

UNIVERSIDAD EAFIT

Tópicos Especiales en Telemática

Laboratorio 1-1

Juan Manuel Gómez - Ingeniería de Sistemas

(jmgomezp@eafit.edu.co)

Profesor: Edwin Nelson Montoya

Medellín, 7 de noviembre de 2024

[illegible]

```
[AWS] [Services] [Search] [Alt+F] [N. Virginia] [vociabz/user3070189@imgagez@eaft.032023.0.6]
Installed:
git-2.40.1-1.amzn2023.0.3.x86_64      git-core-2.40.1-1.amzn2023.0.3.x86_64      git-core-doc-2.40.1-1.amzn2023.0.3.noarch
perl-Git-2.40.1-1.amzn2023.0.3.noarch  perl-TermReadKey-2.38-9.amzn2023.0.2.x86_64  perl-lib-0.65-477.amzn2023.0.6.x86_64

Complete!
[hadoop@ip-172-31-33-51 ~]$ git clone https://github.com/st0263/eaft.git
Cloning into 'st0263-242'...
remote: Enumerating objects: 1194, done.
remote: Counting objects: 100% (1194/1194), done.
remote: Compressing objects: 100% (922/922), done.
remote: Total 1194 (delta 220), reused 1192 (delta 220), pack-reused 0 (from 0)
Receiving objects: 100% (1194/1194), 30.97 MiB | 35.71 MiB/s, done.
Resolving deltas: 100% (222/222), done.
[hadoop@ip-172-31-33-51 ~]$ ls
st0263-242
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /
Found 4 items
drwxr-xr-x   - hdfs hdfsadmingroup    0 2024-10-18 19:39 /apps
drwxrwxrwt   - hdfs hdfsadmingroup    0 2024-10-18 19:41 /tmp
drwxr-xr-x   - hdfs hdfsadmingroup    0 2024-10-18 19:39 /user
drwxr-xr-x   - hdfs hdfsadmingroup    0 2024-10-18 19:39 /var
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user
Found 9 items
drwxrwxrwx   - hadoop hdfsadmingroup    0 2024-10-18 19:39 /user/hadoop
drwxr-xr-x   - mapped mapped            0 2024-10-18 19:39 /user/history
drwxrwxrwx   - hdfs hdfsadmingroup    0 2024-10-18 19:39 /user/hive
drwxrwxrwx   - hue hue                0 2024-10-18 19:39 /user/hue
drwxrwxrwx   - livy livy              0 2024-10-18 19:43 /user/livy
drwxrwxrwx   - oozie oozie            0 2024-10-18 19:40 /user/oozie
drwxrwxrwx   - root hdfsadmingroup    0 2024-10-18 19:39 /user/root
drwxrwxrwx   - spark spark            0 2024-10-18 19:39 /user/spark
drwxrwxrwx   - zeppelin hdfsadmingroup  0 2024-10-18 19:39 /user/zeppelin
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop/datasets
ls: /user/hadoop/datasets: No such file or directory
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -mkdir /user/hadoop/datasets
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop/datasets
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -mkdir /user/hadoop/datasets/gutenberg-small
[hadoop@ip-172-31-33-51 ~]$ []

I-06e3ca77a5ed0019a
PublicIPs: 54.162.32.231 PrivateIPs: 172.31.33.51
```

```
aws Services [Alt+S] N. Virginia voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036

[hadoop@ip-172-31-33-51 ~]$ pwd
/home/hadoop
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -put /home/hadoop/st0263-242/bigdata/datasets/gutenberg-small/*.txt /user/hadoop/datasets/gutenberg-small/
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop/datasets/gutenberg-small
Found 16 items
-rw-r--r-- 1 hadoop hdfsadmin group 5878 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnLetters.txt
-rw-r--r-- 1 hadoop hdfsadmin group 21586 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmin group 1653 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hdfsadmin group 262083 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnsInauguralAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hdfsadmin group 4093 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmin group 516298 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__SpeechesandLettersofAbrahamLincoln1832-1865.txt
-rw-r--r-- 1 hadoop hdfsadmin group 167895 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__StateoftheUnionAddresses.txt
-rw-r--r-- 1 hadoop hdfsadmin group 3928 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheEmancipationProclamation.txt
-rw-r--r-- 1 hadoop hdfsadmin group 45664 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress
.txt
-rw-r--r-- 1 hadoop hdfsadmin group 459006 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hdfsadmin group 505150 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hdfsadmin group 254941 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hdfsadmin group 209643 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hdfsadmin group 692051 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hdfsadmin group 601102 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hdfsadmin group 478689 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume7.txt
[hadoop@ip-172-31-33-51 ~]$
```

I-06e3ca77a5ed0019a

PublicIPs: 54.162.32.231 PrivateIPs: 172.31.33.51

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

```
aws Services [Alt+S] N. Virginia voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036

.txt
-rw-r--r-- 1 hadoop hdfsadmin group 459006 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume1.txt
-rw-r--r-- 1 hadoop hdfsadmin group 505150 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume2.txt
-rw-r--r-- 1 hadoop hdfsadmin group 254941 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume3.txt
-rw-r--r-- 1 hadoop hdfsadmin group 209643 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume4.txt
-rw-r--r-- 1 hadoop hdfsadmin group 692051 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume5.txt
-rw-r--r-- 1 hadoop hdfsadmin group 601102 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume6.txt
-rw-r--r-- 1 hadoop hdfsadmin group 478689 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln__TheWritingsofAbrahamLincolnVolume7.txt
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -copyFromLocal /home/hadoop/st0263-242/bigdata/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__LincolnsInauguralAddressesandLettersSelections.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
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop/datasets/
Found 11 items
-rw-r--r-- 1 hadoop hdfsadmin group 780058 2024-10-18 19:52 /user/hadoop/datasets/airlines.csv
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/all-news
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/covid19
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/flights
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/gutenberg
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/omn
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/otros
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/retail_logs
-rw-r--r-- 1 hadoop hdfsadmin group 534 2024-10-18 19:52 /user/hadoop/datasets/sample_data.csv
-dwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/spark
[hadoop@ip-172-31-33-51 ~]$
```

I-06e3ca77a5ed0019a

PublicIPs: 54.162.32.231 PrivateIPs: 172.31.33.51

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

```
aws Services Search [Alt+S] N. Virginia voclabs/user3070189-jmgomezp@eaffit.edu.co @ 2651-7354-0036

[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -copyFromLocal /home/hadoop/st0263-242/bigdata/datasets/* /user/hadoop/datasets/
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt': File exists
copyFromLocal: /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnFirstInauguralAddress.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
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -ls /user/hadoop/datasets/
Found 11 items
-rw-r--r-- 1 hadoop hdfsadmin group 780058 2024-10-18 19:52 /user/hadoop/datasets/airlines.csv
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/all-news
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/covid19
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/flights
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/gutenberg
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:51 /user/hadoop/datasets/gutenberg-small
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/omn
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/otros
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/retail_logs
-rw-r--r-- 1 hadoop hdfsadmin group 534 2024-10-18 19:52 /user/hadoop/datasets/sample_data.csv
drwxr-xr-x - hadoop hdfsadmin group 0 2024-10-18 19:52 /user/hadoop/datasets/spark

[hadoop@ip-172-31-33-51 ~]$ mkdir mis_datasets
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -copyToLocal /user/hadoop/datasets/gutenberg/gutenberg-small.zip /home/hadoop/mis_datasets/
copyToLocal: /user/hadoop/datasets/gutenberg/gutenberg-small.zip: No such file or directory
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -copyToLocal /user/hadoop/datasets/gutenberg/gutenberg-small /home/hadoop/mis_datasets/
copyToLocal: /user/hadoop/datasets/gutenberg/gutenberg-small: No such file or directory
[hadoop@ip-172-31-33-51 ~]$ hdfs dfs -copyToLocal /user/hadoop/datasets/gutenberg/* /home/hadoop/mis_datasets/
[hadoop@ip-172-31-33-51 ~]$ ls mis_datasets/
gutenberg-small-es.zip gutenberg-txt-es.zip-url.txt
[hadoop@ip-172-31-33-51 ~]$
```

I-06e3ca77a5ed0019a

PublicIPs: 54.162.32.231 PrivateIPs: 172.31.33.51

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences