

UNIVERSIDAD EAFIT
ST1800/ST1801 ALMACENAMIENTO Y RECUPERACIÓN DE INFORMACIÓN
2024-2

Nombre: Andrés Felipe Téllez Rodríguez - email: aftellezr@eafit.edu.co

Laboratorios de clase:

Curso en AWS ACADEMY:

Evidencias de la realización del lab2

1. Evidencias de HDFS y S3 desde EMR

<input type="checkbox"/>	Name	Size	Usuario	Group	Permisos	Date
<input type="checkbox"/>	└		hdfs	hdfsadmingroup	drwxr-xr-x	October 23, 2024 06:58 PM
<input type="checkbox"/>	.		hadoop	hdfsadmingroup	drwxrwxrwx	October 23, 2024 07:06 PM
<input type="checkbox"/>	datasets		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	└		hadoop	hdfsadmingroup	drwxrwxrwx	October 23, 2024 07:06 PM
<input type="checkbox"/>	.		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input checked="" type="checkbox"/>	airlines.csv	761,8 KB	hadoop	hdfsadmingroup	-rw-r--r--	October 23, 2024 07:34 PM
<input type="checkbox"/>	all-news		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:33 PM
<input type="checkbox"/>	covid19		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	flights		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	gutenberg		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	gutenberg-small		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:23 PM
<input type="checkbox"/>	onu		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	otros		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	retail_logs		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM
<input type="checkbox"/>	sample_data.csv	534 bytes	hadoop	hdfsadmingroup	-rw-r--r--	October 23, 2024 07:34 PM
<input type="checkbox"/>	spark		hadoop	hdfsadmingroup	drwxr-xr-x	October 23, 2024 07:34 PM

2. Evidencias de Hive desde EMR

The screenshot shows the Hive web interface. On the left, a sidebar lists tables, including 'hue_tmp_airlines' with its schema: id (string), airline (string), date (string), location (string), rating (string), cabin (string), value (string), recommended (string), and review (string). The main panel displays a query: `SELECT * FROM hue_tmp_airlines;`. Below the query, the execution status is shown as 'OK' with a completion time of 0.8 seconds. The 'Results (100+)' tab is active, showing a table with 11 rows and 6 columns: hue_tmp_airlines.id, hue_tmp_airlines.airline, hue_tmp_airlines.date, hue_tmp_airlines.location, hue_tmp_airlines.rating, and hue_tmp_airlines.cabin. The data includes flights from Delta Air Lines to various locations like Thailand, USA, Ecuador, and UK, with ratings and cabin classes (Economy, Business).

hue_tmp_airlines.id	hue_tmp_airlines.airline	hue_tmp_airlines.date	hue_tmp_airlines.location	hue_tmp_airlines.rating	hue_tmp_airlines.cabin
1	10001	Delta Air Lines	21-Jun-14	Thailand	7
2	10002	Delta Air Lines	19-Jun-14	USA	0
3	10003	Delta Air Lines	18-Jun-14	USA	0
4	10004	Delta Air Lines	17-Jun-14	USA	9
5	10005	Delta Air Lines	17-Jun-14	Ecuador	7
6	10006	Delta Air Lines	17-Jun-14	USA	9
7	10007	Delta Air Lines	14-Jun-14	UK	0
8	10008	Delta Air Lines	14-Jun-14	USA	0
9	10009	Delta Air Lines	13-Jun-14	USA	4
10	10010	Delta Air Lines	13-Jun-14	UK	9
11	10011	Delta Air Lines	11-Jun-14	USA	10

3. Evidencias de spark desde EMR y Google Colab

In [6]: spark

```
<pyspark.sql.session.SparkSession object at 0x7f44fc5b2d90>
```

In [5]: sc

```
<SparkContext master=yarn appName=livy-session-0>
```

In [3]: result = spark.sql("SELECT * FROM hue_tmp_airlines LIMIT 10")
result.show()

```
+-----+-----+-----+-----+-----+-----+-----+-----+
| id | airline | date | location | rating | cabin | value | recommended | review |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 10001 | Delta Air Lines | 21-Jun-14 | Thailand | 7 | Economy | 4 | YES | Flew Mar 30 NRT t... |
| 10002 | Delta Air Lines | 19-Jun-14 | USA | 0 | Economy | 2 | NO | Flight 2463 leavi... |
| 10003 | Delta Air Lines | 18-Jun-14 | USA | 0 | Economy | 1 | NO | Delta Website fro... |
| 10004 | Delta Air Lines | 17-Jun-14 | USA | 9 | Business | 4 | YES | I just returned f... |
| 10005 | Delta Air Lines | 17-Jun-14 | Ecuador | 7 | Economy | 3 | YES | Round-trip flight... |
| 10006 | Delta Air Lines | 17-Jun-14 | USA | 9 | Business | 5 | YES | Narita - Bangkok ... |
| 10007 | Delta Air Lines | 14-Jun-14 | UK | 0 | Economy | 1 | NO | Flight from NY La... |
| 10008 | Delta Air Lines | 14-Jun-14 | USA | 0 | Economy | 1 | NO | Originally I had ... |
| 10009 | Delta Air Lines | 13-Jun-14 | USA | 4 | Business | 2 | NO | We flew paid busi... |
+-----+-----+-----+-----+-----+-----+-----+-----+
```