EAFIT

Special topics in telematics

Laboratory 4

Luisa María Álvarez García – Computer Science

(lmalvarez8@eafit.edu.co)

Teacher: Edwin Nelson Montoya

Medellin, November 11, 2024

1. Into the hive editor in hue, create the usernamedb in case it does not exist, using the command:

```
CREATE DATABASE usernamedb;
Y usarla
USE usernamedb;
SHOW DATABASES;
```

2. After that, create the table HDI, using the command:

```
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;
```

3. And save the information uploaded the last workshop into the onu folder in aws, using the command:

```
LOAD DATA INPATH '/user/hadoop/datasets/onu/hdi-data.csv' INTO TABLE HDI;
```

4. After that, it has to look like this:

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

```
INFO  : Executing command(queryId=hive_20241117184738_121a60cb-35e6-4577-81bc-b1d926ed0ac4): select * from hdi
INFO  : Completed executing command(queryId=hive_20241117184738_121a60cb-35e6-4577-81bc-b1d926ed0ac4); Time taken: 0.001 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
```

Query History | Saved Queries | Results (100+)

	hdi.id	hdi.country	hdi.hdi	hdi.lifeex	hdi.mysch	hdi.eysch	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
7	6	Canada	0.908	81	12	16	35166
8	7	Ireland	0.908	80	11	18	29322
9	8	Liechtenstein	0.905	79	10	14	83717
10	9	Germany	0.905	80	12	15	34854
11	10	Sweden	0.904	81	11	15	35837
12	11	Switzerland	0.903	82	11	15	39924
13	12	Japan	0.901	83	11	15	32295

5. To upload the information into s3, run:

```
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 's3://jupyterbuck/onu/hdi/';
```

6. To make queries first, create the table where the queries will be saved:

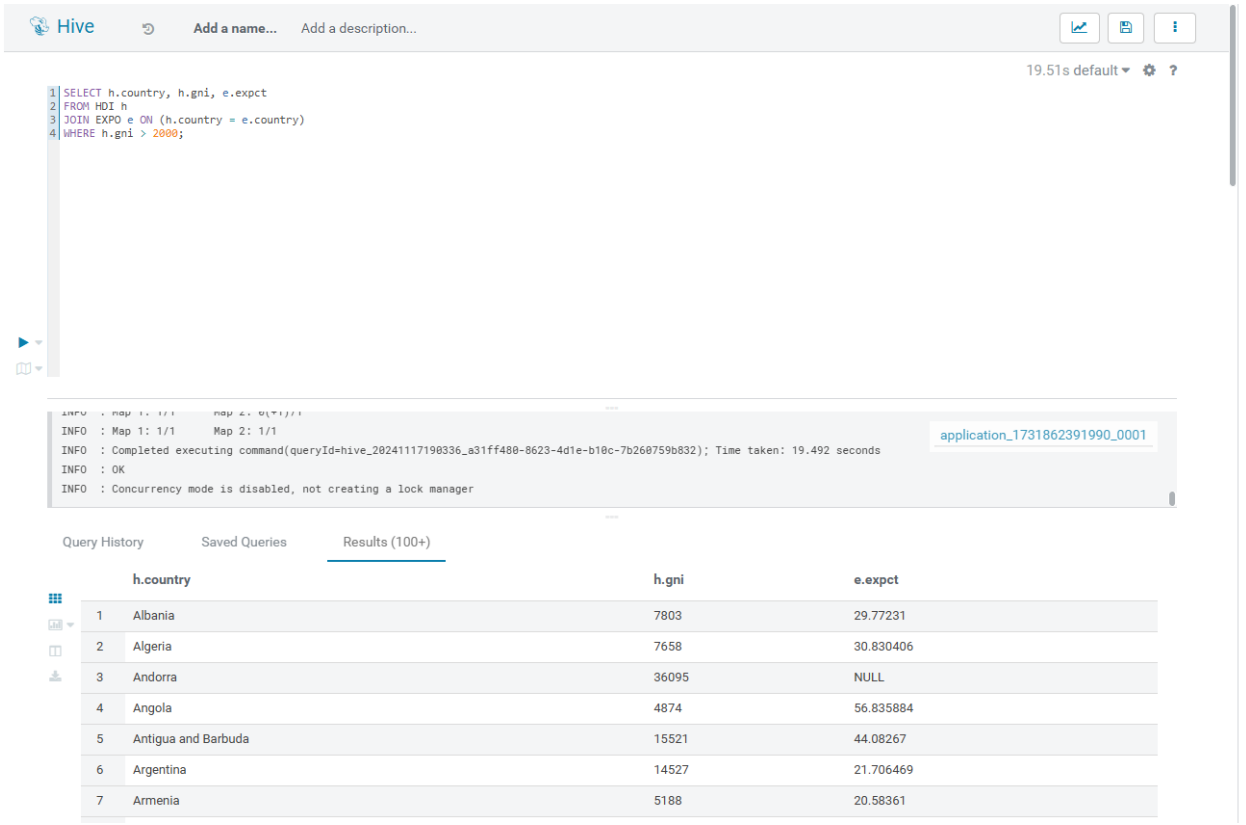
```
USE usernamedb;

CREATE EXTERNAL TABLE EXPO (
  country STRING,
  expct FLOAT
```

```
        )
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY ','
        STORED AS TEXTFILE
LOCATION 's3://jupyterbuck/onu/export/';
```

7. To create a join between two tables, run :

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

That shows:



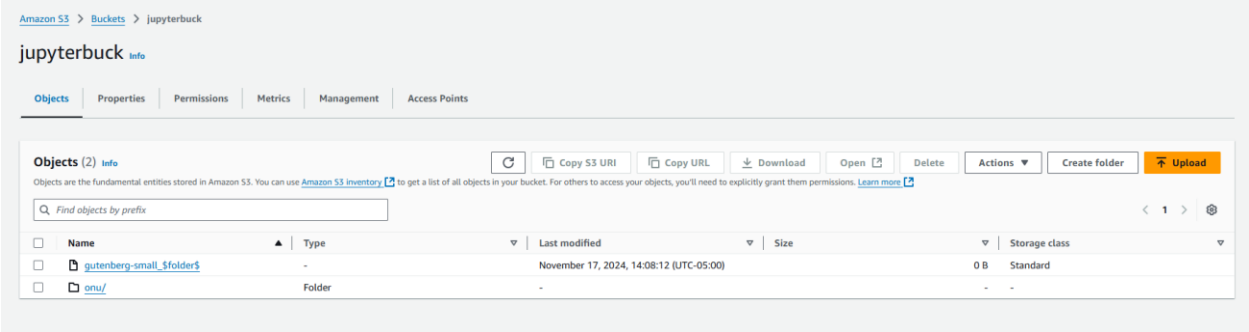
	h.country	h.gni	e.expct
1	Albania	7803	29.77231
2	Algeria	7658	30.830406
3	Andorra	36095	NULL
4	Angola	4874	56.835884
5	Antigua and Barbuda	15521	44.08267
6	Argentina	14527	21.706469
7	Armenia	5188	20.58361

8. To realize the wordcount on hive, first create the table, running:

```
CREATE EXTERNAL TABLE docs (line STRING)
STORED AS TEXTFILE
LOCATION 's3://emontoyadatasets/gutenberg-small/';

SELECT word, count(1) AS count FROM (SELECT explode(split(line,' ')) AS word FROM
docs) w
GROUP BY word
```

```
ORDER BY word DESC LIMIT 10;

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

That will create a s3 file like this:



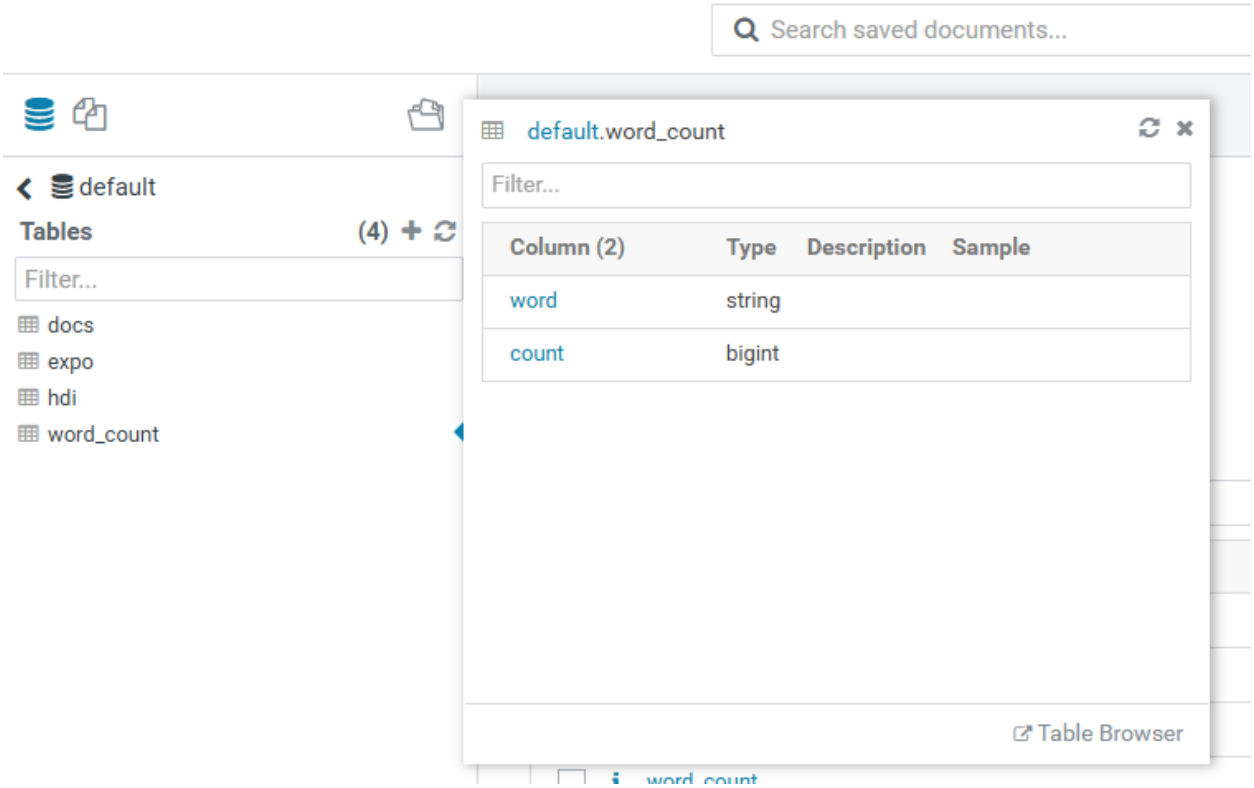
9. Now, to store the output into the *word_count* table, just run:

```
CREATE TABLE word_count (
  word STRING,
  count INT
)
STORED AS TEXTFILE;

INSERT INTO TABLE word_count
SELECT word, count(1) AS count
FROM (SELECT explode(split(line,' ')) AS word FROM docs) w
GROUP BY word
ORDER BY count DESC;
```

10. To create directly the table, run:

```
CREATE TABLE word_count AS
SELECT word, count(1) AS count
FROM (SELECT explode(split(line,' ')) AS word FROM docs) w
GROUP BY word
ORDER BY count DESC;
Y para verificar
SELECT * FROM word_count LIMIT 10;
```

It will show something like:


Search saved documents...
default
Tables
(4)
Filter...
docs
expo
hdi
word_count
default.word_count
Filter...
Column (2)
Type
Description
Sample
word
string
count
bigint
Table Browser