

Creación tablas HDI

1

2

3

4

```
CREATE TABLE HDI (id INT, country STRING, hdi FLOAT, lifeex INT, mysch INT, eysch INT, gni INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE
```

▶

📖

INFO : Starting task [stage-0:DOL] in serial mode

INFO : Completed executing command(queryId=hive\_20241106180313\_2013480e-e938-4ce9-ac67-aaab7d5e4f8); Time taken: 0.563 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

0.32s default

⚙️ ?

1

2

3

4

```
CREATE EXTERNAL TABLE HDI (id INT, country STRING, hdi FLOAT, lifeex INT, mysch INT, eysch INT, gni INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/hadoop/datasets/onu/hdi/'
```

▶

📖

INFO : Starting task [Stage-0:DOL] in serial mode

INFO : Completed executing command(queryId=hive\_20241106184045\_31800d5f-6ff8-4c00-9d1d-1680670007bc); Time taken: 0.349 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

✔ Success.

0.27s default

⚙️ ?

1

2

```
show tables;
```

▼

▼

INFO : Starting task [stage-0:DOL] in serial mode

INFO : Completed executing command(queryId=hive\_20241106180735\_bb0ff4c0-1dc1-45d2-a197-0dc3cc57997b); Time taken: 0.213 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

Query History

Saved Queries

Results (1)

🔍

tab\_name

1 hdi

```
1 desc hdi;
2
```

```
INFO : Executing command(queryId=hive_20241106180910_53ea1b75-4d77-4864-adea-ee7e714a466): desc hdi
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20241106180910_53ea1b75-4d77-4864-adea-ee7e714a466); Time taken: 0.117 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

	col_name	data_type	comment
1	id	int	
2	country	string	
3	hdi	float	
4	lifeex	int	
5	mysch	int	
6	eysoh	int	
7	gni	int	

```
1 select * from hdi;
2
```

```
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20241106180910_53ea1b75-4d77-4864-adea-ee7e714a466): select * from hdi
INFO : Completed executing command(queryId=hive_20241106180910_53ea1b75-4d77-4864-adea-ee7e714a466); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

	hdi.id	hdi.country	hdi.hdi	hdi.lifeex	hdi.mysch	hdi.eysoh	hdi.gni
1	NULL	country	NULL	NULL	NULL	NULL	NULL
2	1	Norway	0.943	81	12	17	47557
3	2	Australia	0.929	81	12	18	34431
4	3	Netherlands	0.91	80	11	16	36402
5	4	United States	0.91	78	12	16	43017
6	5	New Zealand	0.908	80	12	18	23737

Select en tabla HDI

```
1 select country, gni from hdi where gni>2000;
```

```
INFO : Concurrently mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20241106181003_ed8c4d38-e2b6-4946-bb7d-d1819fc3a4b4): select country, gni from hdi where gni>2000
INFO : Completed executing command(queryId=hive_20241106181003_ed8c4d38-e2b6-4946-bb7d-d1819fc3a4b4); Time taken: 0.0 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

Query History Saved Queries Results (100+)

	country	gni
1	Norway	47557
2	Australia	34431
3	Netherlands	36402
4	United States	43017
5	New Zealand	23737
6	Canada	35166
7	Ireland	29322
8	Liechtenstein	83717

## Creación tabla EXPO

```
1 CREATE EXTERNAL TABLE EXPO (country STRING, expct FLOAT)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
3 STORED AS TEXTFILE
4 LOCATION 's3://davidnotebook/datasets/onu/export/'
```

```
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20241106181605_2c992803-fadc-473f-bc32-a6182)
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

✓ Success.

Tables (2) + ↻

Filter...

expo

hdi

## Join tablas

1

SELECT h.country, gni, expct FROM HDI h JOIN EXPO e ON (h.country = e.country) WHERE gni > 2000;

▶

📄

INFO : Map 1: 1/1 Map 2: 1/1

INFO : Completed executing command(queryId=hive\_20241106181706\_acf5e93e-84cf-4169-ac43-5f194df982dd); Time taken: 7.723 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

Query History

Saved Queries

Results (100+)

📄

📄

📄

	h.country	gni	expct
1	Albania	7803	29.77231
2	Algeria	7658	30.830406
3	Andorra	36095	NULL
4	Angola	4874	56.835884
5	Antigua and Barbuda	15521	44.08267
6	Argentina	14527	21.706469

## Wordcount

1

CREATE EXTERNAL TABLE docs (line STRING)

2

STORED AS TEXTFILE

3

LOCATION 's3://davidnotebook/datasets/gutenberg-small/';

▶

📄

LOCATION 's3://davidnotebook/datasets/gutenberg-small/'

INFO : Starting task [Stage-0:DDL] in serial mode

INFO : Completed executing command(queryId=hive\_20241106182602\_513ebb89-14e6-47c4-88aa-3789fa3f3a02); Time taken: 0.402 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

✔ Success.

```

1 SELECT word, count(1) AS count FROM (SELECT explode(split(line, ' ')) AS word FROM docs) w
2 GROUP BY word
3 ORDER BY word DESC LIMIT 10;

```

```

INFO : Map 1: 1/1      Reducer 2: 2/2 Reducer 3: 0/1/1
INFO : Completed executing command(queryId=hive_20241106182719_2d513ef1-f6ab-4a00-b3eb-c9dd3e
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

```

Query History Saved Queries Results (10)

	word	count
1	Æschines,	1
2	zigzag	1
3	zest	1
4	zenith	1
5	zealously	1
6	zealous,	1
7	zealous	5
8	zeal,	3
9	zeal	8

```

1 SELECT word, count(1) AS count FROM (SELECT explode(split(line, ' ')) AS word FROM docs) w
2 GROUP BY word
3 ORDER BY count DESC LIMIT 10;

```

```

INFO : Map 1: 1/1      Reducer 2: 2/2 Reducer 3: 0/1/1
INFO : Completed executing command(queryId=hive_20241106182906_a3aff338-7963-4945-812b-e90619a59bd7); Time taken: 12.100 sec
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

```

Query History Saved Queries Results (10)

	word	count
1	the	44647
2	of	28020
3		27298
4	to	23208
5	and	20444
6	in	13174
7	at	10000

Reto:

Cómo guardar en una tabla los valores del wordcount.

1

2

3

4

5

```
insert into storing
SELECT word, count(1) AS count FROM (SELECT explode(split(line, ' ')) AS word FROM docs) w
GROUP BY word
ORDER BY count;
```

INFO : Executing stats task

INFO : Table default.storing stats: [numFiles=1, numRows=38683, totalSize=438215, rawDataSize=399532]

INFO : Completed executing command(queryId=hive\_20241106183342\_dc3399b1-c222-44d7-8376-59b2af5c2a72); Time taken: 12.104 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

Success.

6

```
select * from storing;
```

INFO : Concurrency mode is disabled, not creating a lock manager

INFO : Executing command(queryId=hive\_20241106183417\_f0ac9619-5e0e-4fb2-87c6-86c142deaaab): select \* from storing

INFO : Completed executing command(queryId=hive\_20241106183417\_f0ac9619-5e0e-4fb2-87c6-86c142deaaab); Time taken: 0.001 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

Query History

Saved Queries

Results (100+)

storing.word

storing.count

1

Æschines,

1

2

"But

1

3

"KANSAS

1

4

"Nothing

1

5

"One

1

6

"PETERSBURG,

1

7

"--]

1