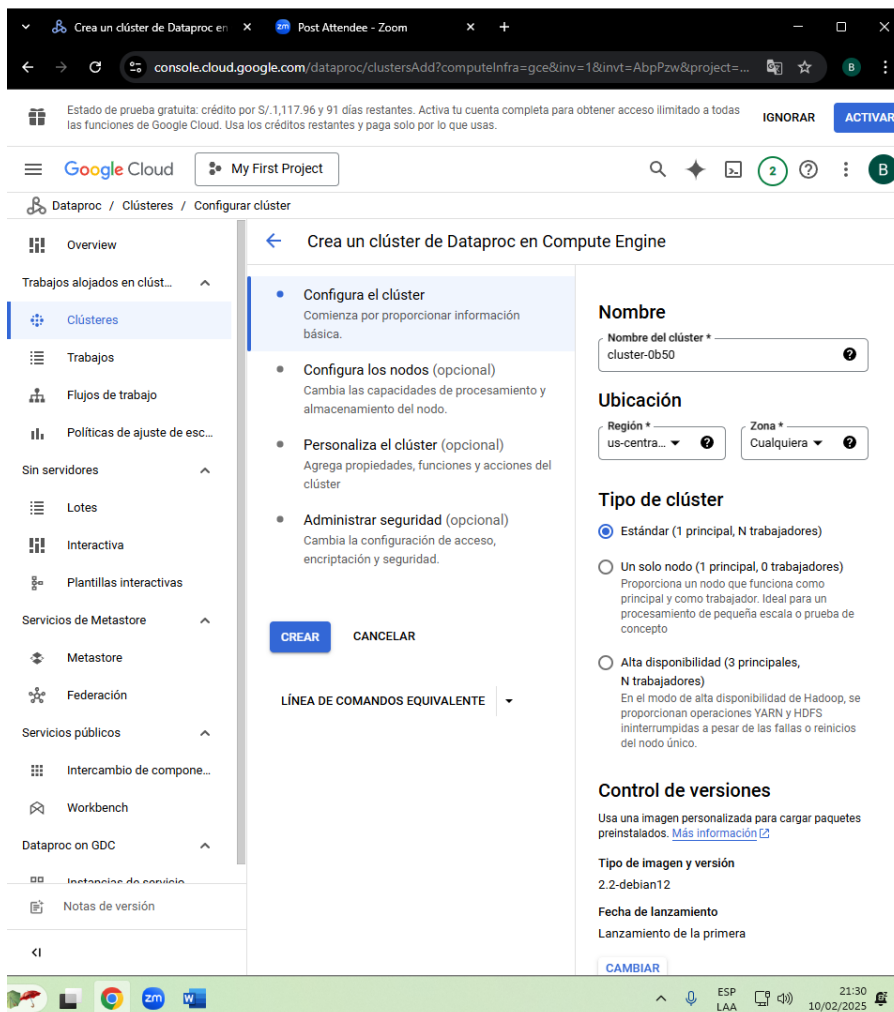
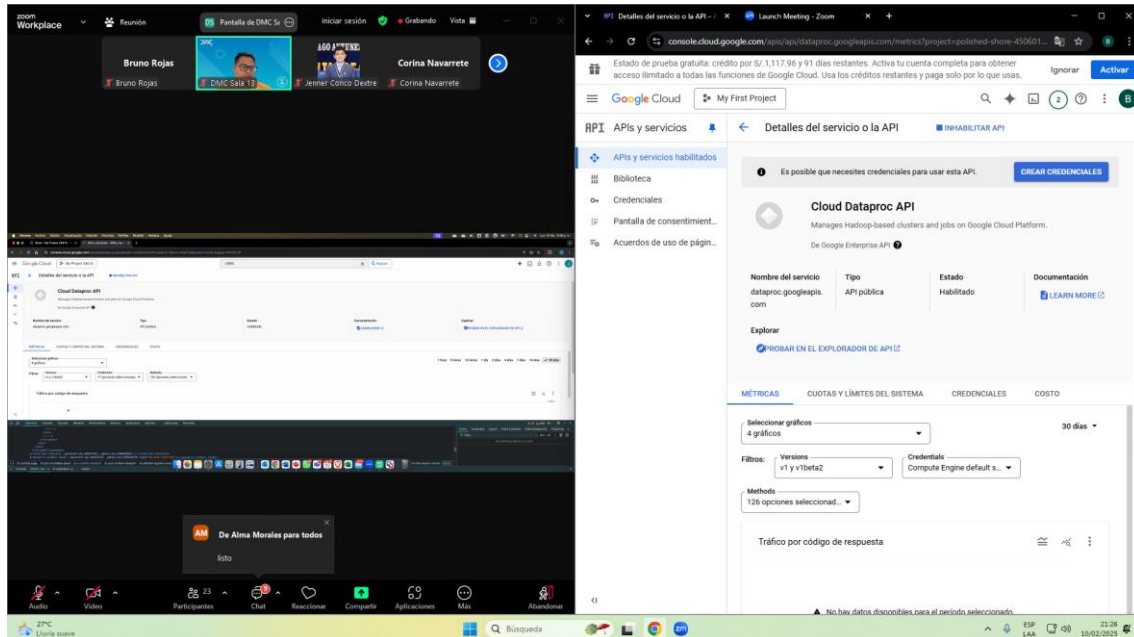


## **Trabajo Especialización Big Data**

<https://github.com/BrunoRojasB/curso-dmc-big-data-processing>

# 1.- HDFS

## Creación de Clúster Dataproc

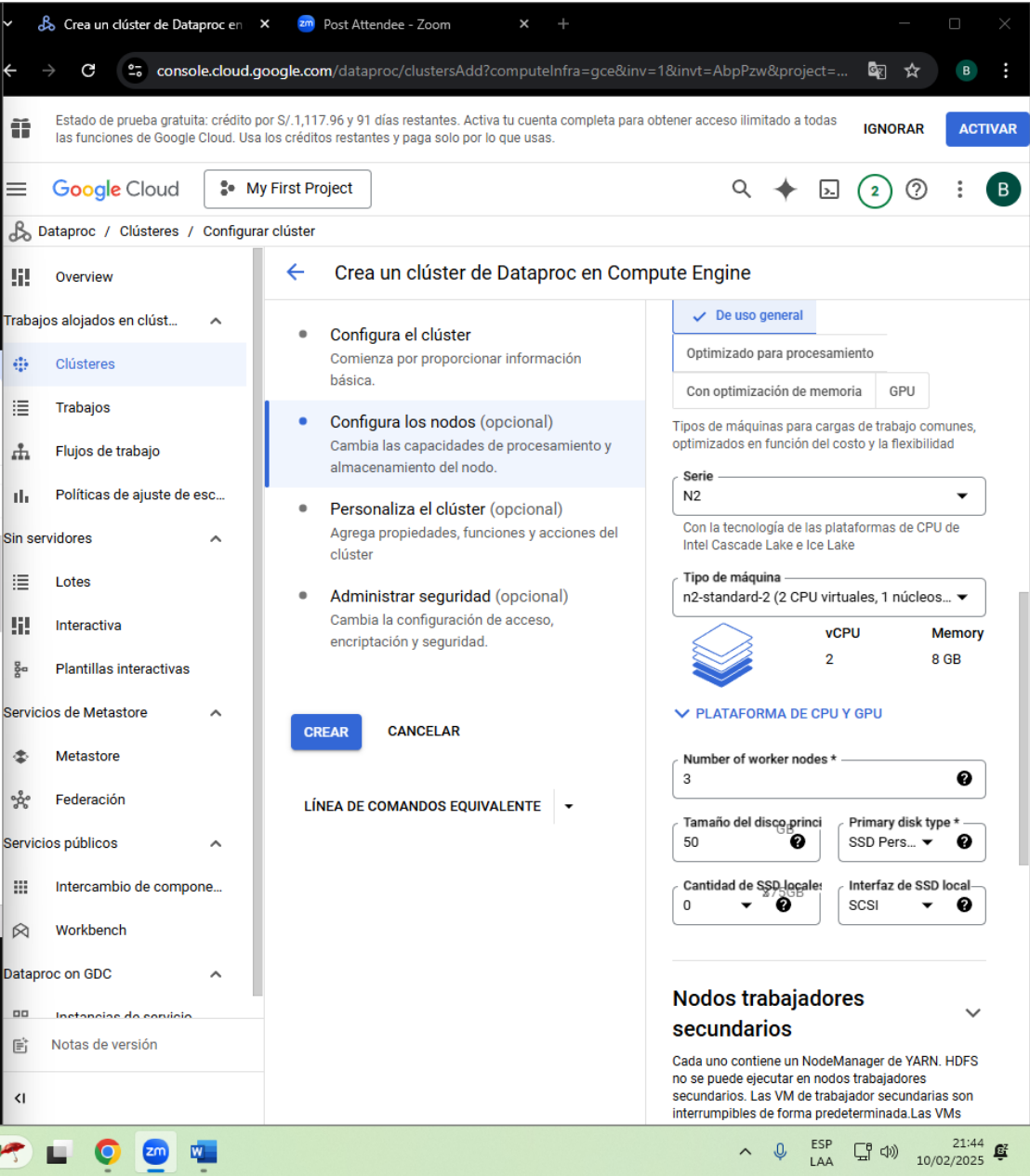


Configuración de Nodos

Tipo de Máquina

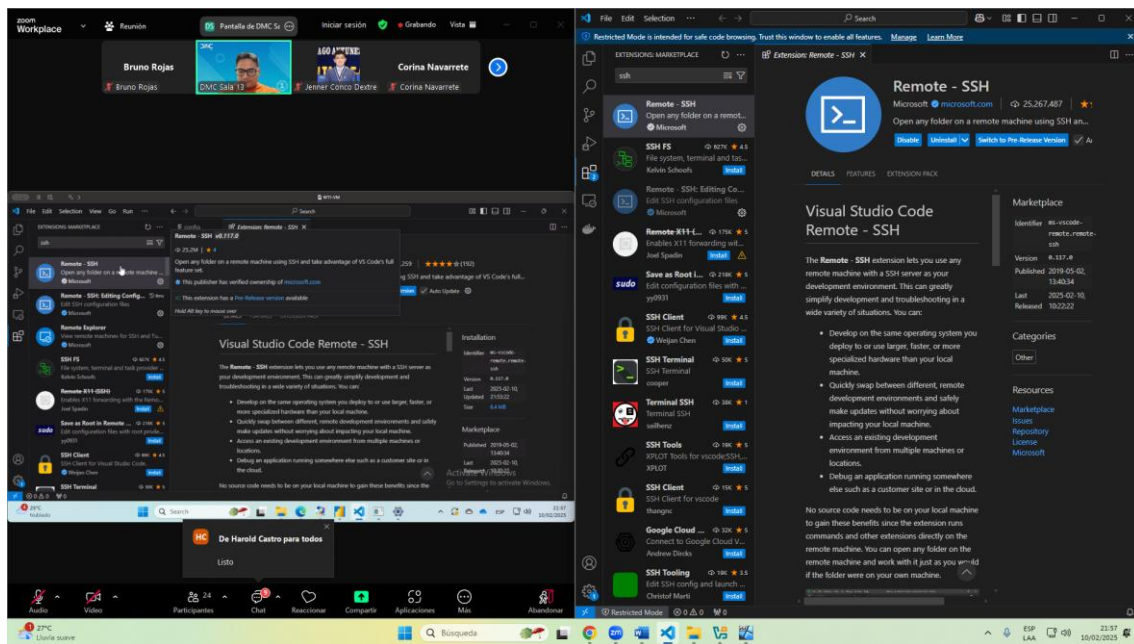
Número de nodos trabajadores

Espacio de disco

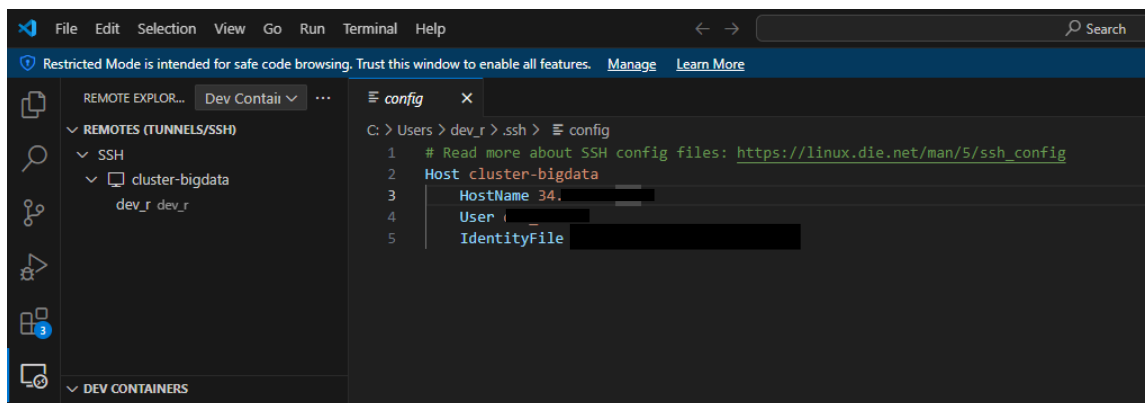


Clústeres								
CREATE CLUSTER ACTUALIZAR INICIAR DETENER BORRAR REGIONES + 5 ALERTAS RECOMENDADAS								
Filtro Busca el clúster por propiedades y presiona Intro								
El servidor no pudo completar tu solicitud.								
	Nombre ↑	Estado	Región	Zona	Total de nodos trabajadores	¿Tiene VMs flexibles?	Eliminación programada	Bucket de etapa de pruebas de Cloud-
	dmc-dev-bdp-15	En ejecución	us-central1	us-central1-f	3	No	Desactivado	dataproc-staging-us-central1-5902-esta1ax8

## Instalación de Remote Explorer en Visual Studio Code



## Configuración de Remote Explorer



Generación de clave SSH .

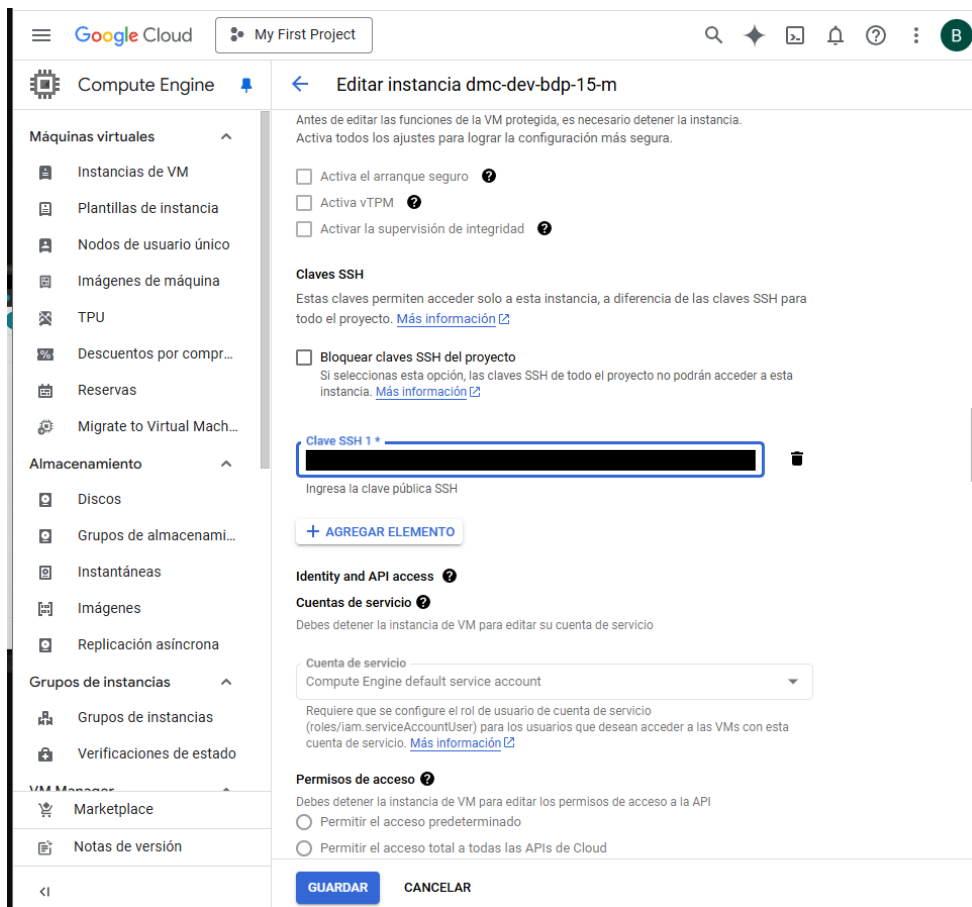
```
Windows PowerShell
Copyright (C) Microsoft Corporation. Todos los derechos reservados.

Instale la versión más reciente de PowerShell para obtener nuevas características y mejoras. https://aka.ms/PSWindows

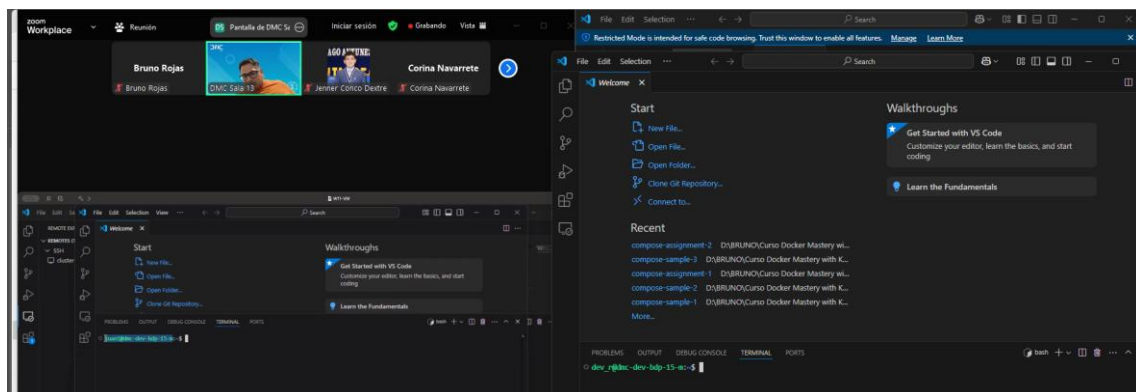
PS C:\> ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (C:\Users\user\ssh):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in C:\Users\user\ssh\id_rsa
Your public key has been saved in C:\Users\user\ssh\id_rsa.pub
The key fingerprint is:
SHA256:
The key's randomart image is:
+---[RSA 3072]-----+
|
|
|
|
|
|
|
|
|
|
+---[SHA256]-----+
```

Mode	LastWriteTime		Length	Name
----	-----		-----	----
-a----	10/02/2025	22:10	2610	id_rsa
-a----	10/02/2025	22:10	576	id_rsa.pub
-a----	2/12/2023	19:20	840	known_hosts
-a----	2/12/2023	18:45	96	known_hosts.old

## Agregar clave SSH a Dataproc



## Nos conectamos al cluster desde Visual Studio Code con Remote Explorer



# Convertir IP Efímera en Estática o Permanente

Google Cloud

My First Project

2

Compute Engine

Máquinas virtuales

Instancias de VM

Plantillas de instancia

Nodos de usuario único

Imágenes de máquina

TPU

Descuentos por compr...

Reservas

Migrate to Virtual Mach...

Almacenamiento

dmc-dev-bdp-...

EDITAR

CÓDIGO EQUIVALENTE

DETALLES

OBSERVABILIDAD

INFORMACIÓN DEL SO

CAPTURA

Tráfico HTTP

Inactivo

Tráfico HTTPS

Inactivo

Permitir las verificaciones de estado del balanceador de cargas

Inactivo

Etiquetas de red

dataproc-notebook-vm

Interfaces de red

principal

Rangos de alias de IP

Tipo de pila de IP

Dirección IP externa

IPv4

ipservermaster

Google Cloud

My First Project

dataproc

Search

Dataproc / Clusters / Cluster: dmc-dev-bdp-15 / VM instances

Overview

Jobs on Clusters

Clusters

Jobs

Workflows

Autoscaling policies

Serverless

Batches

Interactive

Interactive Templates

Metastore Services

Metastore

Federation

Utilities

Cluster details

SUBMIT JOB

REFRESH

START

STOP

DELETE

VIEW LOGS

Failed to validate permissions required for default service account: '590202858150-compute@developer.serviceaccount.com'. Cluster creation could still be successful if required permissions have been document https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/service-accounts#dataproc\_service\_accounts\_2. This could be due to Cloud Resource Manager API hasn't been enabled it by visiting 'https://console.developers.google.com/apis/api/cloudresourcemanager.googleapis.com/overview?project=590202858150'.

MORE

Name

dmc-dev-bdp-15

Cluster UUID

9f0fd831-71f6-41ec-aed7-a3b5d44e3720

Type

Dataproc Cluster

Status

Running

MONITORING

JOB

VM INSTANCES

CONFIGURATION

WEB INTERFACES

Filter

Filter instances

Name

Role

SSH

dmc-dev-bdp-15-m

Master

SSH

dmc-dev-bdp-15-w-0

Worker

dmc-dev-bdp-15-w-1

Worker

dmc-dev-bdp-15-w-2

Worker

Entramos al master y bajamos hasta network interfaces

Google Cloud console showing the details of a VM instance named `dmc-dev-bdp-15-m`. The left sidebar shows the navigation menu with categories like Virtual machines, Storage, Instance groups, VM Manager, Bare Metal Solution, and Settings. The main content area displays the instance details, including the network interfaces and storage configuration.

**Network interfaces**

Name	Network	Subnetwork	Primary internal IP address	Alias IP ranges	IP stack type	External IP address	Network
nic0	default	default	10.128.0.4		IPv4	(34.42.122.86)	Premi

**Storage**

**Boot disk**

Name	Image	Interface type	Size (GB)	Device name	Type	Architecture	Encryption	Mode	Wh
dmc-dev-bdp-15-m	dataproc-2-2-deb12-20250130-035100-rc01	SCSI	50	persistent-disk-0	SSD persistent disk	—	Google-managed	Boot, read/write	Del

La ip que normalmente aparece como ephemeral o efímera

Click derecho sobre nic0 -> Abrir en una nueva ventana

Google Cloud console showing the details of a network interface named `nic0`. The left sidebar shows the navigation menu with categories like VPC Network, IP addresses, Internal ranges, Bring your own IP, Firewall, Routes, VPC network peering, Shared VPC, Serverless VPC access, Packet mirroring, and VPC Flow Logs. The main content area displays the network interface details, including the VM instance details and firewall and routes details.

**Network interface details**

Name	Network	Subnetwork	Primary internal IP address	Alias IP ranges	IP stack type	External IP address	Network Service Tier
nic0	default	default	10.128.0.4	—	IPv4		Premium

**VM instance details**

Name	Zone	Network tags	Service account	IP forwarding
dmc-dev-bdp-15-m	us-central1-f	dataproc-notebook-vm	590202858150-compute@developer.gserviceaccount.com	Off

**Firewall and routes details**

[FIREWALLS](#) [ROUTES](#)



Click IP addresses

The screenshot shows the Google Cloud console interface for IP addresses. The left sidebar lists various VPC network components, with 'IP addresses' selected. The main panel displays a table of IP addresses. The first row, representing the 'ipservermaster' IP (34.42.122.86), is highlighted with a red box. The table includes columns for Name, IP address, Access type, Region, Type, Version, In use by, Subnet, and Actions.

Name	IP address	Access type	Region	Type	Version	In use by	Subnet	Actions
ipservermaster	34.42.122.86	External	us-central1	Static	IPv4	VM instance <a href="#">dmc-dev-bdp-15-m</a> (Zone us-central1-f)		
-	10.128.0.2	Internal	us-central1	Ephemeral	IPv4	VM instance <a href="#">dmc-dev-bdp-15-w-0</a> (Zone us-central1-f)	default	
-	10.128.0.3	Internal	us-central1	Ephemeral	IPv4	VM instance <a href="#">dmc-dev-bdp-15-w-1</a> (Zone us-central1-f)	default	
-	10.128.0.4	Internal	us-central1	Ephemeral	IPv4	VM instance <a href="#">dmc-dev-bdp-15-m</a> (Zone us-central1-f)	default	
-	10.128.0.5	Internal	us-central1	Ephemeral	IPv4	VM instance <a href="#">dmc-dev-bdp-15-w-2</a> (Zone us-	default	

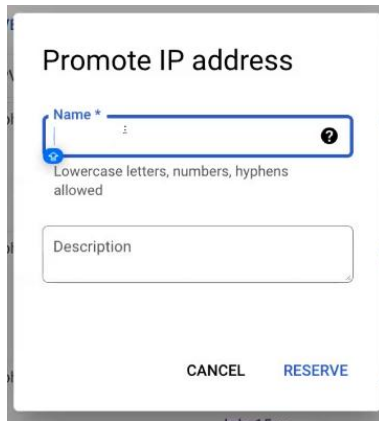
Click en los tres puntitos de la IP que está usando nuestro servidor y dar click en Promote to Static IP Address

This screenshot shows the same IP address table as the previous image. The 'Actions' column for the 'ipservermaster' IP (34.42.122.86) is expanded, revealing a dropdown menu. The option 'Promote to static IP address' is highlighted with a red circle. Other options visible include 'Reassign to another resource' and 'No addresses'.

Debería aparecer promover a ip estática

A close-up view of the 'Promote to static IP address' button, which is circled in red. The button is located in the 'Actions' dropdown menu for the 'ipservermaster' IP.

Le colocamos un nombre y descripción



Promote IP address

Name \*

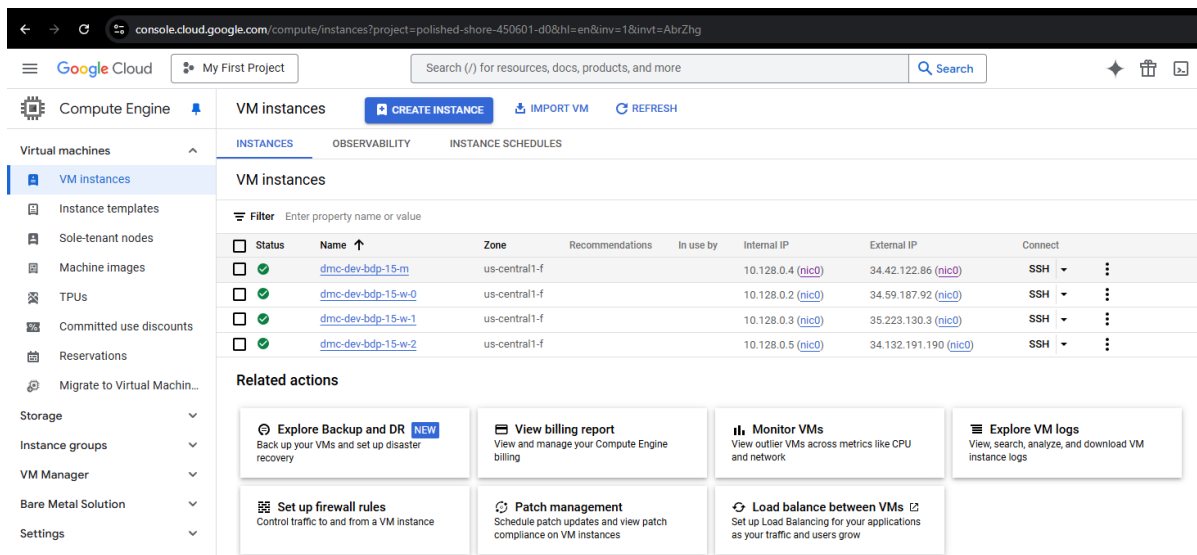
Lowercase letters, numbers, hyphens allowed

Description

CANCEL RESERVE

Luego la ip cambiaría de efímera a estática y es la que usaríamos siempre para conectarnos.

Para añadir una llave ssh



console.cloud.google.com/compute/instances?project=polished-shore-450601-d0&hl=en&inv=1&inv=1&inv=1&inv=1

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Compute Engine VM instances CREATE INSTANCE IMPORT VM REFRESH

Virtual machines VM instances Instance templates Sole-tenant nodes Machine images TPUs Committed use discounts Reservations Migrate to Virtual Machin...

VM instances INSTANCES OBSERVABILITY INSTANCE SCHEDULES

Filter Enter property name or value

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input type="checkbox"/>	<a href="#">dmc-dev-bdp-15-m</a>	us-central1-f			10.128.0.4 (nic0)	34.42.122.86 (nic0)	SSH
<input type="checkbox"/>	<a href="#">dmc-dev-bdp-15-w-0</a>	us-central1-f			10.128.0.2 (nic0)	34.59.187.92 (nic0)	SSH
<input type="checkbox"/>	<a href="#">dmc-dev-bdp-15-w-1</a>	us-central1-f			10.128.0.3 (nic0)	35.223.130.3 (nic0)	SSH
<input type="checkbox"/>	<a href="#">dmc-dev-bdp-15-w-2</a>	us-central1-f			10.128.0.5 (nic0)	34.132.191.190 (nic0)	SSH

Related actions

- Explore Backup and DR NEW Back up your VMs and set up disaster recovery
- View billing report View and manage your Compute Engine billing
- Monitor VMs View outlier VMs across metrics like CPU and network
- Explore VM logs View, search, analyze, and download VM instance logs
- Set up firewall rules Control traffic to and from a VM instance
- Patch management Schedule patch updates and view patch compliance on VM instances
- Load balance between VMs Set up Load Balancing for your applications as your traffic and users grow

Entramos al servidor maestro dmc-dev-bdp-15-m

**Google Cloud** | My First Project | Search (/) for resources, docs, products, and more

---

## Compute Engine

- Virtual machines
  - VM instances**
  - Instance templates
  - Sole-tenant nodes
  - Machine images
  - TPUs
  - Committed use discounts
  - Reservations
  - Migrate to Virtual Machine...
- Storage
- Instance groups
- VM Manager
- Bare Metal Solution
- Settings

[dmc-dev-bdp-15-m - Compute](#)
[IP addresses - VPC Network](#)
[WhatsApp](#)

console.cloud.google.com/compute/instancesDetail/zones/us-central1-f/instances/dmc-dev-bdp-15-m?hl=en&invv=1&invt=AbrZhg&project=polished-shore-450601-d0

dmc-dev-bdp-... [EDIT] [RESET] [CREATE MACHINE IMAGE] [CREATE SIMILAR] ▶ START / RESUME

**DETAILS** OBSERVABILITY OS INFO SCREENSHOT

[SSH] [CONNECT TO SERIAL CONSOLE]

Connecting to serial ports is disabled ⓘ

### Logs

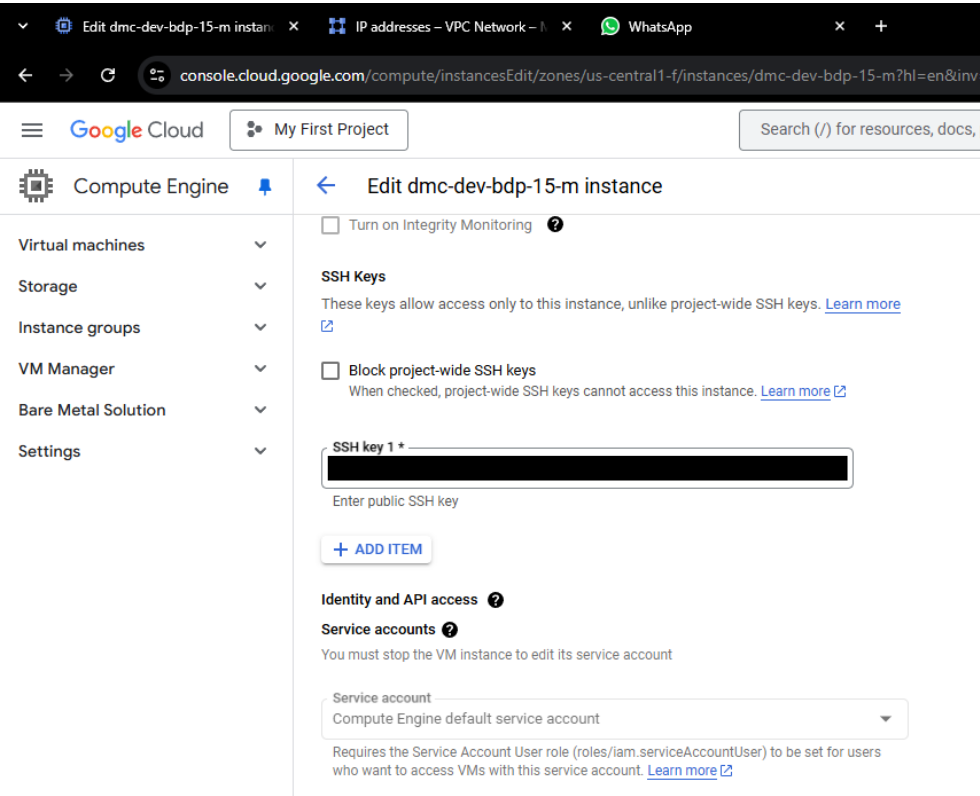
[Logging](#)  
[Serial port 1 \(console\)](#)  
[SHOW MORE](#)

### Basic information

Name	dmc-dev-bdp-15-m
Instance id	6269358924647302186
Description	None
Type	Instance
Status	Running
Creation time	Feb 10, 2025, 9:49:44 PM UTC-05:00
Location ⓘ	us-central1-f
Instance template	None
In use by	None
Physical host ⓘ	None
Maintenance status ⓘ	-
Reservations	Automatically choose (default)
Labels	goog-datap... : enabled   goog-datap... : dmc-dev-bd...   goog-datap... : 9f0fd831-7...   goog-datap... : us-central1
Tags ⓘ	-
Deletion protection	Disabled
Confidential VM service ⓘ	Disabled
Preserved state size	0 GB

Marketplace

Damos click en edit. Bajamos hasta ssh keys. Y damos click en add item y se añade la llave pública.



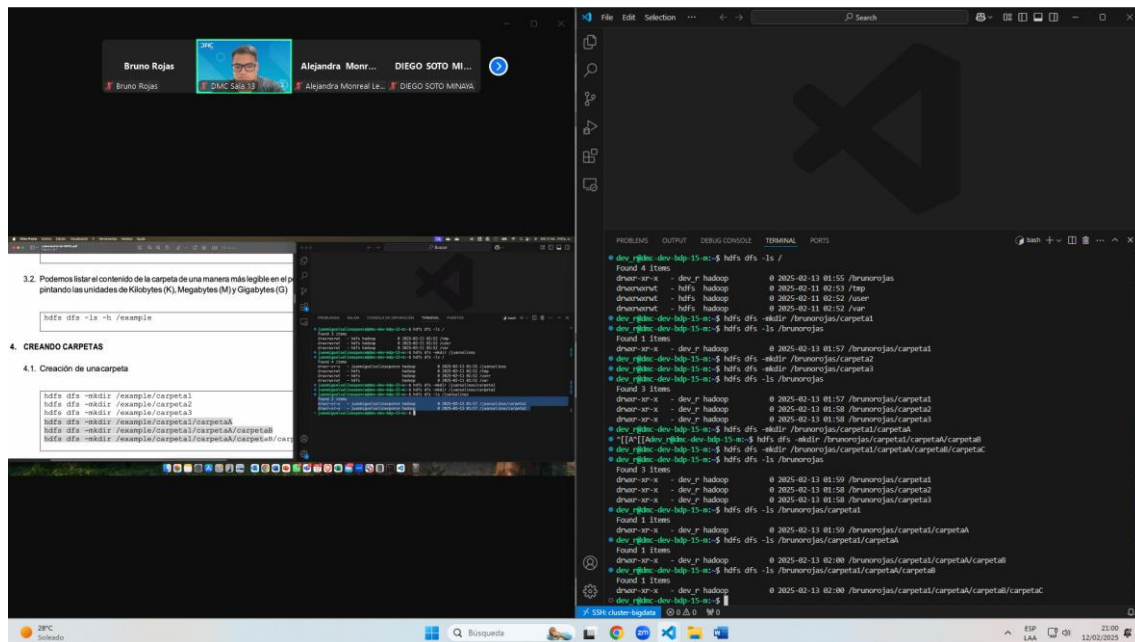
## Comandos en HDFS

Listar carpetas en HDFS: `hdfs dfs -ls /`

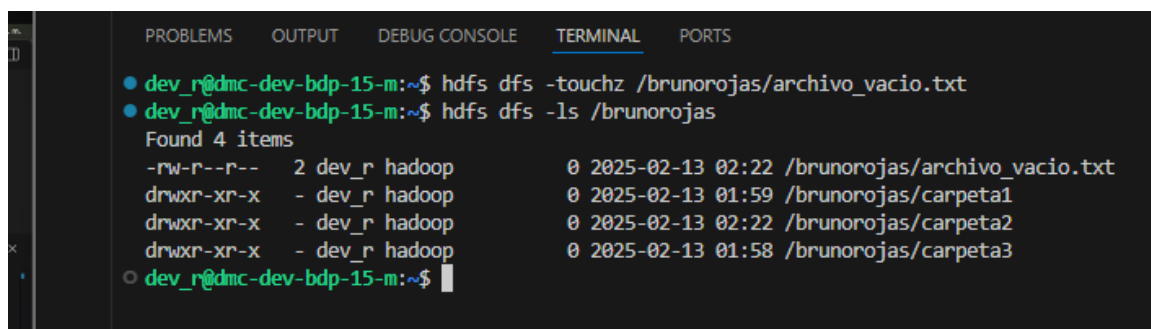
Crear directorios en HDFS: `hdfs dfs -mkdir /brunorojas`

Crear una carpeta dentro de otra: `hdfs dfs -mkdir /brunorojas/carpeta2`

Listar el contenido de la carpeta: `hdfs dfs -ls /brunorojas`



Crear archivo vacío: `hdfs dfs -touchz /brunorojas/archivovacio.txt`



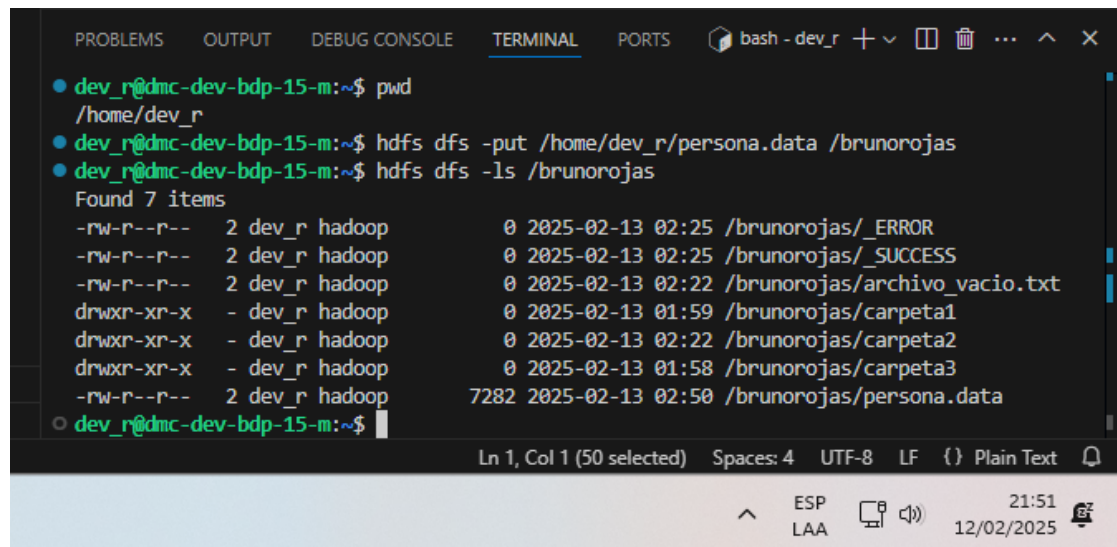
Creamos un archivo txt dentro del servidor para ir guardando nuestros comandos.

Para llevar un archivo de Linux a hdf se usa el comando put

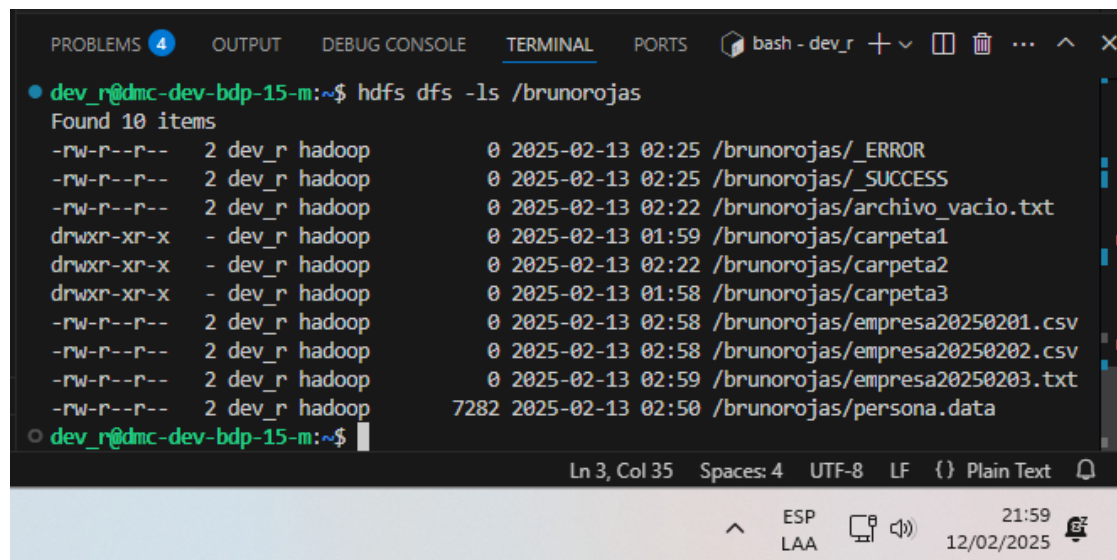
```
hdfs dfs -put /home/dev_r/persona.data /brunorojas
```

```
hdfs dfs -put /home/dev_r/empresa*.csv /brunorojas
```

```
hdfs dfs -put /home/dev_r/empresa* /brunorojas
```



```
dev_r@dmc-dev-bdp-15-m:~$ pwd
/home/dev_r
dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -put /home/dev_r/persona.data /brunorojas
dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -ls /brunorojas
Found 7 items
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:25 /brunorojas/_ERROR
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:25 /brunorojas/_SUCCESS
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:22 /brunorojas/archivo_vacio.txt
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 01:59 /brunorojas/carpeta1
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 02:22 /brunorojas/carpeta2
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 01:58 /brunorojas/carpeta3
-rw-r--r--  2 dev_r hadoop    7282 2025-02-13 02:50 /brunorojas/persona.data
dev_r@dmc-dev-bdp-15-m:~$
```



```
dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -ls /brunorojas
Found 10 items
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:25 /brunorojas/_ERROR
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:25 /brunorojas/_SUCCESS
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:22 /brunorojas/archivo_vacio.txt
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 01:59 /brunorojas/carpeta1
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 02:22 /brunorojas/carpeta2
drwxr-xr-x  - dev_r hadoop          0 2025-02-13 01:58 /brunorojas/carpeta3
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:58 /brunorojas/empresa20250201.csv
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:58 /brunorojas/empresa20250202.csv
-rw-r--r--  2 dev_r hadoop          0 2025-02-13 02:59 /brunorojas/empresa20250203.txt
-rw-r--r--  2 dev_r hadoop    7282 2025-02-13 02:50 /brunorojas/persona.data
dev_r@dmc-dev-bdp-15-m:~$
```

Usamos el comando cat para leer el contenido de un archivo.

hdfs dfs -cat /brunorojas/persona.data

Usamos el commando rm para eliminar un archivo. La f es para forzar el borrado.

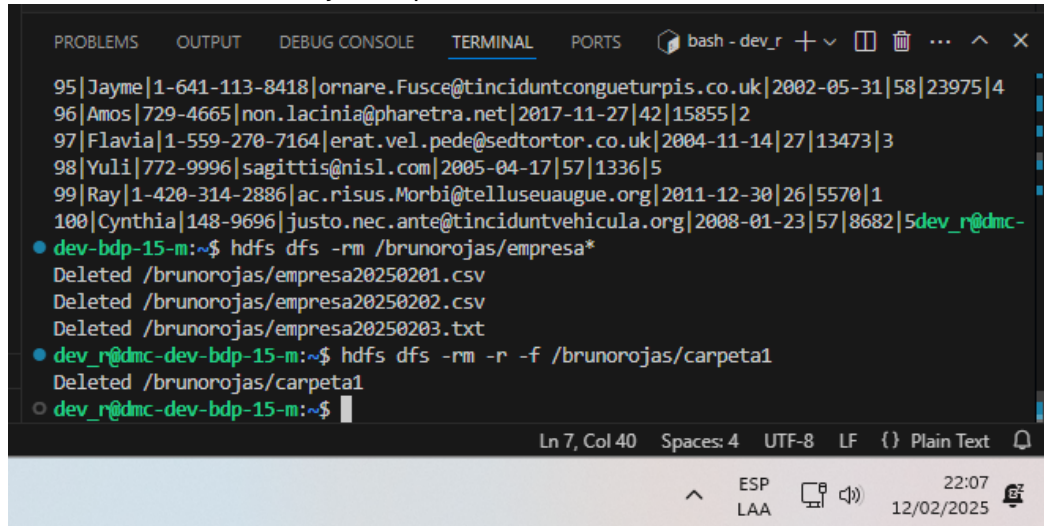
hdfs dfs -rm -f /brunorojas/persona.data

Se puede eliminar usando comodines también.

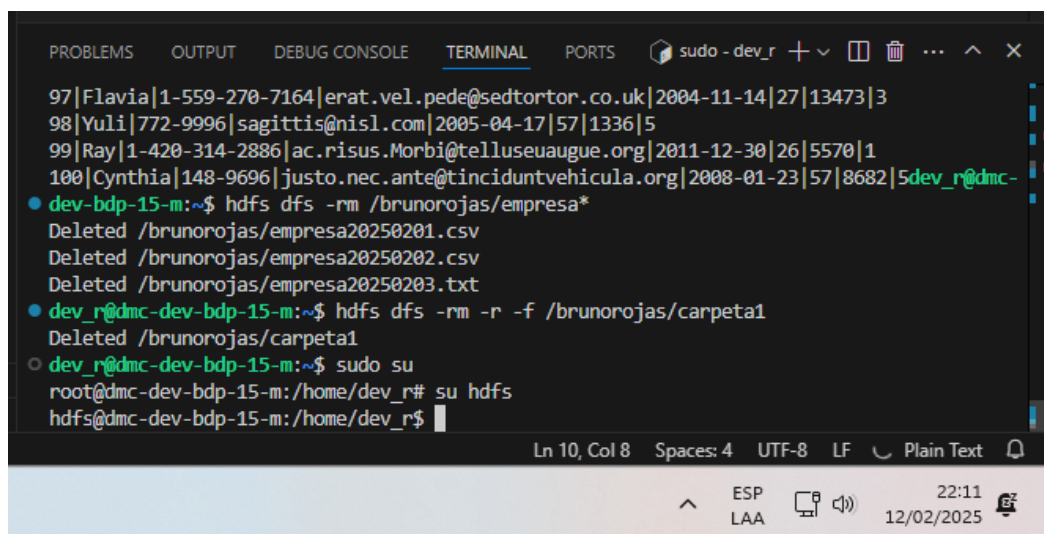
hdfs dfs -rm /brunorojas/empresa\*

Si se quiere eliminar de forma recursiva es decir los archivos o carpetas dentro de una carpeta se usa -r

hdfs dfs -rm -r -f /brunorojas/carpeta1



```
95|Jayme|1-641-113-8418|ornare.Fusce@tinciduntcongueturpis.co.uk|2002-05-31|58|23975|4
96|Amos|729-4665|non.lacinia@pharetra.net|2017-11-27|42|15855|2
97|Flavia|1-559-270-7164|erat.vel.pede@sedtortor.co.uk|2004-11-14|27|13473|3
98|Yuli|772-9996|sagittis@nisl.com|2005-04-17|57|1336|5
99|Ray|1-420-314-2886|ac.risus.Morbi@telluseuauge.org|2011-12-30|26|5570|1
100|Cynthia|148-9696|justo.nec.ante@tinciduntvehicula.org|2008-01-23|57|8682|5dev_r@dmc-
● dev-bdp-15-m:~$ hdfs dfs -rm /brunorojas/empresa*
Deleted /brunorojas/empresa20250201.csv
Deleted /brunorojas/empresa20250202.csv
Deleted /brunorojas/empresa20250203.txt
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -rm -r -f /brunorojas/carpeta1
Deleted /brunorojas/carpeta1
○ dev_r@dmc-dev-bdp-15-m:~$
```



```
97|Flavia|1-559-270-7164|erat.vel.pede@sedtortor.co.uk|2004-11-14|27|13473|3
98|Yuli|772-9996|sagittis@nisl.com|2005-04-17|57|1336|5
99|Ray|1-420-314-2886|ac.risus.Morbi@telluseuauge.org|2011-12-30|26|5570|1
100|Cynthia|148-9696|justo.nec.ante@tinciduntvehicula.org|2008-01-23|57|8682|5dev_r@dmc-
● dev-bdp-15-m:~$ hdfs dfs -rm /brunorojas/empresa*
Deleted /brunorojas/empresa20250201.csv
Deleted /brunorojas/empresa20250202.csv
Deleted /brunorojas/empresa20250203.txt
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -rm -r -f /brunorojas/carpeta1
Deleted /brunorojas/carpeta1
○ dev_r@dmc-dev-bdp-15-m:~$ sudo su
root@dmc-dev-bdp-15-m:/home/dev_r# su hdfs
hdfs@dmc-dev-bdp-15-m:/home/dev_r$
```

Para poder realizar el cambio de permisos de ejecución, escritura y lectura así como del propietario de un archivo o carpeta se debe realizar con el superusuario.

Cambiamos a superusuario en Linux con : sudo su

Cambiamos a superusuario en hdfs con: su hdfs

Estructura del comando chown: chown usuario:grupo ruta

Cambiar de propietario a una carpeta

hdfs dfs -chown dmc:bigdata /brunorojas/carpeta2

Cambiar de propietario de forma recursiva

hdfs dfs -chown -R dmc:bigdata /brunorojas/carpeta2

hdfs dfs -chmod 740 /brunorojas/carpeta2

hdfs dfs -chmod -R 750 /brunorojas/carpeta2



```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS sudo - dev_r + v [ ] [ ] ... ^ X
drwxrwxrwt - hdfs hadoop 0 2025-02-11 02:53 /tmp
drwxrwxrwt - hdfs hadoop 0 2025-02-11 02:52 /user
drwxrwxrwt - hdfs hadoop 0 2025-02-11 02:52 /var
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -chown dmc:bigdata /brunorojas/carpeta2
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -ls /brunorojas
Found 6 items
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:25 /brunorojas/_ERROR
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:25 /brunorojas/_SUCCESS
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:22 /brunorojas/archivo_vacio.txt
drwxr-xr-x - dmc bigdata 0 2025-02-13 02:22 /brunorojas/carpeta2
drwxr-xr-x - dev_r hadoop 0 2025-02-13 01:58 /brunorojas/carpeta3
-rw-r--r-- 2 dev_r hadoop 7282 2025-02-13 02:50 /brunorojas/persona.data
hdfs@dmc-dev-bdp-15-m:/home/dev_r$
```

Ln 11, Col 1 (48 selected) Spaces: 4 UTF-8 LF {} Plain Text

ESP LAA 22:13 12/02/2025

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS sudo - dev_r + v [ ] [ ] ... ^ X
drwxr-xr-x - dev_r hadoop 0 2025-02-13 01:58 /brunorojas/carpeta3
-rw-r--r-- 2 dev_r hadoop 7282 2025-02-13 02:50 /brunorojas/persona.data
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -chown -R dmc:bigdata /brunorojas/carpeta2
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -chmod 740 /brunorojas/carpeta2
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -ls /brunorojas
Found 6 items
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:25 /brunorojas/_ERROR
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:25 /brunorojas/_SUCCESS
-rw-r--r-- 2 dev_r hadoop 0 2025-02-13 02:22 /brunorojas/archivo_vacio.txt
drwxr-xr-x - dmc bigdata 0 2025-02-13 02:22 /brunorojas/carpeta2
drwxr-xr-x - dev_r hadoop 0 2025-02-13 01:58 /brunorojas/carpeta3
-rw-r--r-- 2 dev_r hadoop 7282 2025-02-13 02:50 /brunorojas/persona.data
hdfs@dmc-dev-bdp-15-m:/home/dev_r$
```

Ln 13, Col 41 Spaces: 4 UTF-8 LF {} Plain Text

ESP LAA 22:21 12/02/2025

Para asignar permisos de lectura, escritura y ejecución se usa chmod.

hdfs dfs -chmod 740 /bruno\_rojas/carpeta2

Permisos de lectura escritura y ejecución para DMC

Permisos de lectura para el grupo bigdata

Sin ningún permiso para otros usuarios

```
bash_logout
$ .bashrc
$ .profile
$ wget-hsts
BrunoRojas_tarea.txt
comandos.txt
empresa20250201.csv
empresa20250202.csv
empresa20250203.txt
persona.data
transacciones-2018-01-21.data
transacciones-2018-01-22.data
transacciones-2018-01-23.data

hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -ls -R /bruno_rojas
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:38 /bruno_rojas/_ERROR
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:38 /bruno_rojas/_SUCCESS
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:37 /bruno_rojas/archivo_vacio.txt
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA/carpetaB
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA/carpetaB/carpetaC
drwxr-xr-x - dev_r hadoop 0 2025-03-08 02:26 /bruno_rojas/carpeta3
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 03:02 /bruno_rojas/empresa20250203.txt
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -chmod 740 /bruno_rojas/carpeta2
hdfs@dmc-dev-bdp-15-m:/home/dev_r$ hdfs dfs -ls -R /bruno_rojas
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:38 /bruno_rojas/_ERROR
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:38 /bruno_rojas/_SUCCESS
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 02:37 /bruno_rojas/archivo_vacio.txt
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA/carpetaB
drwxr-xr-x - dmc bigdata 0 2025-03-08 03:16 /bruno_rojas/carpeta2/carpetaA/carpetaB/carpetaC
drwxr-xr-x - dev_r hadoop 0 2025-03-08 02:26 /bruno_rojas/carpeta3
-rw-r--r-- 2 dev_r hadoop 0 2025-03-08 03:02 /bruno_rojas/empresa20250203.txt
hdfs@dmc-dev-bdp-15-m:/home/dev_r$
```



## Permisos con ACL

Estructura del comando setfacl: `hdfs dfs -setfacl -R -m user:nombreusuario:wwx /ruta`  
`hdfs dfs -setfacl -R -m user:leonor:rw- /bruno_rojas/archivo_vacio.txt`  
`hdfs dfs -setfacl -R -m group:ulimar:r-x /bruno_rojas/archivo_vacio.txt`  
`hdfs dfs -getfacl /bruno_rojas/archivo_vacio.txt`

```
PROBLEMS OUTPUT TERMINAL
▼ TERMINAL
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -ls /bruno_rojas
Found 6 items
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 02:38 /bruno_rojas/_ERROR
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 02:38 /bruno_rojas/_SUCCESS
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 02:37 /bruno_rojas/archivo_vacio.txt
drwxr----- - dmc  bigdata          0 2025-03-08 03:16 /bruno_rojas/carpeta2
drwxr-xr-x  - dev_r hadoop          0 2025-03-08 02:26 /bruno_rojas/carpeta3
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 03:02 /bruno_rojas/empresa20250203.txt
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -setfacl -R -m user:leonor:rwx /bruno_rojas/archivo_vacio.txt
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -setfacl -R -m group:ulima:r-x /bruno_rojas/archivo_vacio.txt
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -getfacl /bruno_rojas/archivo_vacio.txt
# file: /bruno_rojas/archivo_vacio.txt
# owner: dev_r
# group: hadoop
user::rw-
user:leonor:rwx
group::r--
group:ulima:r-x
mask::rwx
other::r--
○ dev_r@dmc-dev-bdp-15-m:~$
```

Verificación de la integridad de los datos

`cksum /home/dev_r/persona.data`

`hdfs dfs -cat /brunorojas/persona.data | cksum`

Listando pesos recursivamente

`hdfs dfs -du -s -h '/'`

#14.2 K 28.4 K /brunorojas

Los que están en rojo son el doble porque hay dos réplicas.

```
▼ TERMINAL
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -put /home/dev_r/persona.data /bruno_rojas
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -ls /bruno_rojas
Found 7 items
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 02:38 /bruno_rojas/_ERROR
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 02:38 /bruno_rojas/_SUCCESS
-rw-rw-r--r+ 2 dev_r hadoop          0 2025-03-08 02:37 /bruno_rojas/archivo_vacio.txt
drwxr----- - dmc  bigdata          0 2025-03-08 03:16 /bruno_rojas/carpeta2
drwxr-xr-x  - dev_r hadoop          0 2025-03-08 02:26 /bruno_rojas/carpeta3
-rw-r--r--  2 dev_r hadoop          0 2025-03-08 03:02 /bruno_rojas/empresa20250203.txt
-rw-r--r--  2 dev_r hadoop      7282 2025-03-09 02:36 /bruno_rojas/persona.data
● dev_r@dmc-dev-bdp-15-m:~$ cksum /home/dev_r/persona.data
229408302 7282 /home/dev_r/persona.data
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -cat /bruno_rojas/persona.data | cksum
229408302 7282
● dev_r@dmc-dev-bdp-15-m:~$ hdfs dfs -du -s -h '/'
7.1 K 14.2 K /bruno_rojas
14.2 K 42.7 K /brunorojas
42.1 M 84.2 M /hadoop
0 0 /tmp
13.9 M 27.7 M /user
0 0 /var
○ dev_r@dmc-dev-bdp-15-m:~$
```

Cambiando el número de replicas a 3

`hdfs dfs -setrep -w 3 -R /brunorojas`

Ahora los pesos se multiplan por 3 porque hay tres réplicas.

```
dev_r@dnc-dev-bdp-15-m:~$ hdfs dfs -setrep -w 3 -R /bruno_rojas
setrep: '-R': No such file or directory
Replication 3 set: /bruno_rojas/ ERROR
Replication 3 set: /bruno_rojas/_SUCCESS
Replication 3 set: /bruno_rojas/archivo_vacio.txt
Replication 3 set: /bruno_rojas/empresa20250203.txt
Replication 3 set: /bruno_rojas/persona.data
Waiting for /bruno_rojas/_ERROR ... done
Waiting for /bruno_rojas/_SUCCESS ... done
Waiting for /bruno_rojas/archivo_vacio.txt ... done
Waiting for /bruno_rojas/empresa20250203.txt ... done
Waiting for /bruno_rojas/persona.data .... done
dev_r@dnc-dev-bdp-15-m:~$ hdfs dfs -du -s -h '/*'
7.1 K 21.3 K /bruno_rojas
14.2 K 42.7 K /brunorojas
42.1 M 84.2 M /hadoop
0 0 /tmp
13.9 M 27.7 M /user
0 0 /var
dev_r@dnc-dev-bdp-15-m:~$
```