

HDFS Hadoop en el EMR y comandos varios

```
[hadoop@ip-172-31-71-159 ~]$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
mkdir: '/user/hadoop/datasets/gutenberg-small': File exists
[hadoop@ip-172-31-71-159 ~]$ hdfs dfs -put -st0263-252/bigdata/datasets/gutenberg-small/*.txt /user/hadoop/datasets/gutenberg-small/
[hadoop@ip-172-31-71-159 ~]$ hdfs dfs -ls /user/hadoop/datasets
Found 1 items
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small
[hadoop@ip-172-31-71-159 ~]$ hdfs dfs -ls /user/hadoop/datasets/gutenberg-small
Found 16 items
-rw-r--r-- 1 hadoop hdfsadmingroup 5878 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 21586 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 1653 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 262083 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsInauguralAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 4093 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 516298 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 167893 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_StateoftheUnionAddressees.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 3928 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheEmancipationProclamation.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 45664 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 459096 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 505159 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 254941 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 209643 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 692951 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 601192 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 478689 2025-11-17 02:38 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume7.txt
[hadoop@ip-172-31-71-159 ~]$
```

CloudShell

us-east-1 +

```
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -ls /user/hadoop/datasets
Found 1 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2025-11-10 03:22 /user/hadoop/datasets/gutenberg-small
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -du /user/hadoop/datasets
4229660 4229660 /user/hadoop/datasets/gutenberg-small
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -chmod -R 777 /user/hadoop/datasets
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -cp /user/hadoop/datasets/gutenberg-small /user/telematica/datasets/
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_LincolnsFirstInauguralAddress.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_LincolnsGettysburgAddressGivenNovember-19-1863.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_LincolnsInauguralsAddressesandLettersSelections.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_LincolnsSecondInauguralAddress.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_SpeechesandLettersofAbrahamLincoln1832-1865.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_StateoftheUnionAddresses.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheEmancipationProclamation.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume1.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume2.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume3.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume4.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume5.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume6.txt': File exists
cp: `/user/telematica/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume7.txt': File exists
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -ls /user/telematica/datasets/
Found 2 items
drwxr-xr-x - hadoop telematica 0 2025-11-10 03:54 /user/telematica/datasets/gutenberg-small
drwxr-xr-x - telematica telematica 0 2025-11-10 03:52 /user/telematica/datasets/onus
[hadoop@ip-172-31-78-134 st0263-252]$ hdfs dfs -cat /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt | head
LINCOLN LETTERS
By Abraham Lincoln

Published by The Bibilophile Society

[hadoop@ip-172-31-78-134 st0263-252]$
```

CloudShell

us-east-1 +

```
[hadoop@ip-172-31-78-134 ~]$ hdfs dfs -mkdir /user/hadoop/datasets
mkdir: `/user/hadoop/datasets': File exists
[hadoop@ip-172-31-78-134 ~]$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
mkdir: `/user/hadoop/datasets/gutenberg-small': File exists
[hadoop@ip-172-31-78-134 ~]$ hdfs dfs -ls /user/hadoop
Found 1 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2025-11-10 03:21 /user/hadoop/datasets
[hadoop@ip-172-31-78-134 ~]$
```

Acceso a Hue

The screenshot shows the Hue File Browser interface. The left sidebar has icons for MySQL, Databases, and other connectivity options. The main area is titled "File Browser" and shows the path "/user/telematica". A search bar at the top right contains the placeholder "server_name". Below the search bar are buttons for "Upload" and "New". The main content area displays a table of files and directories:

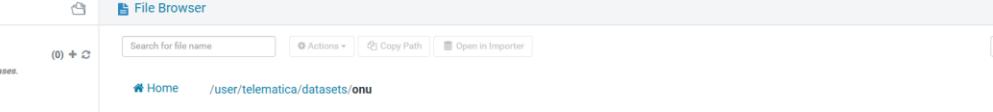
Name	Size	User	Group	Permissions	Date
hdbs		hdbsadmin	group	drwxr-xr-x	November 09, 2025 07:53 PM
telematica		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM
.hive.Jars		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM
datasets		telematica	telematica	drwxr-xr-x	November 09, 2025 07:54 PM

At the bottom, there is a page navigation bar showing "Page 1 of 1" and various navigation icons.

This screenshot shows the Hue File Browser interface again, but this time the path is "/user/telematica/datasets". The left sidebar and search bar are identical to the first screenshot. The main content area displays a table of datasets:

Name	Size	User	Group	Permissions	Date
gutenberg-small		hadoop	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
ona		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM

At the bottom, there is a page navigation bar showing "Page 1 of 1" and various navigation icons.



The screenshot shows the Apache Hue interface for a MySQL database named 'telematica'. The left sidebar lists databases, and the main area displays a file browser for the directory '/user/telematica/datasets/onu'. The browser includes search, actions, copy path, and open importer buttons. It lists two files: 'export-data.csv' (4.3 KB) and 'hdh-data.csv' (9.0 KB), both owned by 'telematica' with permissions 'drwxr-xr-x' and 'rw-r--r--' respectively, and created on November 09, 2025.

Name	Size	User	Group	Permissions	Date
export-data.csv	4.3 KB	telematica	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
hdh-data.csv	9.0 KB	telematica	telematica	-rw-r--r--	November 09, 2025 08:25 PM

The screenshot shows the Apache Hue File Browser interface. The left sidebar has a 'MySQL Databases' section with '(0) +'. Below it, an error message says 'Error loading databases.' The main area is titled 'File Browser' and shows a list of files under '/user/telematica/datasets/gutenberg-small'. The columns are Name, Size, User, Group, Permissions, and Date. The files listed are all text files related to Abraham Lincoln's speeches and letters, such as 'AbrahamLincoln__LincolnLetters.txt', 'AbrahamLincoln__Lincoln'sFirstInauguralAddress.txt', and 'AbrahamLincoln__Lincoln'sGettysburgAddressGivenNovember-19-1863.txt'. The permissions for most files are -rw-r--r--.

Name	Size	User	Group	Permissions	Date
AbrahamLincoln__LincolnLetters.txt	5.7 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__Lincoln'sFirstInauguralAddress.txt	21.1 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__Lincoln'sGettysburgAddressGivenNovember-19-1863.txt	1.6 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__Lincoln'sInauguralAddressesandLettersSelections.txt	255.9 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__Lincoln'sSecondInauguralAddress.txt	4.0 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__SpeechesandLettersofAbrahamLincoln1832-1865.txt	504.2 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__StateoftheUnionAddresses.txt	164.0 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheEmancipationProclamation.txt	3.8 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt	44.6 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume1.txt	448.2 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume2.txt	493.3 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume3.txt	249.0 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume4.txt	204.7 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume5.txt	675.8 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume6.txt	587.0 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM
AbrahamLincoln__TheWritingsofAbrahamLincolnVolume7.txt	467.5 KB	hadoop	telematica	-rw-r--r--	November 09, 2025 07:54 PM