

```
ubuntu@servidor_ubuntu: ~/herramientas_big_data
remote: Enumerating objects: 206, done.
remote: Total 206 (delta 0), reused 0 (delta 0), pack-reused 206 (from 1)
Receiving objects: 100% (206/206), 18.96 MiB | 3.52 MiB/s, done.
Resolving deltas: 100% (82/82), done.
ubuntu@servidor_ubuntu:~$ ls
dockerdata herramientas_big_data imagenes snap test.txt
ubuntu@servidor_ubuntu:~$ cd herramientas_big_data/
ubuntu@servidor_ubuntu:~/herramientas_big_data$ sudo docker-compose -f docker-compose-v1.yml up -d
Creating network "herramientasbigdata_default" with the default driver
Creating volume "herramientasbigdata_hadoop_historyserver" with default driver
Creating volume "herramientasbigdata_hadoop_namenode" with default driver
Creating volume "herramientasbigdata_hadoop_datanode" with default driver
Creating volume "herramientasbigdata_hadoop_datasets" with default driver
Pulling namenode (bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8)...
2.0.0-hadoop3.2.1-java8: Pulling from bde2020/hadoop-namenode
3192219afd04: Pull complete
7127a1d8cced: Pull complete
883a89599900: Pull complete
77920a3e82af: Pull complete
92329e81aec4: Extracting [=>] 8.913MB/359.7MB
f373218fec59: Download complete
aa53513fe997: Download complete
8b1800105b98: Download complete
c3a84a3e49c8: Download complete
a65640a64a76: Download complete
facffb3a6de3: Download complete
c71a6df73788: Download complete
73b8c0ccb707: Download complete
```

```
ubuntu@servidor_ubuntu: ~/herramientas_big_data
KEYS boot entrypoint.sh hadoop home lib64 mnt proc run sbin sys usr
bin dev etc hadoop-data lib media opt root run.sh srv tmp var
root@51945896087a:/# cd home
root@51945896087a:/home# ls
Datasets
root@51945896087a:/home# remove Datasets
bash: remove: command not found
root@51945896087a:/home# rm Datasets
rm: cannot remove 'Datasets': Is a directory
root@51945896087a:/home# rmdir Datasets
root@51945896087a:/home# dir
root@51945896087a:/home# exit
exit
ubuntu@servidor_ubuntu:~$ sudo docker cp ./Datasets namenode:/home/Datasets
lsstat /home/ubuntu/Datasets: no such file or directory
ubuntu@servidor_ubuntu:~$ cd herramientas_big_data/
ubuntu@servidor_ubuntu:~/herramientas_big_data$ sudo docker cp ./Datasets namenode:/home/Datasets
ubuntu@servidor_ubuntu:~/herramientas_big_data$ dir
Datasets Paso04_ConConsulta.hql docker-compose.yml
Generacion_Ventas.ipynb Paso05.py ejemploNeo4J.txt
Mongo Paso06_GeneracionVentasNuevasPorDia.py hadoop-hive.env
Parquet Paso06_IncrementalVentas.py hadoop.env
Paso00.sh README.md hbase-distributed-local.env
Paso01.sh docker-compose-kafka.yml iris.hql
Paso02.hql docker-compose-v1.yml pruebaPySpark.py
Paso02_ConConsultas.hql docker-compose-v2.yml pruebaScala.scala
Paso03.hql docker-compose-v3.yml pyspark-ETL.ipynb
Paso04.hql docker-compose-v4.yml
ubuntu@servidor_ubuntu:~/herramientas_big_data$
```

Browse Directory

/data

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	root	supergroup	15.93 KB	Aug 25 15:43	3	128 MB	airports.csv	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	calendario	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	canaldeventa	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	cliente	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	compra	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	data_nvo	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	empleado	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	gasto	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	root	supergroup	4.7 KB	Aug 25 15:43	3	128 MB	iris.csv	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	root	supergroup	15.43 KB	Aug 25 15:43	3	128 MB	iris.json	<input type="checkbox"/>

```
ubuntu@servidor_ubuntu: ~/herramientas_big_data
71dccc16415: Pull complete
9e0b2dae03c0: Pull complete
74d67ce0424c: Pull complete
91b3d508e3eb: Pull complete
d1ef019eb47e: Pull complete
Digest: sha256:620267768985bb57e52a86db9263a354e92d0202319d835678852539b21e0895
Status: Downloaded newer image for bde2020/hive:2.3.2-postgresql-metastore
Pulling hive-metastore-postgresql (bde2020/hive-metastore-postgresql:2.3.0)...
2.3.0: Pulling from bde2020/hive-metastore-postgresql
5c90d4a2d1a8: Pull complete
22337bfd13a9: Pull complete
c3961b297acc: Pull complete
5a17453338b4: Pull complete
6364e0d7a283: Pull complete
58c25f5c0dad: Pull complete
f0e675ce88d9: Pull complete
10f26c680a34: Pull complete
873d2c220bff: Pull complete
fd10fb78ded6: Pull complete
ff1356ba118b: Pull complete
8161ea5e47f1: Pull complete
b399213c70b6: Pull complete
08bd4e9a6388: Pull complete
Digest: sha256:9ab91699d15131b874829e6572006cd9d9f1cca413f438b6f21c65b412152bf1
Status: Downloaded newer image for bde2020/hive-metastore-postgresql:2.3.0
resourceManager is up-to-date
datanode is up-to-date
historyserver is up-to-date
Creating hive-metastore-postgresql ...
namenode is up-to-date
nodemanager is up-to-date
Creating hive-metastore ...
Creating hive-metastore-postgresql
Creating hive-server ...
Creating hive-metastore
Creating hive-metastore ... done
ubuntu@servidor_ubuntu:~/herramientas_big_data$
```

DS-M4-Herramientas x Browsing HDFS x WhatsApp x Meet: qni-fyea-h x Google Meet x ChatGPT x + -

No es seguro 192.168.100.55:9870/explorer.html/#

PALTAS DE LA ABUE... Home - Henry Prep Course: Carrer... Gonzalo Posse wayfarer - home gr... GPT GitHub United Afro House... Recorder Todos los marcadores

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:43	0	0 B	data	<input type="button" value="Trash"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 18:43	0	0 B	data2	<input type="button" value="Trash"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 15:18	0	0 B	rmstate	<input type="button" value="Trash"/>
<input type="checkbox"/>	drwx-wx-wx	root	supergroup	0 B	Aug 25 17:01	0	0 B	tmp	<input type="button" value="Trash"/>
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Aug 25 18:39	0	0 B	user	<input type="button" value="Trash"/>

Showing 1 to 5 of 5 entries 1

```
ubuntu@servidor_ubuntu: ~/herramientas_big_data
at org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)
at org.apache.hadoop.hdfs.DistributedFileSystem.mkdirsInternal(DistributedFileSystem.java:1061)
at org.apache.hadoop.hdfs.DistributedFileSystem.mkdirs(DistributedFileSystem.java:1036)
at org.apache.hadoop.hive.ql.session.SessionState.createPath(SessionState.java:747)
at org.apache.hadoop.hive.ql.session.SessionState.createSessionDirs(SessionState.java:670)
at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:582)
... 9 more
root@1065c266c38c:/opt# hdfs dfsadmin -safemode leave
Safe mode is OFF
root@1065c266c38c:/opt# hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop-2.7.4/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in file:/opt/hive/conf/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different
execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
OK
default
integrador
integrador2
Time taken: 2.43 seconds, Fetched: 3 row(s)
hive> quit;
root@1065c266c38c:/opt#
```

Browser tabs: (3) WhatsApp, Meet: nyc-feha-gwi, ChatGPT, DS-M4-Herramientas_Big_D, Browsing HDFS

Address bar: No es seguro 192.168.100.74:9870/explorer.html#/data3/compra

Browse Directory

/data3/compra Go! [Icons]

Show 25 entries Search: []

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:51	0	0 B	fecha=2015
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:51	0	0 B	fecha=2016
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:52	0	0 B	fecha=2017
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:52	0	0 B	fecha=2018
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:52	0	0 B	fecha=2019
drwxr-xr-x	root	supergroup	0 B	Aug 25 22:52	0	0 B	fecha=2020

Showing 1 to 6 of 6 entries Previous 1 Next

Hadoop, 2019.

ubuntu@servidor_ubuntu: ~/herramientas_big_data

```
hive> CREATE INDEX index_fechaentrega ON TABLE venta(Fecha_Entrega)
> AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
> WITH DEFERRED REBUILD ;
OK
Time taken: 1.641 seconds
hive> show INDEXES ON venta;
OK
index_fechaentrega      venta      fecha_entrega      integrador2__venta_index_fechaentrega__
compact
index_venta_sucursal    venta      idsucursal          integrador2__venta_index_venta_sucursal
compact
Time taken: 1.415 seconds, Fetched: 2 row(s)
hive> DESCRIBE FORMATTED venta;
OK
# col_name              data_type              comment
idventa                  int
fecha                    date
fecha_entrega            date
idcanal                  int
idcliente                int
idsucursal               int
idempleado               int
idproducto               int
precio                   float
cantidad                 int

# Detailed Table Information
Database:                 integrador2
```