

Tutorial: Manejo de HDFS en Amazon EMR

1. Conexión al Clúster de Amazon EMR:

Inicia una conexión segura utilizando SSH con tu llave de acceso:

```
ssh -i [ruta-de-tu-llave.pem] hadoop@[dirección-IP-del-master-node]
```

Ejemplo de conexión:

```
ssh -i bigdata-key-pair.pem hadoop@ec2-100-26-198-124.compute-1.amazonaws.com
```

2. Operaciones con Archivos en HDFS Mediante la Terminal:

Preparación de Datos:

Prepara un espacio en HDFS para tus datasets creando un directorio llamado 'datasets' dentro de tu directorio home (/user/hadoop).

Asegúrate de que tus datos ya estén descomprimidos y listos para su procesamiento.

Ejemplo de Datasets y Comandos Básicos:

Para obtener los datasets, clona el repositorio con el siguiente comando:

```
git clone https://github.com/hfbanilatq/hfbanilatq-st0263
```

Navega al directorio de datasets:

```
cd hfbanilatq-st0263/reto3.1/
```

Lista los archivos presentes en el directorio raíz de HDFS:

```
hdfs dfs -ls /
```

```
[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -ls /  
Found 4 items  
drwxr-xr-x - hdfs hdfsadmin group 0 2023-11-25 01:10 /apps  
drwxrwxrwt - hdfs hdfsadmin group 0 2023-11-25 01:12 /tmp  
drwxr-xr-x - hdfs hdfsadmin group 0 2023-11-25 01:10 /user  
drwxr-xr-x - hdfs hdfsadmin group 0 2023-11-25 01:10 /var
```

Inspecciona los contenidos de otros directorios relevantes con comandos similares:

```
hdfs dfs -ls /user
```

```
hdfs dfs -ls /user/hadoop
```

```
hdfs dfs -ls /user/hadoop/datasets
```

```
[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -ls /user
Found 9 items
drwxrwxrwx - hadoop hdfsadmingroup 0 2023-11-25 01:10 /user/hadoop
drwxr-xr-x - mapred mapred 0 2023-11-25 01:10 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2023-11-25 01:10 /user/hive
drwxrwxrwx - hue hue 0 2023-11-25 01:10 /user/hue
drwxrwxrwx - livy livy 0 2023-11-25 01:10 /user/livy
drwxrwxrwx - oozie oozie 0 2023-11-25 01:11 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2023-11-25 01:10 /user/root
drwxrwxrwx - spark spark 0 2023-11-25 01:10 /user/spark
drwxrwxrwx - zeppelin hdfsadmingroup 0 2023-11-25 01:10 /user/zeppelin

[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -ls /user/hadoop
[hadoop@ip-10-0-6-239 datasets]$ |
[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -ls /user/hadoop/datasets
ls: '/user/hadoop/datasets': No such file or directory
```

Gestión de tu Directorio de Datasets:

Crea tu directorio 'datasets' en HDFS con:

```
hdfs dfs -mkdir /user/hadoop/datasets
```

Transfiere tus archivos desde el servidor local al directorio HDFS que acabas de crear. Por ejemplo:

```
hdfs dfs -put [ruta-local] /user/hadoop/datasets/
```

```
[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -mkdir /user/hadoop/datasets
[hadoop@ip-10-0-6-239 datasets]$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
[hadoop@ip-10-0-6-239 datasets]$ cd ..
[hadoop@ip-10-0-6-239 bigdata]$ hdfs dfs -put datasets/gutenberg-small/*.txt /user/hadoop/datasets/gutenberg-small/
```

Si tus datos están en Amazon S3, puedes transferirlos directamente con:

Primero debes subir los archivos a S3 para eso emplea la consola de aws, para crear un carpeta en el bucket de s3 llamado datasets, luego sube el archivo airline.csv

Archivos y carpetas (1 Total, 761.8 KB)

Eliminar

Agregar archivos

Agregar carpeta

Se cargarán todos los archivos y las carpetas de esta tabla.

Q Buscar por nombre

< 1 >

<input type="checkbox"/>	Nombre	Carpeta	Tipo	Tamaño
<input type="checkbox"/>	airlines.csv	-	text/csv	761.8 KB

Destino

Destino

s3://hfbailats3/datasets/

► Detalles del destino

Los ajustes del bucket que afectan a los objetos nuevos almacenados en el destino especificado.

Se ha realizado la carga correctamente

Consulte los detalles a continuación.

Resumen

Destino

s3://hfbailats3/datasets/

Realizado correctamente

1 archivo, 761.8 KB (100.00%)

Con errores

0 archivos, 0 B (0%)

Archivos y carpetas

Configuración

Archivos y carpetas (1 Total, 761.8 KB)

Q Buscar por nombre

< 1 >

Nombre	Carpeta	Tipo	Tamaño	Estado	Error
airlines.csv	-	text/csv	761.8 KB	Realizado correctam...	-

Luego emplea:

hadoop distcp s3://[ruta-en-s3] /destino-en-hdfs

```
[hadoop@ip-10-0-6-239 bigdata]$ hadoop distcp s3://hfbailats3/datasets/airlines.csv /tmp
2023-11-25 01:45:14,532 INFO tools.DistCp: Input Options: DistCpOptions{atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false, append=false, useDiff=false, useDiff=false, fromSnapshot=null, toSnapshot=null, skipCRC=false, blocking=true, numListStatusThreads=0, maxMaps=20, mapBandwidth=0.0, copyStrategy='uniformsize', preserveStatus=[], atomicWorkPath=null, logPath=null, sourceFileListing=null, sourcePaths=[s3://hfbailats3/datasets/airlines.csv], targetPath=/tmp, filtersFile='null', blocksPerChunk=0, copyBufferSize=8192, verboseLog=false, directWrite=false, useIterator=false}, sourcePaths=[s3://hfbailats3/datasets/airlines.csv], targetPathExists=true, preserveRawXattrs=false
2023-11-25 01:45:14,802 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at ip-10-0-6-239.ec2.internal/10.0.6.239:8032
2023-11-25 01:45:14,851 INFO client.AHSProxy: Connecting to Application History server at ip-10-0-6-239.ec2.internal/10.0.6.239:10200
2023-11-25 01:45:18,988 INFO tools.SimpleCopyListing: Starting: Building listing using multi threaded approach for s3://hfbailats3/datasets
2023-11-25 01:45:18,989 INFO tools.SimpleCopyListing: Building listing using multi threaded approach for s3://hfbailats3/datasets: duration 0:00.001s
2023-11-25 01:45:19,114 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0
2023-11-25 01:45:19,114 INFO tools.SimpleCopyListing: Build file listing completed.
2023-11-25 01:45:19,115 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
2023-11-25 01:45:19,115 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
2023-11-25 01:45:19,204 INFO tools.DistCp: Number of paths in the copy list: 1
2023-11-25 01:45:19,226 INFO tools.DistCp: Number of paths in the copy list: 1
2023-11-25 01:45:19,242 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at ip-10-0-6-239.ec2.internal/10.0.6.239:8032
2023-11-25 01:45:19,243 INFO client.AHSProxy: Connecting to Application History server at ip-10-0-6-239.ec2.internal/10.0.6.239:10200
2023-11-25 01:45:19,313 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1700874674194_0002
2023-11-25 01:45:19,890 INFO mapreduce.JobSubmitter: number of splits:1
2023-11-25 01:45:20,018 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1700874674194_0002
2023-11-25 01:45:20,018 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-11-25 01:45:20,224 INFO conf.Configuration: resource-types.xml not found
2023-11-25 01:45:20,225 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-11-25 01:45:20,347 INFO impl.YarnClientImpl: Submitted application application_1700874674194_0002
2023-11-25 01:45:20,394 INFO mapreduce.Job: The url to track the job: http://ip-10-0-6-239.ec2.internal:20888/proxy/application_1700874674194_0002/
```

Para una copia de datos recursiva utiliza:

hdfs dfs -copyFromLocal [ruta-local] /user/hadoop/datasets/

```

AC[hadoop@ip-10-0-6-239 bigdata]$ hdfs dfs -copyFromLocal datasets/* /user/hadoop/datasets/
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnLetters.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsFirstInauguralAddress.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsGettysburgAddressGivenNovember-19-1863.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsInauguralsAddressesandLettersSelections.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsSecondInauguralAddress.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___SpeechesandLettersofAbrahamLincoln1832-1865.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___StateoftheUnionAddresses.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheEmancipationProclamation.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume1.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume2.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume3.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume4.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume5.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume6.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume7.txt': File exists

```

Lista los archivos en tu directorio 'datasets' para verificar la transferencia:

hdfs dfs -ls /user/hadoop/datasets

```

-rw-r--r-- 1 hadoop hdfsadmin group 780058 2023-11-25 01:49 /user/hadoop/datasets/airlines.csv
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/all-news
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/covid19
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/flights
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/gutenberg
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/otus
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/otus_logs
-rw-r--r-- 1 hadoop hdfsadmin group 534 2023-11-25 01:49 /user/hadoop/datasets/sample_data.csv
drwxr-xr-x - hadoop hdfsadmin group 0 2023-11-25 01:49 /user/hadoop/datasets/spark
[hadoop@ip-10-0-6-239 bigdata]$ hdfs dfs -ls /user/hadoop/datasets/gutenberg-small/
Found 16 items
-rw-r--r-- 1 hadoop hdfsadmin group 5878 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnLetters.txt
-rw-r--r-- 1 hadoop hdfsadmin group 21586 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmin group 1653 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hdfsadmin group 262083 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsInauguralsAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hdfsadmin group 4093 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmin group 516298 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 hadoop hdfsadmin group 187895 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___StateoftheUnionAddresses.txt
-rw-r--r-- 1 hadoop hdfsadmin group 3928 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheEmancipationProclamation.txt
-rw-r--r-- 1 hadoop hdfsadmin group 45664 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt
-rw-r--r-- 1 hadoop hdfsadmin group 459006 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hdfsadmin group 505150 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hdfsadmin group 254941 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hdfsadmin group 209643 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hdfsadmin group 692051 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hdfsadmin group 601102 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hdfsadmin group 478689 2023-11-25 01:32 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln___TheWritingsofAbrahamLincolnVolume7.txt

```

Recuperación de Archivos desde HDFS:

Para copiar archivos de vuelta al servidor local, usa:

hdfs dfs -get /user/hadoop/datasets/* ~/[tu-username]/mis_datasets/

Otra opción para recuperar archivos es:

hdfs dfs -copyToLocal /user/hadoop/datasets/[nombre-del-archivo] ~/[tu-username]/mis_datasets/

Verifica la transferencia con:

ls -l ~/[tu-username]/mis_datasets

```

[hadoop@ip-10-0-6-239 ~]$ hdfs dfs -get /user/hadoop/datasets/gutenberg-small/* ~/hfbailatq/mis_dataset/
[hadoop@ip-10-0-6-239 ~]$ hdfs dfs -copyToLocal /user/hadoop/datasets/gutenberg/gutenberg-small.zip ~/hfbailatq/mis_dataset/
copyToLocal: '/user/hadoop/datasets/gutenberg/gutenberg-small.zip': No such file or directory
[hadoop@ip-10-0-6-239 ~]$ cd hfbailatq/
[hadoop@ip-10-0-6-239 hfbailatq]$ ls -l mis_datasets
ls: cannot access mis_datasets: No such file or directory
[hadoop@ip-10-0-6-239 hfbailatq]$ ls -l mis_dataset
total 4160
-rw-r--r-- 1 hadoop hadoop 5878 Nov 25 01:52 AbrahamLincoln__LincolnLetters.txt
-rw-r--r-- 1 hadoop hadoop 21586 Nov 25 01:52 AbrahamLincoln__LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hadoop 1653 Nov 25 01:52 AbrahamLincoln__LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hadoop 262083 Nov 25 01:52 AbrahamLincoln__LincolnsInauguralAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hadoop 4093 Nov 25 01:52 AbrahamLincoln__LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hadoop 516298 Nov 25 01:52 AbrahamLincoln__SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 hadoop hadoop 167895 Nov 25 01:52 AbrahamLincoln__StateoftheUnionAddresses.txt
-rw-r--r-- 1 hadoop hadoop 3928 Nov 25 01:52 AbrahamLincoln__TheEmancipationProclamation.txt
-rw-r--r-- 1 hadoop hadoop 45664 Nov 25 01:52 AbrahamLincoln__TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt
-rw-r--r-- 1 hadoop hadoop 459006 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hadoop 505150 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hadoop 254941 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hadoop 209643 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hadoop 692051 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hadoop 601102 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hadoop 478689 Nov 25 01:52 AbrahamLincoln__TheWritingsofAbrahamLincolnVolume7.txt
[hadoop@ip-10-0-6-239 hfbailatq]$

```

Experimenta con otros comandos:

Aplica diversos comandos de HDFS para gestionar tus archivos, como:

`hdfs dfs -du [ruta]` # Muestra el uso de disco

`hdfs dfs -mv [origen] [destino]` # Mueve archivos

`hdfs dfs -cp [origen] [destino]` # Copia archivos

`hdfs dfs -rm [ruta]` # Elimina archivos

`hdfs dfs -put [src-local] [dest-hdfs]` # Copia local a HDFS

`hdfs dfs -cat [nombre-de-archivo]` # Muestra contenido de archivo

```

[hadoop@ip-10-0-6-239 hfbailatq]$ hdfs dfs -cat /user/hadoop/datasets/airline.csv
cat: '/user/hadoop/datasets/airline.csv': No such file or directory
[hadoop@ip-10-0-6-239 hfbailatq]$ /user/hadoop/datasets/airlines.csv
-bash: /user/hadoop/datasets/airlines.csv: No such file or directory
[hadoop@ip-10-0-6-239 hfbailatq]$ hdfs dfs -cat /user/hadoop/datasets/airlines.csv
id,airline,date,location,rating,cabin,value,recommended,review
10001,Delta Air Lines,21-Jun-14,Thailand,7,Economy,4,YES,Flew Mar 30 NRT to BKK. All flights were great. Flight was on-time and the in-flight entertainment was great. Apart from the meals - so
me Thai passengers cannot eat beef so the flight crews tried to ask other passengers who could eat beef and changed the meals around. We feel disappointed with their food services.
10002,Delta Air Lines,19-Jun-14,USA,0,Economy,2,NO,Flight 2463 leaving West Palm Beach (PBI) at 2:42pm on June 15 arriving at New York LaGuardia (LGA) at 5:30pm. I was slated to take flight 39
07 from LaGuardia to Pittsburgh (PIT) leaving at 6:30pm arriving in Pittsburgh at 8:35pm. Every time I fly Delta I end up getting stuck in whatever airport my layover is for 12+ hours due to c
ancellations or delays. There is no attempt to accommodate me the customer with a hotel for the night a partial refund or even a blanket or pillow for my extended stay in the airport.
10003,Delta Air Lines,18-Jun-14,USA,0,Economy,1,NO,Delta Website froze 4 times trying to set up flight to three different locations. Had to call and set up over the phone only to find out at t

```

Además, gestiona los permisos y propiedad de los archivos con:

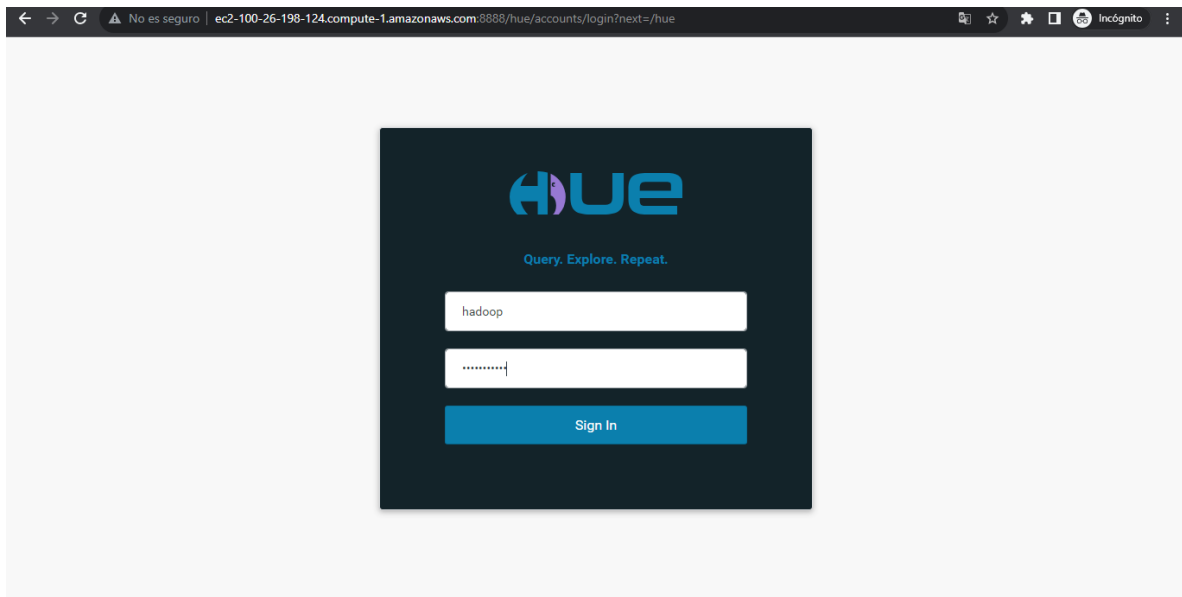
`hdfs dfs -chmod [-R] [modo] [archivo]`

`hdfs dfs -chown [username] [archivo]`

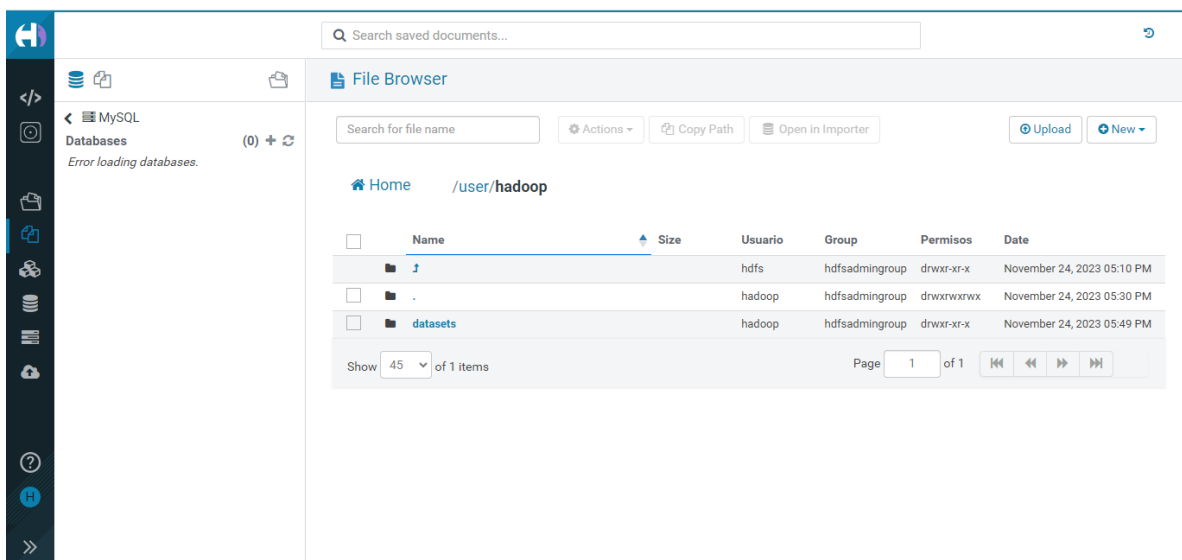
`hdfs dfs -chgrp [grupo] [archivo]`

3. Administración de Archivos a través de HUE en Amazon EMR:

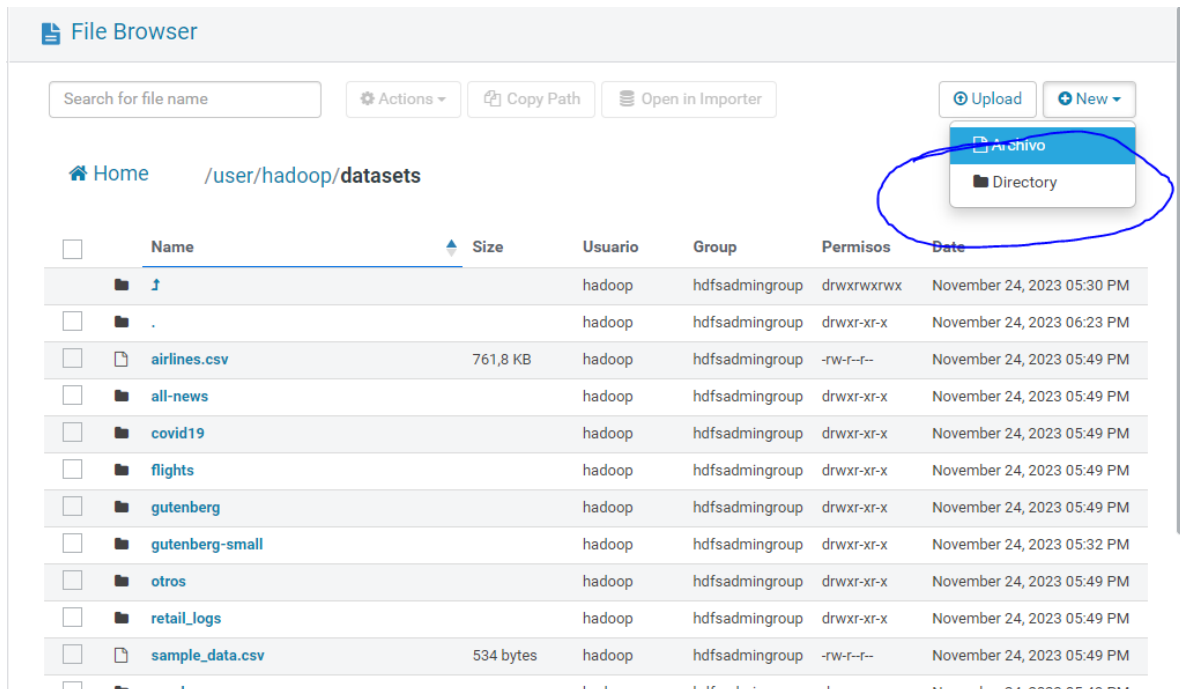
Inicia sesión en la interfaz de HUE para una gestión más visual de los archivos.



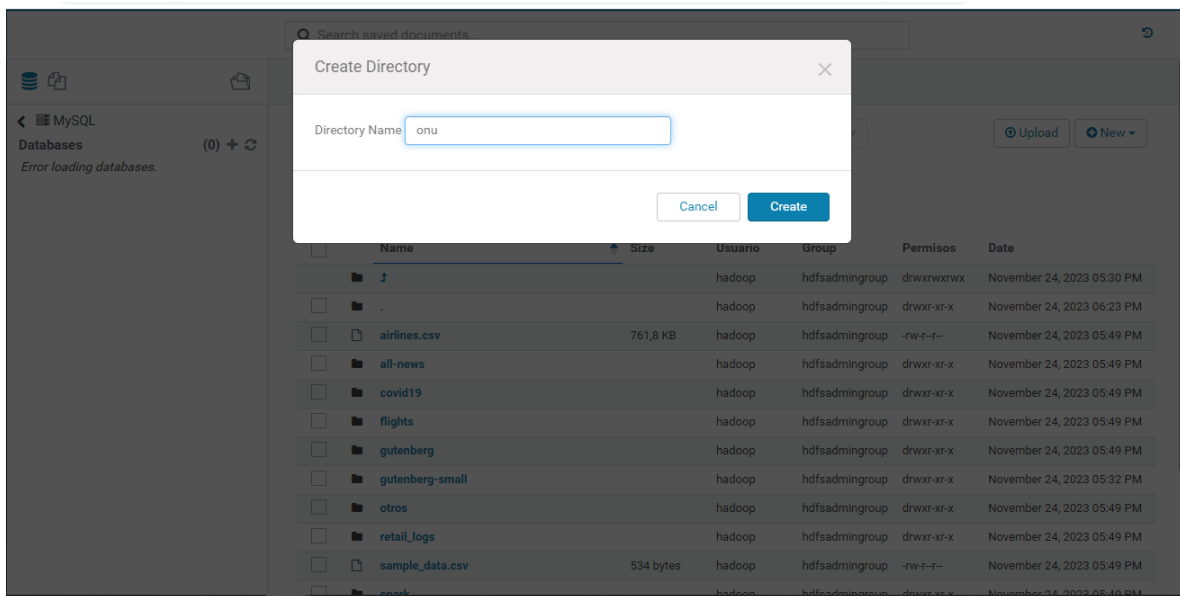
Navega por el sistema de archivos, crea directorios y sube archivos directamente desde la interfaz de HUE.



Como ya subimos todos los archivos usando la terminal, voy a eliminar la carpeta ONU para crearla nuevamente y subir los archivos en esta.



Damo nombre a la carpeta:



La seleccionamos y damos clic en upload para subir archivos

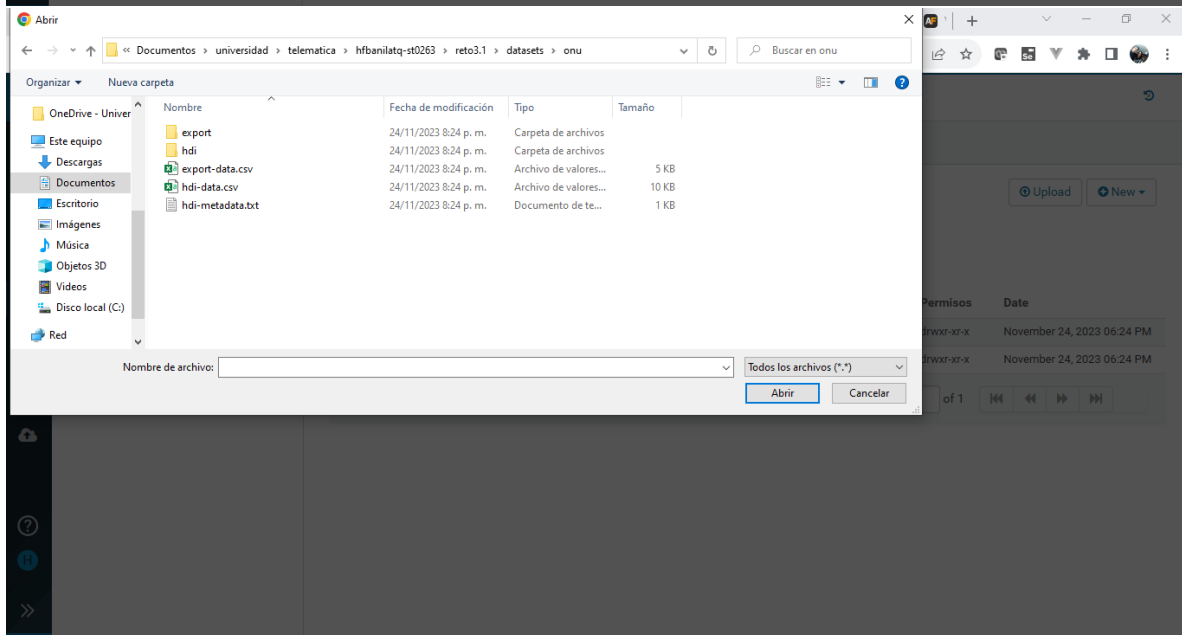
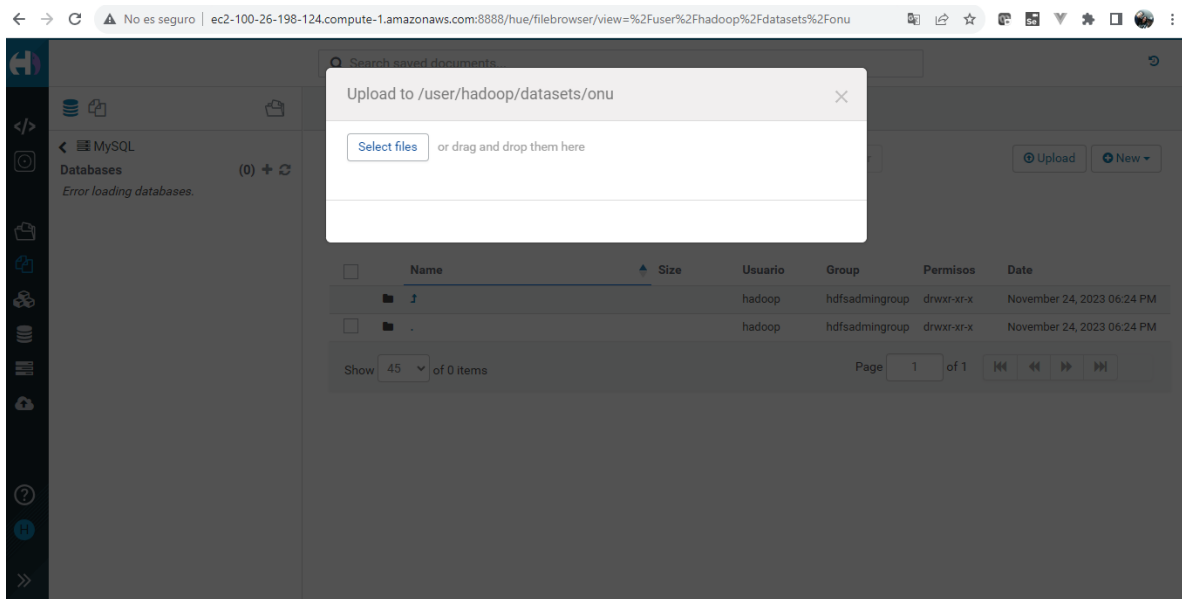
File Browser

[Home](#)[/user/hadoop/datasets/onu](#)

<input type="checkbox"/>	Name	Size	Usuario	Group	Permisos	Date
<input type="checkbox"/>	 ↑		hadoop	hdfsadmingroup	drwxr-xr-x	November 24, 2023 06:24 PM
<input type="checkbox"/>	 .		hadoop	hdfsadmingroup	drwxr-xr-x	November 24, 2023 06:24 PM

Show of 0 items

Page of 1



Con lo anterior quedan los archivos subidos:

File Browser

Search for file name

Actions

Copy Path





Open in Importer

Upload

New


Home

/user/hadoop/datasets/onu

	Name	Size	Usuario	Group	Permisos	Date
<input type="checkbox"/>	 ↑		hadoop	hdfsadmingroup	drwxr-xr-x	November 24, 2023 06:24 PM
<input type="checkbox"/>	 .		hadoop	hdfsadmingroup	drwxr-xr-x	November 24, 2023 06:27 PM
<input type="checkbox"/>	 export-data.csv	4,3 KB	hadoop	hdfsadmingroup	-rw-r--	November 24, 2023 06:27 PM
<input type="checkbox"/>	 hdi-data.csv	9,0 KB	hadoop	hdfsadmingroup	-rw-r--	November 24, 2023 06:27 PM

Show 45 of 2 items

Page 1 of 1




Ahora para ver el contenido de los archivos simplemente damos clic en el archivo que queremos ver

File Browser

Back

Home

Page 1 to 3 of 3



Edit file

Refresh

View as binary

Download

Last modified 24/11/2023 21:27 -05:00

Usuario hadoop

Group hdfsadmingroup

Size 9.02 KB

Mode 100644

/ user/ hadoop/ datasets/ onu/ hdi-data.csv

```
id,country,hdi,lifeex,myschool,eyschool,gni,gni2,nihdi
1,Norway,0.943,81.1,12.6,17.3,47557,6,0.975
2,Australia,0.929,81.9,12.18,34431,16,0.979
3,Netherlands,0.91,80.7,11.6,16.8,36402,9,0.944
4,United States,0.91,78.5,12.4,16,43017,6,0.931
5,New Zealand,0.908,80.7,12.5,18,23737,30,0.978
6,Canada,0.908,81,12.1,16,35166,10,0.944
7,Ireland,0.908,80.6,11.6,18,29322,19,0.959
8,Liechtenstein,0.905,79.6,10.3,14.7,83717,-6,0.877
9,Germany,0.905,80.4,12.2,15.9,34854,8,0.94
10,Sweden,0.904,81.4,11.7,15.7,35837,4,0.936
11,Switzerland,0.903,82.3,11,15.6,39924,0,0.926
12,Japan,0.901,83.4,11.6,15.1,32295,11,0.94
13,Hong Kong China (SAR),0.898,82.8,10,15.7,44805,-4,0.91
14,Iceland,0.898,81.8,10.4,18,29354,11,0.943
15,Korea (Republic of),0.897,80.6,11.6,16.9,28230,12,0.945
16,Denmark,0.895,78.8,11.4,16.9,34347,3,0.926
17,Israel,0.888,81.6,11.9,15.5,25849,14,0.939
18,Belgium,0.886,80,10.9,16.1,33357,2,0.914
19,Austria,0.885,80.9,10.8,15.3,35719,-4,0.908
20,France,0.884,80.5,10.6,15.6,33128,1,0.908
```

Y listo, estos son los métodos con los cuales podemos subir archivos en HDFS.