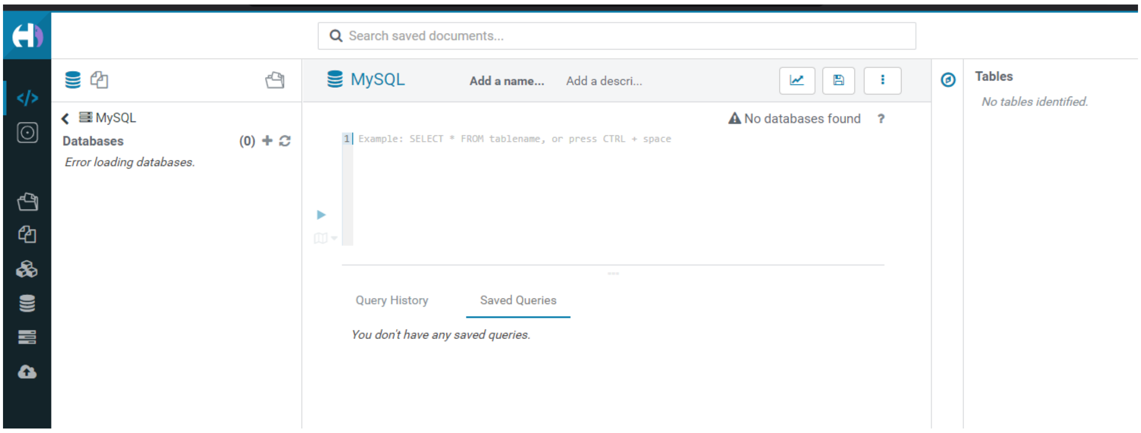
Creación Hadoop:



No tables identified.

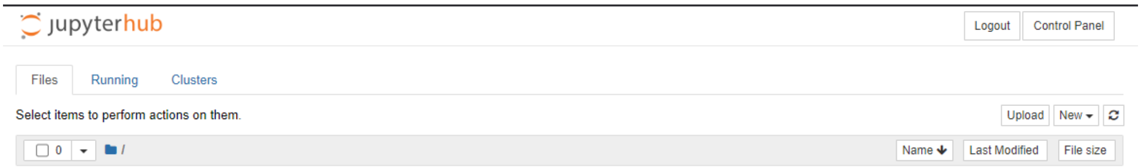
Error loading databases.

No databases found

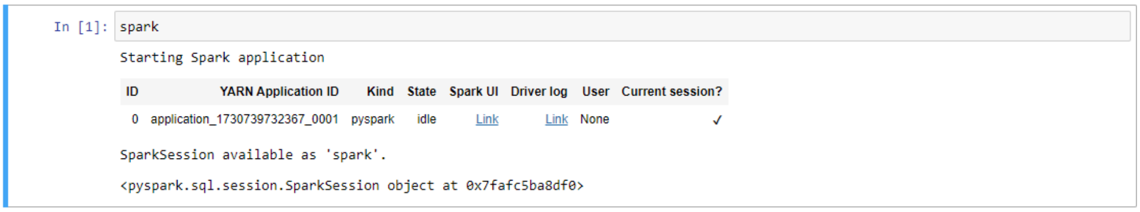
```
1 Example: SELECT * FROM tablename, or press CTRL + space
```

You don't have any saved queries.

Creación JupyterHub



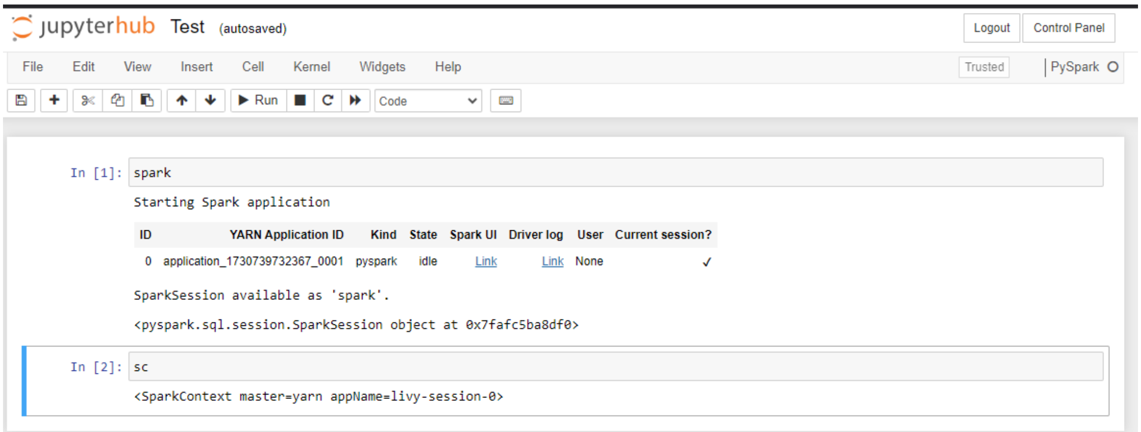
Select items to perform actions on them.



```
In [1]: spark
Starting Spark application
```

ID	YARN Application ID	Kind	State	Spark UI	Driver log	User	Current session?
0	application_1730739732367_0001	pyspark	idle	Link	Link	None	✓

```
SparkSession available as 'spark'.
<pyspark.sql.session.SparkSession object at 0x7fafc5ba8df0>
```



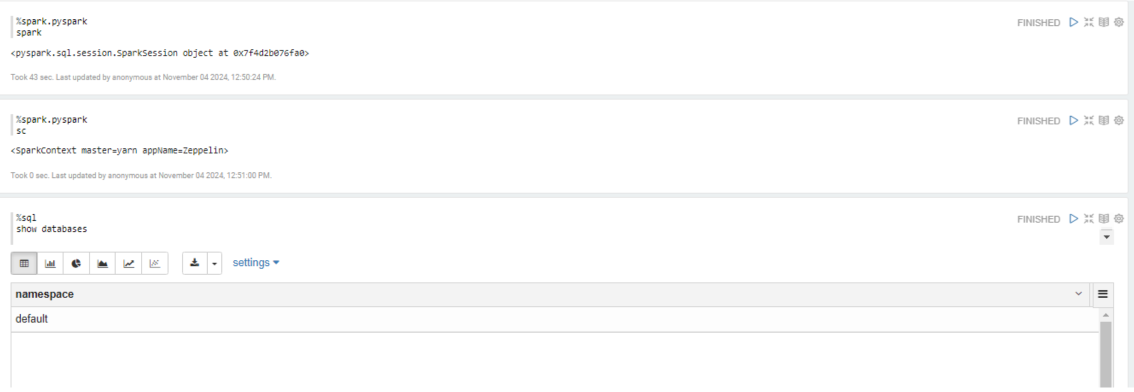
```
In [1]: spark
Starting Spark application
```

ID	YARN Application ID	Kind	State	Spark UI	Driver log	User	Current session?
0	application_1730739732367_0001	pyspark	idle	Link	Link	None	✓

```
SparkSession available as 'spark'.
<pyspark.sql.session.SparkSession object at 0x7fafc5ba8df0>

In [2]: sc
<SparkContext master=yarn appName=livy-session-0>
```

Usando Zeppelin



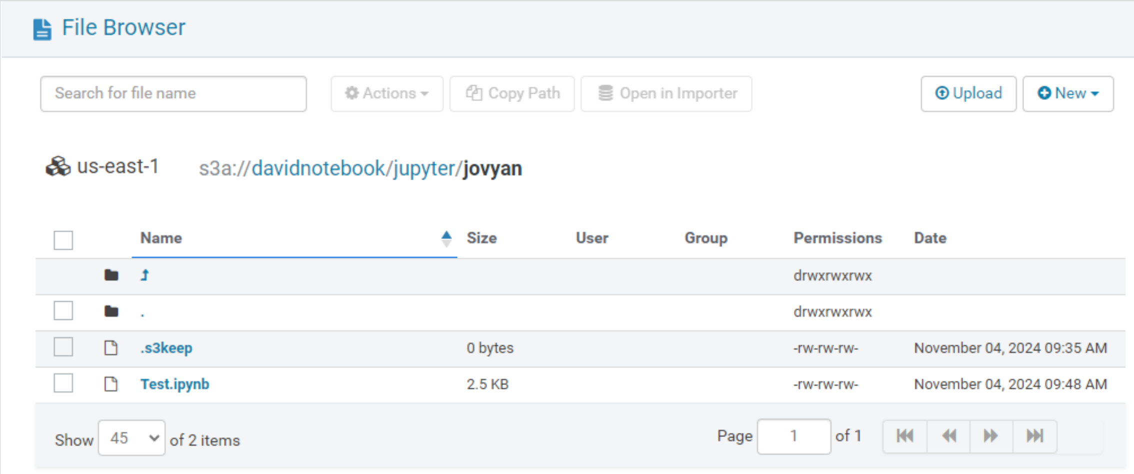
```
%spark.pyspark
spark
<pyspark.sql.session.SparkSession object at 0x7f4d2b076fa0>
Took 43 sec. Last updated by anonymous at November 04 2024, 12:50:24 PM.
```

```
%spark.pyspark
sc
<SparkContext master=yarn appName=Zeppelin>
Took 0 sec. Last updated by anonymous at November 04 2024, 12:51:00 PM.
```

```
%sql
show databases
```

namespace
default

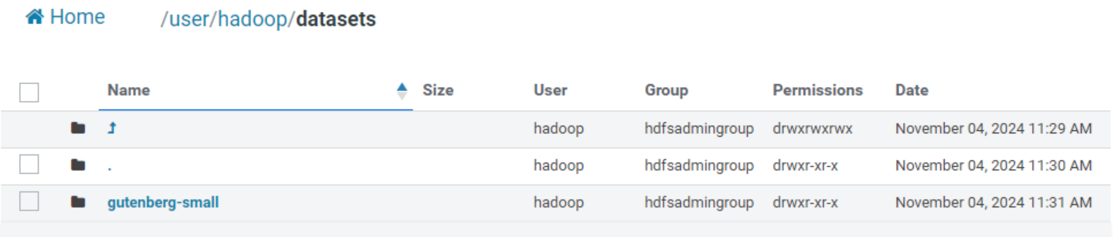
Archivos S3 desde Hue.



s3a://davidnotebook/jupyter/**jovyan**

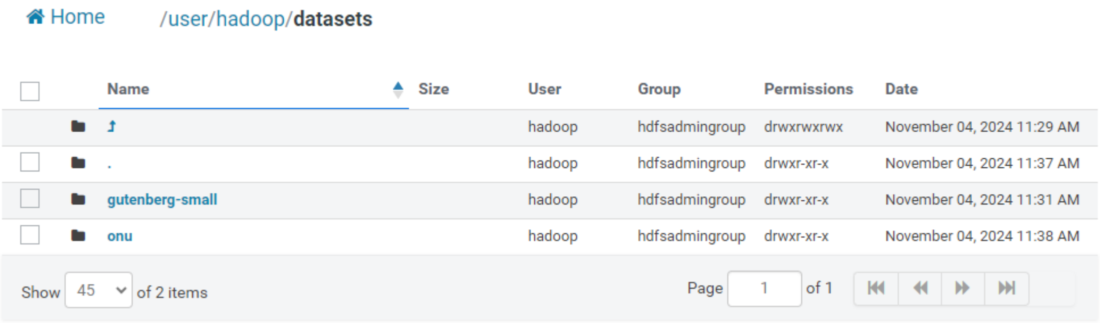
☐		Name	Size	User	Group	Permissions	Date
		↥				drwxrwxrwx	
☐		.				drwxrwxrwx	
☐		.s3keep	0 bytes			-rw-rw-rw-	November 04, 2024 09:35 AM
☐		Test.ipynb	2.5 KB			-rw-rw-rw-	November 04, 2024 09:48 AM

Archivos montados en datasets de hadoop.



/user/hadoop/**datasets**

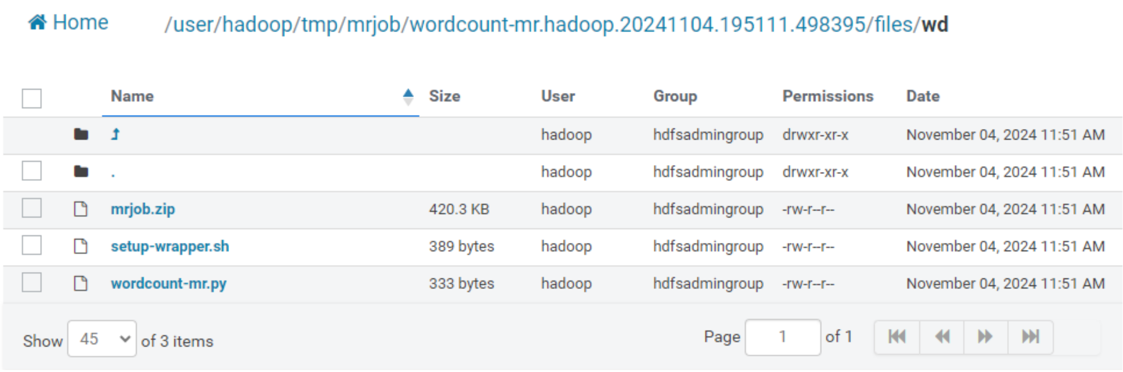
☐		Name	Size	User	Group	Permissions	Date
		↥		hadoop	hdfsadmingroup	drwxrwxrwx	November 04, 2024 11:29 AM
☐		.		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:30 AM
☐		gutenberg-small		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:31 AM



/user/hadoop/**datasets**

☐		Name	Size	User	Group	Permissions	Date
		↥		hadoop	hdfsadmingroup	drwxrwxrwx	November 04, 2024 11:29 AM
☐		.		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:37 AM
☐		gutenberg-small		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:31 AM
☐		onu		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:38 AM

Wordcount:

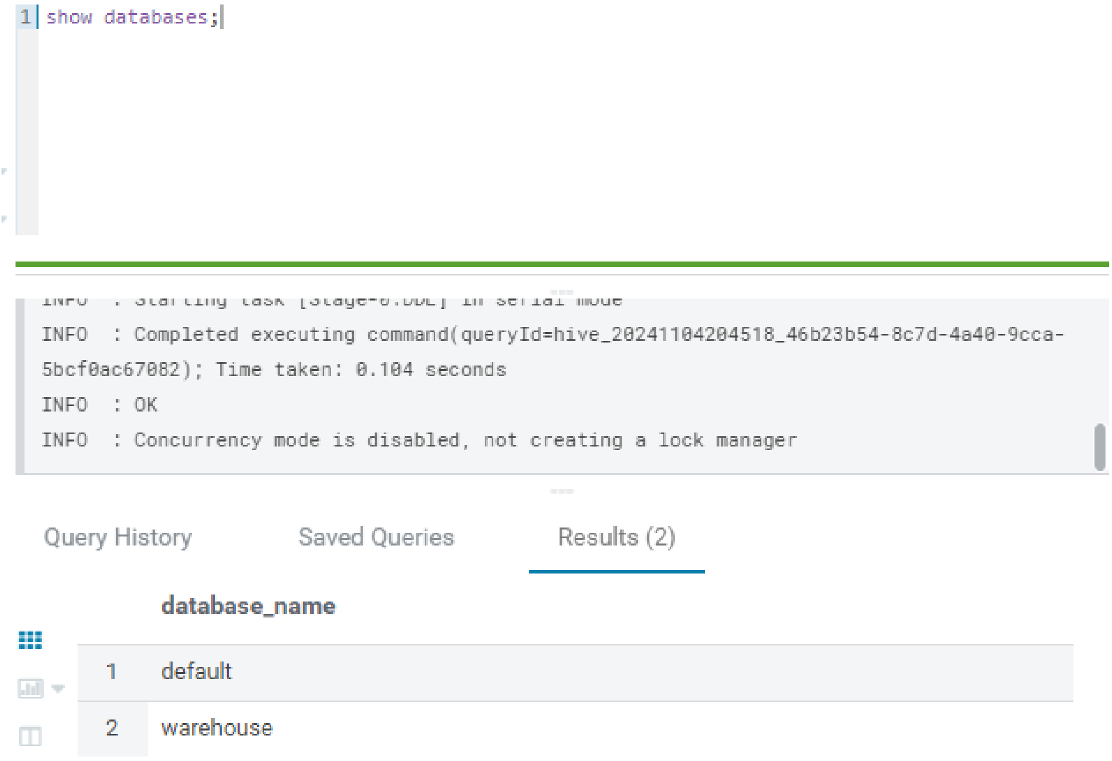


Home /user/hadoop/tmp/mrjob/wordcount-mr.hadoop.20241104.195111.498395/files/**wd**

□		Name	Size	User	Group	Permissions	Date
		↥		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:51 AM
□		.		hadoop	hdfsadmingroup	drwxr-xr-x	November 04, 2024 11:51 AM
□		mrjob.zip	420.3 KB	hadoop	hdfsadmingroup	-rw-r--r--	November 04, 2024 11:51 AM
□		setup-wrapper.sh	389 bytes	hadoop	hdfsadmingroup	-rw-r--r--	November 04, 2024 11:51 AM
□		wordcount-mr.py	333 bytes	hadoop	hdfsadmingroup	-rw-r--r--	November 04, 2024 11:51 AM

Show 45 of 3 items Page 1 of 1

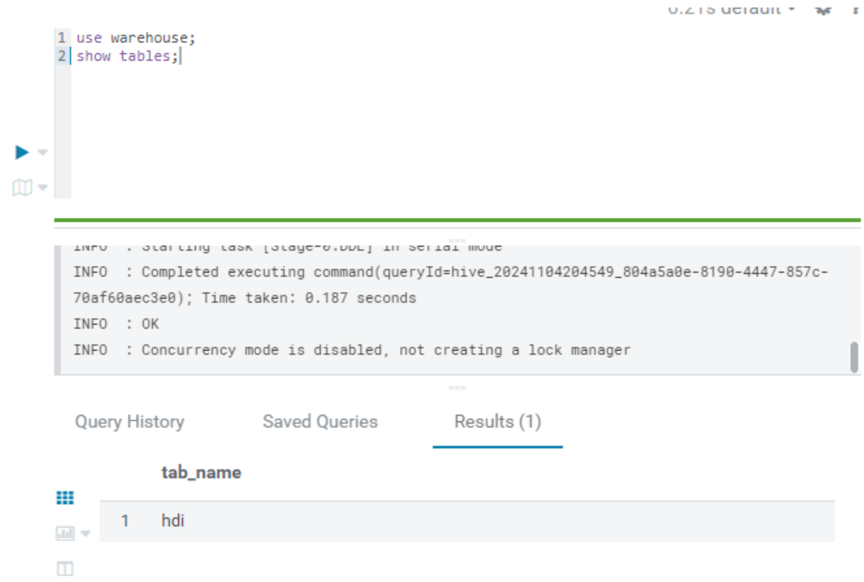
Hive SparkSQL



```
show databases;
```

```
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20241104204518_46b23b54-8c7d-4a40-9cca-
5bcf0ac67082); Time taken: 0.104 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
```

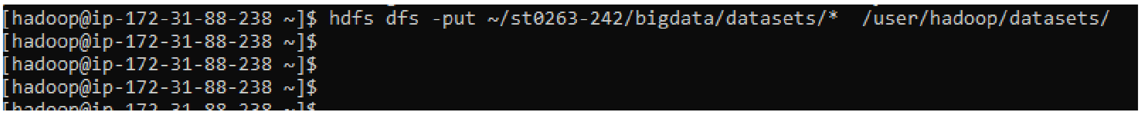
	database_name
1	default
2	warehouse



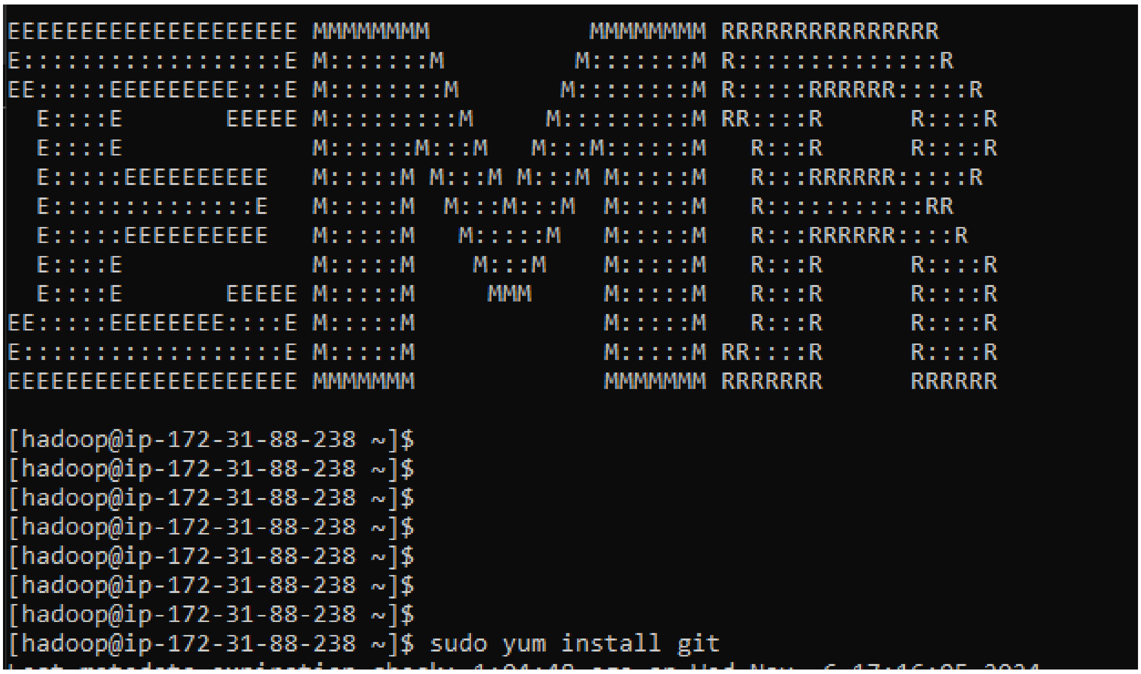
```
use warehouse;
show tables;
```

```
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20241104204549_804a5a0e-8190-4447-857c-
70af60aec3e0); Time taken: 0.187 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
```

	tab_name
1	hdi

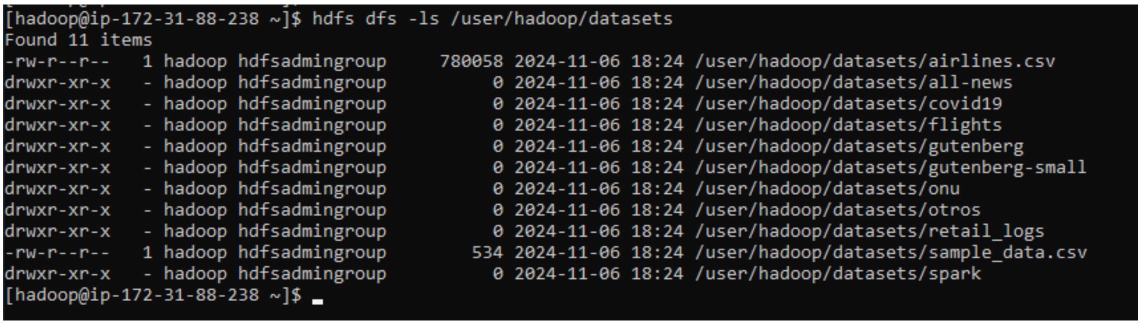


```
[hadoop@ip-172-31-88-238 ~]$ hdfs dfs -put ~/st0263-242/bigdata/datasets/*  /user/hadoop/datasets/
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
```



```
EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E::::::::::::::E   M:::::M  M:::M:::M  M:::::M   R:::::::::::RR
  E:::::EEEEEEEEEE   M:::::M   M:::::M   M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$
[hadoop@ip-172-31-88-238 ~]$ sudo yum install git
Last metadata expiration check: 1:04:48 ago on Wed Nov  6 17:16:05 2024.
```



```
[hadoop@ip-172-31-88-238 ~]$ hdfs dfs -ls /user/hadoop/datasets
Found 11 items
-rw-r--r--   1 hadoop hdfsadmingroup     780058 2024-11-06 18:24 /user/hadoop/datasets/airlines.csv
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/all-news
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/covid19
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/flights
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/gutenberg
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/gutenberg-small
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/onu
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/otros
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/retail_logs
-rw-r--r--   1 hadoop hdfsadmingroup        534 2024-11-06 18:24 /user/hadoop/datasets/sample_data.csv
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-06 18:24 /user/hadoop/datasets/spark
[hadoop@ip-172-31-88-238 ~]$
```