

HDFS Hadoop en el EMR y comandos varios

[illegible]

```

[hadoo@ip-172-31-71-159 ~]$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
mkdir: /user/hadoop/datasets/gutenberg-small: File exists
[hadoo@ip-172-31-71-159 ~]$ hdfs dfs -put ~/st0263-252/bigdata/datasets/gutenberg-small/*.txt /user/hadoop/datasets/gutenberg-small/

[hadoo@ip-172-31-71-159 ~]$
[hadoo@ip-172-31-71-159 ~]$ hdfs dfs -ls /user/hadoop/datasets
Found 1 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small

[hadoo@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 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 21586 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 1653 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 262083 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsInauguralAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 4093 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 516298 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 167895 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_StatsoftheUnionAddresses.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 3928 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheEmancipationProclamation.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 45664 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 459006 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 505150 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 254941 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 289643 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 629951 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 601182 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 478689 2025-11-17 02:30 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume7.txt

[hadoo@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__LincolnLetters.txt': File exists
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/onu
[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 ~]$
[hadoop@ip-172-31-78-134 ~]$
```

Acceso a Hue

The screenshot shows the Hue File Browser interface. The left sidebar contains a navigation menu with icons for Home, File Browser, Databases, and other tools. The main panel displays the file browser for the path `/user/telematica`. At the top, there is a search bar and buttons for 'Upload' and 'New'. Below this, a table lists the files and directories in the current directory.

Name	Size	User	Group	Permissions	Date
<code>.</code>		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM
<code>.hiveJars</code>		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM
<code>datasets</code>		telematica	telematica	drwxr-xr-x	November 09, 2025 07:54 PM

At the bottom of the table, it indicates 'Show 45 of 2 items' and 'Page 1 of 1'.

The screenshot shows the Hue File Browser interface, now displaying the subdirectory `/user/telematica/datasets`. The table lists the files and directories within this subdirectory.

Name	Size	User	Group	Permissions	Date
<code>.</code>		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM
<code>gutenberg-small</code>		hadoop	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
<code>onu</code>		telematica	telematica	drwxr-xr-x	November 09, 2025 07:52 PM

At the bottom of the table, it indicates 'Show 45 of 2 items' and 'Page 1 of 1'.

MySQL
Databases
Error loading databases.

(0) +

File Browser

Search for file name

Actions

Copy Path

Open in Importer

Upload

New

Home

/user/telematica/datasets/onu

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	.		telematica	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
<input type="checkbox"/>	.		telematica	telematica	drwxr-xr-x	November 09, 2025 08:25 PM
<input type="checkbox"/>	export-data.csv	4.3 KB	telematica	telematica	-rw-r--	November 09, 2025 08:25 PM
<input type="checkbox"/>	hdi-data.csv	9.0 KB	telematica	telematica	-rw-r--	November 09, 2025 08:25 PM

Show 45 of 2 items

Page 1 of 1

MySQL
Databases
Error loading databases.

(0) +

File Browser

Search for file name

Actions

Copy Path

Open in Importer

Upload

New

Home

/user/telematica/datasets/gutenberg-small

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	.		telematica	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
<input type="checkbox"/>	.		hadoop	telematica	drwxr-xr-x	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__LincolnLetters.txt	5.7 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__LincolnsFirstInauguralAddress.txt	21.1 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__LincolnsGettysburgAddressGivenNovember-19-1863.txt	1.6 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__LincolnsInauguralAddressesandLettersSelections.txt	255.9 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__LincolnsSecondInauguralAddress.txt	4.0 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__SpeechesandLettersofAbrahamLincoln1832-1865.txt	504.2 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__StateoftheUnionAddresses.txt	164.0 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheEmancipationProclamation.txt	3.8 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt	44.6 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume1.txt	448.2 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume2.txt	493.3 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume3.txt	249.0 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume4.txt	204.7 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume5.txt	675.8 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume6.txt	587.0 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM
<input type="checkbox"/>	Abraham.Lincoln__TheWritingsofAbrahamLincolnVolume7.txt	467.5 KB	hadoop	telematica	-rw-r--	November 09, 2025 07:54 PM

Show 45 of 16 items

Page 1 of 1