentro del cluster.

```
hdfs dfs -cat /data/weblogs/weblogs.csv
```

```
hadoop@fedora:~  
hadoop@fedora:~  
2015-09-08 00:00:00,Y9SP0A6IW7,music_fun,Purchased,NMQR38S3DI,278,The adventures of Arthur Conan Doyle,Political,15,25.99,99554,Completed,15  
2015-12-08 00:00:00,ZJX5W46091,music_fun,Purchased,BTDRGL712H,541,Science in Dispute,Romance,74,300.95,98765,Paid,225  
2015-04-08 00:00:00,JRK07IRCIJ,women_in_australian_politics,Added to Cart,BRTDFS241G,9010,Sea of Poppies,Philosophy,1,140,967445,Issued,120  
2015-04-08 00:00:00,JRK07IRCIJ,eddie's_world,Added to Cart,ABERKF334I,667,Sea of Poppies,Drive_books,10,25,97342,Paid,120  
2015-02-08 00:00:00,9HEX4GFUH3,science_in_dispute,Added to Cart,DS54EX316,154,Understanding American politics,Cook,21,25,88734,Paid,335  
2015-07-08 00:00:00,T7DVIB2J8L,book_of_mint_a01,Purchased,DS54EX316,667,History of political economy,Drama,39,15,967445,Requested,299.99  
2015-12-08 00:00:00,ZJX5W46091,book_of_mint_a01,Added to Cart,ABERKF334I,123,The voyages of Captain Cook,World_History,9,15,99568,Failed,9.99  
2015-09-08 00:00:00,VS753ZSZJA,new_integrated_science_a04,Browsed,ABERKF334I,154,Science in Dispute,Fiction,18,5.99,97342,Issued,149.99  
2015-04-08 00:00:00,JRK07IRCIJ,women_in_australian_politics,Browsed,BTDRGL712H,123,The voyages of Captain Cook,Cook,39,300.95,99456,Cancelled,225  
2015-01-10 00:00:00,IAYIINRJT0,the_book_lady,Purchased,ARBUER437Y,61,The adventures of Arthur Conan Doyle,Political,5,5.99,98453,Requested,15  
2015-05-08 00:00:00,OCLA2XGTB8,the_book_lady,Added to Cart,ABERKF334I,123,The adventures of Arthur Conan Doyle,Drive_books,15,25.99,967445,Cancelled,625  
2015-04-08 00:00:00,JRK07IRCIJ,music_fun,Purchased,ARBUER437Y,278,Advances in school psychology,Music,5,225,98453,Issued,325.39  
2015-09-08 00:00:00,VS753ZSZJA,women_in_australian_politics,Purchased,MKJUDF993M,667,History of political economy,Cook,18,305,98765,Issued,225  
2015-09-08 00:00:00,VS753ZSZJA,the_book_lady,Browsed,KRFSTI561J,9010,The adventures of Arthur Conan Doyle,Management,11,5.99,97342,Paid,125  
2015-09-08 00:00:00,VS753ZSZJA,science_in_dispute,Purchased,ARBUER437Y,123,The voyages of Captain Cook,Philosophy,61,2.99,87556,Paid,299.99  
2015-09-08 00:00:00,Y9SP0A6IW7,new_integrated_science_a04,Purchased,JLTRBT354D,10,History of political economy,Management,21,25,99568,Cancelled,335  
2015-09-08 00:00:00,VS753ZSZJA,women_in_australian_politics,Added to Cart,ARBUER437Y,61,Advances in school psychology,Fiction,11,2.99,99554,Paid,225  
2015-04-08 00:00:00,JRK07IRCIJ,the_art_of_deception_a14,Added to Cart,BRTDFS241G,541,Understanding American politics,Music,61,300.95,88734,Cancelled,9.99  
2015-04-08 00:00:00,JRK07IRCIJ,texas_politics_a04,Purchased,DGHRIT776Q,10,Understanding American politics,Religion,1,5.99,98765,Paid,75  
2015-05-08 00:00:00,OCLA2XGTB8,women_in_australian_politics,Browsed,ARBUER437Y,10,Advances in school psychology,World_History,61,25,99554,Paid,225  
2015-07-08 00:00:00,T7DVIB2J8L,eddie's_world,Added to Cart,BTDRGL712H,278,The voyages of Captain Cook,Science,21,45.25,967445,Cancelled,335  
2015-09-08 00:00:00,Y9SP0A6IW7,new_integrated_science_a04,Purchased,JLTRBT354D,541,The voyages of Captain Cook,Science,11,25.99,967445,Paid,225  
2015-02-08 00:00:00,9HEX4GFUH3,music_fun,Added to Cart,BRTDFS241G,322,Space fact and fiction,Art,10,45.25,98453,Failed,149.99  
2015-09-08 00:00:00,Y9SP0A6IW7,science_in_dispute,Browsed,DS54EX316,107,History of political economy,Political,9,25,97342,Requested,125  
2015-02-08 00:00:00,9HEX4GFUH3,childrens_fiction_series,Browsed,BTDRGL712H,107,Advances in school psychology,Non_Fiction,15,140,95567,Requested,299.99  
2015-12-08 00:00:00,ZJX5W46091,the_art_of_deception_a14,Purchased,DGHRIT776Q,9010,Space fact and fiction,Programming,1,5.99,98453,Cancelled,75  
[hadoop@fedora ~]$
```

Para usar Hive basta con utilizar el siguiente comando:

```
cd /home/hadoop/hive/bin
```

La ruta de acceso será relativa al sitio de instalación de hive y su nombre de usuario.

```
./hive
```

Lo siguiente es hecho en Apache Ambari Hive View. La guía la puedes ver en

<https://learn.microsoft.com/es-es/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view>

Lo primero será acceder a nuestra cuenta de Azure con el siguiente enlace:

<https://portal.azure.com/>




portal.azure.com/#home

Microsoft Azure

Buscar recursos, servicios y documentos (G+/)


Le damos la bienvenida a Azure

¿No tiene ninguna suscripción? Consulte las opciones siguientes.




**Empiece con una prueba gratuita de Azure**  
Obtenga un crédito gratuito de 200 USD para los productos y servicios de Azure, más 12 meses de **servicios gratuitos** populares.

[Inicio](#)



**Administrar Azure Active Directory**  
Administre el acceso, establezca directivas inteligentes y mejore la seguridad con Azure Active Directory.

[Ver](#) [Más información](#)



**Acceder a las ventajas para alumnos**  
Obtenga software gratuito o crédito de Azure, o acceda a Azure Dev Tools for Teaching después de verificar su estado académico.

[Explorar](#) [Más información](#)

Servicios de Azure

[Crear un](#)

[Clústeres de](#)

[Todos los](#)

[Centro de](#)

[Máquinas](#)

[App Services](#)

[Cuentas de](#)

[SQL Database](#)

[Azure](#)

[Más servicios](#)

[https://portal.azure.com/#blade/Microsoft\\_Azure\\_Billing/FreeServicesBlade](https://portal.azure.com/#blade/Microsoft_Azure_Billing/FreeServicesBlade)



Luego podemos consultar la enumeración y visualización de clústeres para poder ver las instrucciones de la creación de un cluster, esto en el enlace: <https://learn.microsoft.com/es-es/azure/hdinsight/hdinsight-administer-use-portal-linux#showClusters>


Microsoft Azure

Buscar recursos, servicios y documentos (G+/)

Inicio >


Le damos la bienvenida a Azure

¿No tiene ninguna suscripción? Consulte las opciones siguientes.



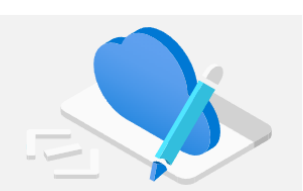
**Empiece con una prueba gratuita de Azure**  
Obtenga un crédito gratuito de 200 USD para los productos y servicios de Azure, más 12 meses de **servicios gratuitos** populares.

[Inicio](#)



**Administrar Azure Active Directory**  
Administre el acceso, establezca directivas inteligentes y mejore la seguridad con Azure Active Directory.

[Ver](#) [Más información](#)



**Acceder a las ventajas para alumnos**  
Obtenga software gratuito o crédito de Azure, o acceda a Azure Dev Tools for Teaching después de verificar su estado académico.

[Explorar](#) [Más información](#)

Nota: Se recomienda acceder a las ventajas para alumnos.

The screenshot shows the Microsoft Azure Education portal. The header includes the Microsoft Azure logo and a search bar. The main content area is titled 'Education | Información general'. On the left, there is a sidebar with 'Recursos de aprendizaje' (Learning resources) including Roles, Software, Aprendizaje (Learning), and Plantillas (Templates). Below this is a link for '¿Necesita ayuda?' (Need help?) and 'Soporte técnico' (Technical support). The main content area is divided into several sections: 'Detalles de la oferta para estudiantes' (Student offer details) showing 100 US\$ of 100 US\$ in credits, 365 days until expiration, and 0.00 US\$ for September costs; 'Soluciones populares' (Popular solutions) with links for Docker, Node.js, Machine Learning, and web site deployment; 'Servicios gratuitos' (Free services) including Azure Virtual Machines; and 'Software libre' (Open source software) including SQL Server 2019 Developer.

*NOTA: El ejercicio de los compañeros fue realizado en este ambiente posiblemente porque se llegó al mismo estado de instalación que el presentado en esta guía, por tanto lo que vera a continuación es la preparación de Apache Hive por Azure.*


The screenshot shows the Microsoft Azure HDInsight clusters page. The header includes the Microsoft Azure logo and a search bar. The main content area is titled 'Clústeres de HDInsight'. Below the title, there is a message indicating an error: 'Se produjo un error al intentar capturar recursos. Detalles adicionales de la API subyacente que pueden resultar útiles: Proporcionar la información siguiente al pedir soporte: timestamp = 2022-09-25T04:03:07.4750897Z, correlationId = 4b9d22a6-a262-4696-8346-e3f059e73b39. (Código: AccessDenied) Se denegó el acceso al recurso solicitado. Puede que el usuario no tenga permisos suficientes. (Código: AccessDenied). Pruebe a actualizar la página. Los recursos no se ven afectados por el problema, simplemente no se puede mostrar esta vista en este momento. Para ver una lista de recursos, seleccione Vista simplificada.' Below the error message, there is a table of clusters with columns: Nombre, Tipo de clúster, Estado, Grupo de recursos, Ubicación, and Versión de clúster. The table is currently empty, and a message states 'No hay Clústeres de HDInsight para mostrar'. Below the message, there is a link to 'Crear Clúster de HDInsight'.


Desde Paneles de clúster, seleccione Vistas de Ambari. Cuando se le solicite autenticarse, use el nombre de cuenta y la contraseña de inicio de sesión del clúster (el valor predeterminado es `admin`) que proporcionó al crear el clúster. También puede ir a <https://CLUSTERNAME.azurehdinsight.net/#/main/views> en el explorador, donde CLUSTERNAME es el nombre del clúster.


En la lista de vistas se debe seleccionar la VISTA DE HIVE.



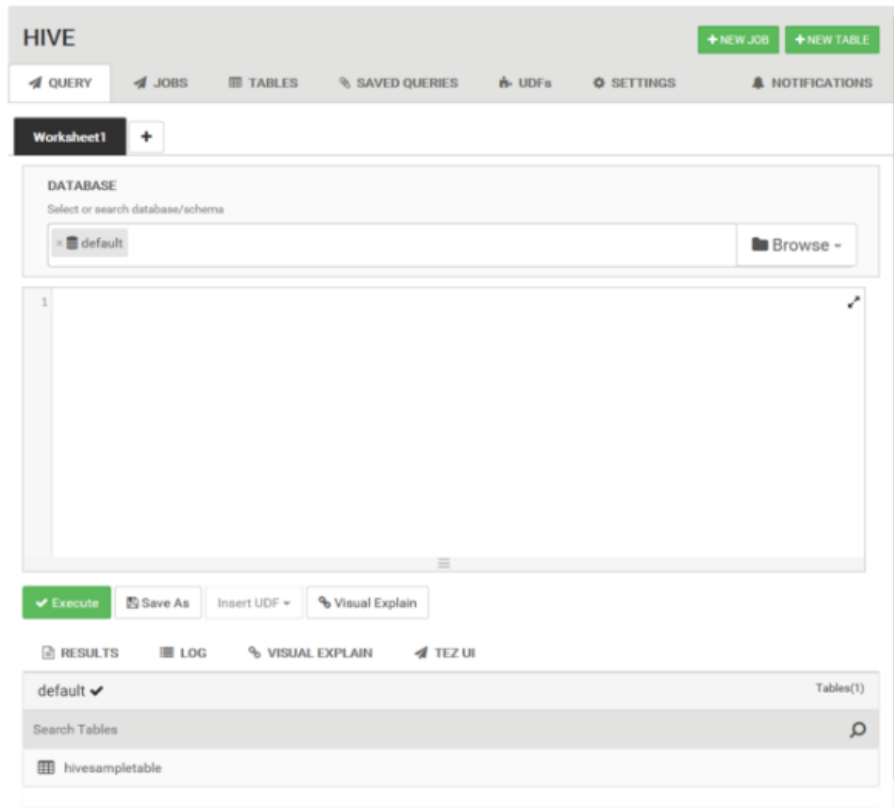
## Your Views

 **YARN Queue Manager** (1.0.0)  
Manage YARN Capacity Scheduler Queues

 **Hive View 2.0** (2.0.0)  
This view instance is auto created when the Hive service is added to a cluster.

 **Tez View** (0.7.0.2.6.0.10-20)  
Monitor and debug all Tez jobs, submitted by Hive queries and Pig scripts (auto-created)

La pagina de vista de Hive sera similar a la presentada en la siguiente imagen.



En la versión estudiantil de servicios de Azure no podemos subir bases de datos CSV pero podemos hacer consultas de las ya existentes.

