

UNIVERSIDAD EAFIT

Tópicos Especiales En Telemática

Laboratorio 3-1 HDFS

Miguel Angel Hoyos Velasquez

Medellín, 18 de Noviembre 2024

```
EFIT Interactiva - Universidad I... (3) WhatsApp
st0263-242/bigdata/01-hdfs ar... Correo: Miguel Angel Hoyos Vo... Launch AWS Academy Learner I... EC2 | us-east-1
us-east-1.console.aws.amazon.com/ec2-instance-connect/shAddressFamily=ipv4&connType=standard&instanceId=i-0f0146bfe3bbbba9380sUs...
AWS Servicios Buscar [Alt+S]
Norte de Virginia voclabs/user3070118-mahoyosv@eafit.edu.co @ 8433-4340-7800

Keyboard shortcut
To tab out of the terminal window and select the next button element, press the left and right Shift keys together.
Close permanently

EEEEEEEEEEEEEEEEEEEE MWWWMM RRRRRR RRRRR
[root@ip-172-31-2-212 ~]# ls
st0263-242
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-10-18 16:14 /apps
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-10-18 16:10 /tmp
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-10-18 16:14 /user
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-10-18 16:14 /var
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /user
Found 9 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2024-10-18 16:14 /user/hadoop
drwxr-xr-x - mapred mapred 0 2024-10-18 16:14 /user/history
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-10-18 16:14 /user/hive
drwxr-xr-x - hue hue 0 2024-10-18 16:14 /user/hue
drwxr-xr-x - livy livy 0 2024-10-18 16:14 /user/livy
drwxr-xr-x - oozie oozie 0 2024-10-18 16:15 /user/oozie
drwxr-xr-x - root root 0 2024-10-18 16:14 /user/root
drwxr-xr-x - spark spark 0 2024-10-18 16:14 /user/spark
drwxr-xr-x - zeppelin hdfsadmingroup 0 2024-10-18 16:14 /user/zeppelin
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /user/hadoop
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /user/hadoop/datasets
ls: /user/hadoop/datasets: No such file or directory
[root@ip-172-31-2-212 ~]# hdfs dfs -mkdir /user/hadoop/datasets
[root@ip-172-31-2-212 ~]# usermaster$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
bash: usermaster$: command not found
[root@ip-172-31-2-212 ~]# hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
[root@ip-172-31-2-212 ~]# ls -l
total 16
drwxr-xr-x 8 root root 16384 Oct 18 16:25 st0263-242
[root@ip-172-31-2-212 ~]# cd st0263-242/
[root@ip-172-31-2-212 st0263-242]# ls
Laboratorio-MW Laboratorio-RPC README.template.md README.md bigdata docker-nginx-wordpress-ssl-letsencrypt eks-wp
[root@ip-172-31-2-212 st0263-242]# cd bigdata/
[root@ip-172-31-2-212 bigdata]# ls
lab-aws-emr 01-hdfs README.md datasets lab3-0-aws-emr.txt lab3-1-hdfs-s3.txt st0263-242-unidad3-labs-student.docx
[root@ip-172-31-2-212 bigdata]# cd datasets/
[root@ip-172-31-2-212 datasets]# ls
airlines.csv all-news covid19 flights gutenberg gutenberg-small onu otros retail_logs sample_data.csv spark
[root@ip-172-31-2-212 datasets]# cd gutenberg-small/
[root@ip-172-31-2-212 gutenberg-small]# hdfs dfs -put *.txt /user/hadoop/datasets/gutenberg-small
[root@ip-172-31-2-212 gutenberg-small]#
```

```
CloudShell Comentarios
© 2024 Amazon Web Services, Inc. o sus filiales. Privacidad Términos Preferencias de cookies

EFIT Interactiva - Universidad I... (3) WhatsApp
st0263-242/bigdata/01-hdfs ar... Correo: Miguel Angel Hoyos Vo... Launch AWS Academy Learner I... EC2 | us-east-1
us-east-1.console.aws.amazon.com/ec2-instance-connect/shAddressFamily=ipv4&connType=standard&instanceId=i-0f0146bfe3bbbba9380sUs...
AWS Servicios Buscar [Alt+S]
Norte de Virginia voclabs/user3070118-mahoyosv@eafit.edu.co @ 8433-4340-7800

Keyboard shortcut
To tab out of the terminal window and select the next button element, press the left and right Shift keys together.
Close permanently

drwxr-xr-x - root hdfsadmingroup 0 2024-10-18 16:14 /user/root
drwxr-xr-x - spark spark 0 2024-10-18 16:14 /user/spark
drwxr-xr-x - zeppelin hdfsadmingroup 0 2024-10-18 16:14 /user/zeppelin
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /user/hadoop
[root@ip-172-31-2-212 ~]# hdfs dfs -ls /user/hadoop/datasets
ls: /user/hadoop/datasets: No such file or directory
[root@ip-172-31-2-212 ~]# hdfs dfs -mkdir /user/hadoop/datasets
[root@ip-172-31-2-212 ~]# usermaster$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
bash: usermaster$: command not found
[root@ip-172-31-2-212 ~]# hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
[root@ip-172-31-2-212 ~]# ls -l
total 16
drwxr-xr-x 8 root root 16384 Oct 18 16:25 st0263-242
[root@ip-172-31-2-212 ~]# cd st0263-242/
[root@ip-172-31-2-212 st0263-242]# ls
Laboratorio-MW Laboratorio-RPC README.template.md README.md bigdata docker-nginx-wordpress-ssl-letsencrypt eks-wp
[root@ip-172-31-2-212 st0263-242]# cd bigdata/
[root@ip-172-31-2-212 bigdata]# ls
lab-aws-emr 01-hdfs README.md datasets lab3-0-aws-emr.txt lab3-1-hdfs-s3.txt st0263-242-unidad3-labs-student.docx
[root@ip-172-31-2-212 bigdata]# cd datasets/
[root@ip-172-31-2-212 datasets]# ls
airlines.csv all-news covid19 flights gutenberg gutenberg-small onu otros retail_logs sample_data.csv spark
[root@ip-172-31-2-212 datasets]# cd gutenberg-small/
[root@ip-172-31-2-212 gutenberg-small]# hdfs dfs -put *.txt /user/hadoop/datasets/gutenberg-small
[root@ip-172-31-2-212 gutenberg-small]# hdfs dfs -ls /user/hadoop/datasets/gutenberg-small
Found 16 items
-rw-r--r-- 1 root hdfsadmingroup 5878 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnLetters.txt
-rw-r--r-- 1 root hdfsadmingroup 21580 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnFirstInauguralAddress.txt
-rw-r--r-- 1 root hdfsadmingroup 1053 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 root hdfsadmingroup 262083 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnInauguralAddressesandLettersSelections.txt
-rw-r--r-- 1 root hdfsadmingroup 4093 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnSecondInauguralAddress.txt
-rw-r--r-- 1 root hdfsadmingroup 516298 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 root hdfsadmingroup 167895 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__StateoftheUnionAddresses.txt
-rw-r--r-- 1 root hdfsadmingroup 3922 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheEmancipationProclamation.txt
-rw-r--r-- 1 root hdfsadmingroup 45664 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheFriendshipServiceofGeneralZacharyTaylorAnAddress.txt
-rw-r--r-- 1 root hdfsadmingroup 459060 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 root hdfsadmingroup 585150 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 root hdfsadmingroup 254941 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 root hdfsadmingroup 208643 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 root hdfsadmingroup 692093 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 root hdfsadmingroup 691162 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 root hdfsadmingroup 478689 2024-10-18 16:31 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume7.txt
[root@ip-172-31-2-212 gutenberg-small]#
```


File Browser

Search for file name: Actions Copy Path Open in Importer Upload New

us-east-1 s3a://

	Name	Size	Usuario	Group	Permisos	Date
<input type="checkbox"/>	.				drwxrwxr-x	
<input type="checkbox"/>	aws-logs-843343407800-us-east-1				drwxrwxr-x	
<input type="checkbox"/>	mahoyev				drwxrwxr-x	

Show 45 of 2 items Page 1 of 1

File Browser

Search for file name: Actions Copy Path Open in Importer Upload New

Inicio /user/hadoop/datasets

	Name	Size	Usuario	Group	Permisos	Date
<input type="checkbox"/>	1		hadoop	hdfsadmin	drwxrwxr-x	October 18, 2024 12:07 PM
<input type="checkbox"/>	.		hadoop	hdfsadmin	drwxrwxr-x	October 18, 2024 12:07 PM

Show 45 of 0 items Page 1 of 1

Create Directory

Directory Name:

Cancel Create

File Browser interface showing the directory structure of /user/hadoop/datasets. The interface includes a search bar, a file list table, and a sidebar with navigation options.

Search for file name: [] Actions: [] Copy Path: [] Open in Importer: []

Inicio /user/hadoop/datasets

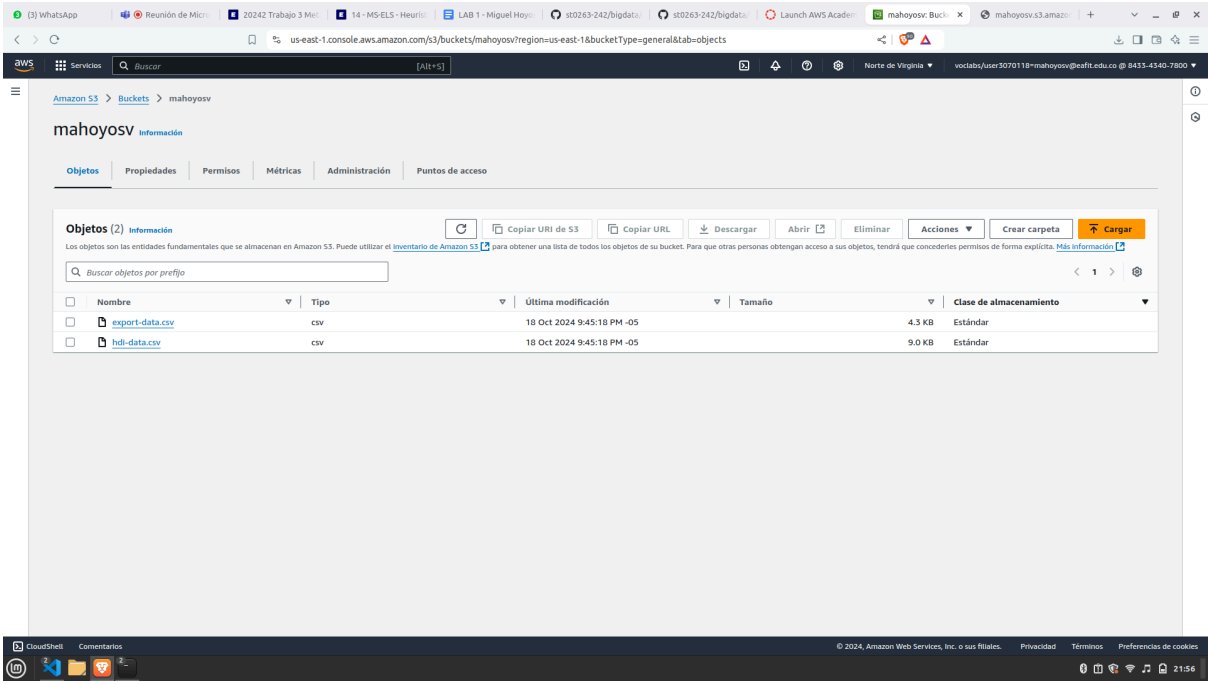
Name	Size	Usuario	Group	Permisos	Date
.		hadoop	hdfsadmingroup	drwxrwxr-x	October 18, 2024 12:07 PM
ONU		hadoop	hdfsadmingroup	drwxr-xr-x	October 18, 2024 12:13 PM
export-data.csv	4,3 KB	hadoop	hdfsadmingroup	-rwxr-xr-x	October 18, 2024 12:13 PM
hdi-data.csv	9,0 KB	hadoop	hdfsadmingroup	-rwxr-xr-x	October 18, 2024 12:13 PM

Show 45 of 3 items Page 1 of 1

File Browser interface showing the contents of the file export-data.csv. The interface includes a sidebar with navigation options and a large text area displaying the file content.

Inicio /user/hadoop/datasets/export-data.csv

country,export
Afghanistan,15.48717184
Albania,29.77238976
Algeria,38.83848547
American Samoa,
Andorra,
Angola,56.83588283
Antigua and Barbuda,44.88266774
Argentina,21.79646876
Armenia,20.5836199
Aruba,
Australia,19.78824202
Austria,33.97185586
Azerbaijan,55.12575386
Bahamas, The,42.82174624
Bahrain,
Bangladesh,19.48616841
Barbados,47.34395912
Belarus,54.6236892
Belgium,80.08874773
Belize,
Benin,14.12394984
Bermuda,
Bhutan,
Bolivia,41.18745687
Bosnia and Herzegovina,35.92835962
Botswana,32.64188364
Brazil,11.15298
Brunei Darussalam,
Bulgaria,57.79868296
Burkina Faso,
Burundi,
Cambodia,54.88278279
Cameroon,28.92224376
Canada,29.43868726
Cape Verde,39.25519746
Cayman Islands,
Central African Republic,



Link Bucket:https://mahoyosv.s3.amazonaws.com/