Last login: Tue Apr 8 21:23:35 2025 from 192.168.1.84

vboxuser@bigdata :~ $ su-hadoop

su-hadoop: command not found

vboxuser@bigdata :~ $ su - hadoop

Password:

hadoop@bigdata :~ $ start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [bigdata]

Starting resourcemanager

Starting nodemanagers

hadoop@bigdata :~ $ hdfs dfs -mkdir /Tarea4

hadoop@bigdata :~ $

vboxuser@bigdata :~ $ su-hadoop

su-hadoop: command not found

vboxuser@bigdata :~ $ su - hadoop

Password:

hadoop@bigdata :~ $ start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [bigdata]

Starting resourcemanager

Starting nodemanagers

hadoop@bigdata :~ $ hdfs dfs -mkdir /Tarea4

hadoop@bigdata :~ $ https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-s

ongs-2023

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023:

No such file or directory

hadoop@bigdata :~ $ https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed

-spotify-songs-2024

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-so

ngs-2024: No such file or directory

hadoop@bigdata :~ $ wget https://www.kaggle.com/datasets/nelgiriyewithana/most-str

eamed-spotify-songs-2024

hadoop@bigdata :~ $ https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-s

ongs-2023

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023:

No such file or directory

hadoop@bigdata :~ $ https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed

-spotify-songs-2024

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-so

ngs-2024: No such file or directory

hadoop@bigdata :~ $ wget https://www.kaggle.com/datasets/nelgiriyewithana/most-str

eamed-spotify-songs-2024

-- 2025-04-11 03:35:18 -- https://www.kaggle.com/datasets/nelgiriyewithana/most-s

treamed-spotify-songs-2024

Resolving www.kaggle.com (www.kaggle.com) ... 35.244.233.98

Connecting to www.kaggle.com (www.kaggle.com) |35.244.233.98|:443 ... connected.

HTTP request sent, awaiting response ... 200 OK

Length: unspecified [text/html]

Saving to: 'most-streamed-spotify-songs-2024'

[ <= >

2025-04-11 03:35:19 (14,4 MB/s) - 'most-streamed-spotify-songs-2024' saved [1274

9]

nadoop@bigdata :~ $

most-streamed-spoti

12,45K

--.- KB/s

in 0,001s

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023:

No such file or directory

hadoop@bigdata :~ $ https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed

-spotify-songs-2024

-bash: https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-so

ngs-2024: No such file or directory

hadoop@bigdata :~ $ wget https://www.kaggle.com/datasets/nelgiriyewithana/most-str

eamed-spotify-songs-2024

-- 2025-04-11 03:35:18 -- https://www.kaggle.com/datasets/nelgiriyewithana/most-s

treamed-spotify-songs-2024

Resolving www.kaggle.com (www.kaggle.com) ... 35.244.233.98

Connecting to www.kaggle.com (www.kaggle.com) | 35.244.233.98|:443 ... connected.

HTTP request sent, awaiting response ... 200 OK

Length: unspecified [text/html]

Saving to: 'most-streamed-spotify-songs-2024'

most-streamed-spoti

2025-04-11 03:35:19 (14,4 MB/s) - 'most-streamed-spotify-songs-2024' saved [1274

9]

hadoop@bigdata :~ $ pwd

/home/hadoop

hadoop@bigdata :~ $

[ <= >

12,45K

--.- KB/s

in 0,001s

hadoop@bigdata :~ $ pwd

/home/hadoop

hadoop@bigdata :~ $ hdfs dfs -put /home/hadoop/rows.csv /Tarea4

hadoop@bigdata :~ $ hdfs dfs -put /home/hadoop/rows. csv /Tarea4

put: `/Tarea4/rows.csv': File exists

hadoop@bigdata :~ $ hdfs dfs -put /home/hadoop/^Csv /Tarea4

hadoop@bigdata :~ $ hdfs dfs-put/home/hadoop/most-streamed-spotify-songs-2024/Tare

a4

/home/hadoop/hadoop/libexec/hadoop-functions.sh: line 2386: HDFS\_DFS-PUT/HOME/HA

DOOP/MOST-STREAMED-SPOTIFY-SONGS-2024/TAREA4 USER: invalid variable name

ERROR: dfs-put/home/hadoop/most-streamed-spotify-songs-2024/Tarea4 is not COMMAN

D nor fully qualified CLASSNAME.

Usage: hdfs [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]

OPTIONS is none or any of:

-- buildpaths

-- config dir

-daemon (start | status | stop)

-- debug

attempt to add class files from build tree

Hadoop config directory

operate on a daemon

turn on shell script debug mode

version

Daemon Commands :

balancer

datanode

dfsrouter

diskbalancer

httpfs

journalnode

mover

hamenode

hfs3

portmap

secondarynamenode

sps

zkfc

SUBCOMMAND may print help when invoked w/o parameters or with -h.

hadoop@bigdata :~ $ hdfs dfs -ls /Tarea4

Found 1 items

-rw-r -- r -- 1 hadoop supergroup

hadoop@bigdata :~ $

directory contents with a snapshot

print the version

run a cluster balancing utility

run a DFS datanode

run the DFS router

Distributes data evenly among disks on a given node

run HttpFS server, the HDFS HTTP Gateway

run the DFS journalnode

run a utility to move block replicas across storage types

run the DFS namenode

run an NFS version 3 gateway

run a portmap service

run the DFS secondary namenode

run external storagepolicysatisfier

run the ZK Failover Controller daemon

271214 2025-04-11 03:37 /Tarea4/rows.csv

Importamos librerias necesarias

rom pyspark. sql import SparkSession, functions as F

Inicializa la sesión de Spark

park = SparkSession.builder. appName ('Tarea4') .getOrCreate ()

Define la ruta del archivo .csv en HDFS

file path = 'hdfs://localhost : 9000/Tarea4/rows. csv'

Lee el archivo .csv

if = spark. read. format ('csv' ) . option ('header', 'true') . option ('inferSchema', 't>

Imprimimos el esquema para verificar los nombres y tipos de columnas

rint ("Esquema del DataFrame : ")

if.printSchema ()

Muestra las primeras filas del DataFrame

print ("\nPrimeras filas del DataFrame: ")

lf.show (truncate=False)

Active ESM Apps para recibir futuras actualizaciones de seguridad adicionales.

Vea https://ubuntu.com/esm o ejecute «sudo pro status>

New release '24.04.2 LTS' available.

Run 'do-release-upgrade' to upgrade to it.

Last login: Fri Apr 11 04:01:03 2025 from 192.168.1.84

vboxuser@bigdata :~ $ nano tarea4.py

vboxuser@bigdata :~ $ nano tarea4.py

vboxuser@bigdata :~ $ nano tarea4.py

vboxuser@bigdata :~ $ nano tarea4.py

vboxuser@bigdata :~ $ python3 tarea4.py

25/04/11 04:55:59 WARN Utils: Your hostname, bigdata resolves to a loopback addr

ess: 127.0.1.1; using 192.168.1.56 instead (on interface enp0s3)

25/04/11 04:55:59 WARN Utils: Set SPARK\_LOCAL\_IP if you need to bind to another

address

Setting default log level to "WARN".

To adjust logging level use sc.setLogLevel (newLevel). For SparkR, use setLogLeve

1 (newLevel) .

25/04/11 04:56:00 WARN NativeCodeLoader: Unable to load native-hadoop library fo

r your platform ... using builtin-java classes where applicable

4430050}

Sent: {'sensor id': 8, 'temperature': 29.68, 'humidity': 31.37, 'timestamp' : 174

4430051}

Sent: {'sensor id': 7, 'temperature' : 24.15, 'humidity' : 65. 78, 'timestamp' : 174

4430052}

Sent: {'sensor id': 7, 'temperature': 27.29, 'humidity' : 34.25, 'timestamp' : 174

4430053}

Sent: {'sensor id': 8, 'temperature' : 23.85, 'humidity' : 65.29, 'timestamp' : 174

4430054}

Sent: {'sensor id': 7, 'temperature' : 28.11, 'humidity' : 47.48, 'timestamp' : 174

4430055}

Sent: {'sensor\_id' : 2, 'temperature' : 24.3, 'humidity' : 33.96, 'timestamp' : 1744

430056}

Sent: {'sensor id' : 9, 'temperature' : 22.34, 'humidity' : 42.85, 'timestamp' : 174

4430057}

Sent: {'sensor id': 10, 'temperature' : 21.19, 'humidity' : 53.21, 'timestamp' : 17

44430058}

Sent: {'sensor id': 9, 'temperature' : 20.76, 'humidity': 39.46, 'timestamp' : 174

4430059}

Sent: {'sensor id' : 2, 'temperature' : 22.33, 'humidity' : 34. 77, 'timestamp' : 174

4430060}

Sent: {'sensor\_id': 10, 'temperature': 22.89, 'humidity' : 48.87, 'timestamp': 17

44430061}

avg (temperature) |

{2025-04-12 03:46 ... |

{2025-04-12 03:52 ... |

{2025-04-12 03:42 ... |

{2025-04-12 03:49 ..

{2025-04-12 03:45.

{2025-04-12 03:46.

{2025-04-12 03:41 ..

{2025-04-12 03:42.

{2025-04-12 03:47.

{2025-04-12 03:53.

{2025-04-12 03:49.

{2025-04-12 03:51 ..

{2025-04-12 03:43

{2025-04-12 03:46.

{2025-04-12 03:53.

{2025-04-12 03:47.

{2025-04-12 03:41 ... |

{2025-04-12 03:47 ... |

{2025-04-12 03:42 ... |

{2025-04-12 03:52 ...

window| sensor id|

5|23.390000025431316| 51.96333312988281|

3|23.203332901000977| 53.8999989827474|

3|25.355454531582918| 42.31727287986062|

4|25.549999872843426|54. 716667811075844|

10|

4|24. 638749837875366|56.334999561309814|

5|23.204999923706055|47.144999504089355|

1|23.744000244140626| 46.89000015258789|

8| 24.88800048828125| 51.36199951171875|

8|24.397500038146973| 44.90499973297119|

3|25.699999809265137| 55.07499885559082|

7|27.550000190734863| 56.73750019073486|

8|24.135000228881836| 47.50500043233236|

9|24.804285866873606| 48.07857131958008|

9|25.619999885559082| 50.90750074386597|

3|24.532856804983957| 49.25999995640346|

3|26.515000025431316| 55.40833346048991|

5| 22.40999984741211| 65.04000091552734|

10|26.426666895548504| 43.84000015258789|

2|24.549999952316284|

avg (humidity) |

22.5| 45.00799980163574|

51.4449987411499