

PRACTICA HUE HIVE SANTIAGO QUISPE ABANTO 27 oct 2025

1. Descargar el Archivo de Datos

Primero, descarga el archivo CSV desde el repositorio de GitHub.

Descarga el archivo, por ejemplo, usando wget

wget <https://github.com/owid/co2-data/raw/master/owid-co2-data.csv>

2. Subir el Archivo a HDFS

Una vez descargado el archivo, súbelo al sistema de archivos distribuido HDFS para que [Hive](#) pueda acceder a él.

```
hdfs dfs -mkdir -p /curso/datos/co2
```

```
# Sube el archivo CSV a HDFS
```

```
hdfs dfs -put owid-co2-data.csv /curso/datos/co2
```

3. Crear la Tabla externa en [Hive](#) a partir de datos ya existentes

A continuación, crea una tabla en [Hive](#) que refleje la estructura del archivo CSV. Dado que el archivo tiene múltiples columnas, definiremos cada una con su tipo de datos correspondiente.

```
CREATE DATABASE IF NOT EXISTS bigdata;  
USE bigdata;
```

```
CREATE EXTERNAL TABLE IF NOT EXISTS owid_co2_data (  
  country STRING,  
  year INT,  
  iso_code STRING,  
  population DOUBLE,  
  gdp DOUBLE,  
  cement_co2 DOUBLE,  
  cement_co2_per_capita DOUBLE,  
  co2 DOUBLE,  
  co2_growth_abs DOUBLE,  
  co2_growth_prct DOUBLE,  
  co2_including_luc DOUBLE,  
  co2_including_luc_growth_abs DOUBLE,  
  co2_including_luc_growth_prct DOUBLE,  
  co2_including_luc_per_capita DOUBLE,
```

co2_including_luc_per_gdp DOUBLE,
co2_including_luc_per_unit_energy DOUBLE,
co2_per_capita DOUBLE,
co2_per_gdp DOUBLE,
co2_per_unit_energy DOUBLE,
coal_co2 DOUBLE,
coal_co2_per_capita DOUBLE,
consumption_co2 DOUBLE,
consumption_co2_per_capita DOUBLE,
consumption_co2_per_gdp DOUBLE,
cumulative_cement_co2 DOUBLE,
cumulative_co2 DOUBLE,
cumulative_co2_including_luc DOUBLE,
cumulative_coal_co2 DOUBLE,
cumulative_flaring_co2 DOUBLE,
cumulative_gas_co2 DOUBLE,
cumulative_luc_co2 DOUBLE,
cumulative_oil_co2 DOUBLE,
cumulative_other_co2 DOUBLE,
energy_per_capita DOUBLE,
energy_per_gdp DOUBLE,
flaring_co2 DOUBLE,
flaring_co2_per_capita DOUBLE,
gas_co2 DOUBLE,
gas_co2_per_capita DOUBLE,
ghg_excluding_lucf_per_capita DOUBLE,
ghg_per_capita DOUBLE,
land_use_change_co2 DOUBLE,
land_use_change_co2_per_capita DOUBLE,
methane DOUBLE,
methane_per_capita DOUBLE,
nitrous_oxide DOUBLE,
nitrous_oxide_per_capita DOUBLE,
oil_co2 DOUBLE,
oil_co2_per_capita DOUBLE,
other_co2_per_capita DOUBLE,
other_industry_co2 DOUBLE,
primary_energy_consumption DOUBLE,
share_global_cement_co2 DOUBLE,
share_global_co2 DOUBLE,
share_global_co2_including_luc DOUBLE,

```

share_global_coal_co2 DOUBLE,
share_global_cumulative_cement_co2 DOUBLE,
share_global_cumulative_co2 DOUBLE,
share_global_cumulative_co2_including_luc DOUBLE,
share_global_cumulative_coal_co2 DOUBLE,
share_global_cumulative_flaring_co2 DOUBLE,
share_global_cumulative_gas_co2 DOUBLE,
share_global_cumulative_luc_co2 DOUBLE,
share_global_cumulative_oil_co2 DOUBLE,
share_global_cumulative_other_co2 DOUBLE,
share_global_flaring_co2 DOUBLE,
share_global_gas_co2 DOUBLE,
share_global_luc_co2 DOUBLE,
share_global_oil_co2 DOUBLE,
share_global_other_co2 DOUBLE,
share_of_temperature_change_from_ghg DOUBLE,
temperature_change_from_ch4 DOUBLE,
temperature_change_from_co2 DOUBLE,
temperature_change_from_ghg DOUBLE,
temperature_change_from_n2o DOUBLE,
total_ghg DOUBLE,
total_ghg_excluding_lucf DOUBLE,
trade_co2 DOUBLE,
trade_co2_share DOUBLE
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/curso/datos/co2';

```

Notas:

- **EXTERNAL TABLE:** Utilizamos una tabla externa para que [Hive](#) no administre físicamente los datos. Los datos permanecen en HDFS incluso si la tabla se elimina.
- **Tipos de Datos:** Se han asignado tipos de datos básicos (STRING, INT, DOUBLE). Puedes ajustar estos tipos según la naturaleza de los datos para optimizar el almacenamiento y las consultas.
- **FIELDS TERMINATED BY ',':** Especifica que el delimitador de campos en el CSV es una coma.

4. Verificar la Tabla y los Datos

Después de crear la tabla, verifica que [Hive](#) la haya reconocido correctamente y que los datos se hayan cargado.

-- Mostrar las primeras 10 filas de la tabla

```
SELECT * FROM owid_co2_data LIMIT 10;
```

1. Captura de pantalla que muestre las 15 primeras filas de la tabla

Query History Saved Queries Results (15)						
	owid_co2_data.country	owid_co2_data.year	owid_co2_data.iso_code	owid_co2_data.population	owid_co2_data.gdp	owid_co
1	country	NULL	iso_code	NULL	NULL	NULL
2	Afghanistan	1750	AFG	2802560	NULL	0
3	Afghanistan	1751	AFG	NULL	NULL	0
4	Afghanistan	1752	AFG	NULL	NULL	0
5	Afghanistan	1753	AFG	NULL	NULL	0
6	Afghanistan	1754	AFG	NULL	NULL	0
7	Afghanistan	1755	AFG	NULL	NULL	0
8	Afghanistan	1756	AFG	NULL	NULL	0
9	Afghanistan	1757	AFG	NULL	NULL	0
10	Afghanistan	1758	AFG	NULL	NULL	0
11	Afghanistan	1759	AFG	NULL	NULL	0
12	Afghanistan	1760	AFG	2866255	NULL	0
13	Afghanistan	1761	AFG	NULL	NULL	0
14	Afghanistan	1762	AFG	NULL	NULL	0
15	Afghanistan	1763	AFG	NULL	NULL	0

-- Contar el número total de registros

```
SELECT COUNT(*) FROM owid_co2_data;
```

2. ¿Cuántas filas tiene la tabla?

```
1 SELECT COUNT(*) FROM owid_co2_data;
```

```
INFO : Ended JOB = job_1701352353571_0001
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulat
INFO : Total MapReduce CPU Time Spent: 6 seconds
INFO : Completed executing command(queryId=hadoop.
```

Query History

Saved Queries

Result

_c0



1	50192
---	-------


5. Realizar Consultas Básicas con HiveQL

A continuación, se presentan algunos ejemplos de consultas que puedes realizar sobre la tabla `owid_co2_data`.

3. Mostrar una captura de pantalla del resultado de cada consulta

a. Obtener las Emisiones Totales de CO₂ por País y Año

```
SELECT country, year, co2
FROM owid_co2_data
WHERE co2 IS NOT NULL
ORDER BY country, year;
```



	country	year	co2
1	Afghanistan	1949	0.015
2	Afghanistan	1950	0.084
3	Afghanistan	1951	0.092
4	Afghanistan	1952	0.092
5	Afghanistan	1953	0.106
6	Afghanistan	1954	0.106
7	Afghanistan	1955	0.154
8	Afghanistan	1956	0.183
9	Afghanistan	1957	0.293
10	Afghanistan	1958	0.33
11	Afghanistan	1959	0.385
12	Afghanistan	1960	0.414
13	Afghanistan	1961	0.491
14	Afghanistan	1962	0.689
15	Afghanistan	1963	0.707
16	Afghanistan	1964	0.839
17	Afghanistan	1965	1.007
18	Afghanistan	1966	1.001

b. Calcular el Promedio de Emisiones de CO₂ por País

```

SELECT country, AVG(co2) AS promedio_co2
FROM owid_co2_data
WHERE co2 IS NOT NULL
GROUP BY country
ORDER BY promedio_co2 DESC;

```

	country	promedio_co2
1	World	6614.439000000001
2	OECD (GCP)	5631.715844827587
3	Non-OECD (GCP)	4499.257729885058
4	High-income countries	4159.9862189781015
5	Europe (GCP)	3125.268913793104
6	Asia (GCP)	3010.8939655172408
7	Upper-middle-income countries	2901.4217586206896
8	North America (GCP)	2807.2870804597706
9	China	2329.3293846153847
10	Asia	2195.718529197081
11	North America	2077.1419539748954
12	Europe	2000.7230000000002
13	United States	1927.9139419642868
14	European Union (28)	1380.3948065693428
15	European Union (27)	1286.4244655172413
16	Asia (excl. China and India)	971.6252664233575
17	Europe (excl. EU-27)	911.4876897810217
18	Russia	705.0379534883722

c. Encontrar los 10 Países con Mayor Emisión de CO₂ en un Año Específico

Por ejemplo, para el año 2020:

```
SELECT country, co2
FROM owid_co2_data
WHERE year = 2020
ORDER BY co2 DESC
LIMIT 10;
```

	country	co2
1	Vatican	NULL
2	OECD (Jones et al.)	NULL
3	Least developed countries (Jones et al.)	NULL
4	Ryukyu Islands	NULL
5	Kuwaiti Oil Fires	NULL
6	Monaco	NULL
7	San Marino	NULL
8	International transport	NULL
9	World	35126.527
10	Non-OECD (GCP)	23102.908

d. Analizar la Relación entre GDP y Emisiones de CO₂ per Cápita

```
SELECT gdp, co2_per_capita
FROM owid_co2_data
WHERE gdp IS NOT NULL AND co2_per_capita IS NOT NULL;
```

	gdp	co2_per_capita
1	9421400064	0.011
2	9692279808	0.012
3	10017325056	0.011
4	10630519808	0.013
5	10866360320	0.013
6	11078185984	0.018
7	11581236224	0.022
8	11578973184	0.034
9	12238841856	0.038
10	12579874816	0.043
11	13033253888	0.046
12	13146287104	0.053
13	13367633920	0.073
14	13630298112	0.074
15	13870503936	0.085

e. Obtener las Emisiones Totales de CO₂ de ESPAÑA en 2015

```
1 SELECT co2
2 FROM owid_co2_data
3 WHERE country = 'Spain' and year = 2015
```

```
INFO : Compiling command(queryId=hadoop_202510270833
FROM owid_co2_data
WHERE country = 'Spain' and year = 2015
INFO : No Stats for default@owid_co2_data, Columns:
INFO : Semantic Analysis Completed (astId=1, false)
```

Query History

Saved Queries

Results

co2



1	270.74
---	--------

f. Encontrar los 5 Países con Mejor Emisión de CO₂ en 2018

```
SELECT country, co2
FROM owid_co2_data
WHERE year = 2018 and co2 is not NULL
ORDER BY co2 DESC LIMIT 5
```

```
INFO : Compiling command(queryId=hadoop_20251027084131_30099cfd-465d-44ac-a9a9-82c0024033fa):
```

Query History

Saved Queries

Results (5)

country

co2

1	World	36730.426
2	Non-OECD (GCP)	22884.207
3	Asia	20335.523
4	Asia (GCP)	17738.896
5	Upper-middle-income countries	15855.511