Universidad Francisco Marroquín Facultad de Ciencias Computer Science Big Data Primer Semestre 2020

Sophia Gamarro: 20170356

**FORMAT** 

# Hoja de Trabajo No. 6

6.

[cloudera@quickstart ~]\$ hdfs dfs -put /home/cloudera/Desktop/ejercicios/data/ra tings\_2012.txt analyst/dualcore/ [cloudera@quickstart ~]\$ hdfs dfs -put /home/cloudera/Desktop/ejercicios/data/ra tings\_2013.txt analyst/dualcore/ [cloudera@quickstart ~]\$ ■

7. Search data

Jobs 

Ouery

Search data

Jobs 

Ouery

Search data

Jobs 

Ouery

Indicate the search data

Indicate

#### Load data in analyst.ratings INFO : Loading data to table analyst.ratings from hdfs://quickstart.cloudera:8 020/user/cloudera/analyst/dualcore/ratings\_2012.txt INFO : Starting task [Stage-1:STATS] in serial mode INFO : Table analyst.ratings stats: [numFiles=1, totalSize=27007] INFO : Completed executing command(queryId=hive\_20200306162121\_67759ede-9438-4 1c8-933d-7f1548d6707c); Time taken: 0.98 seconds INFO : OK Load data in analyst.ratings 3 minutes ago Creating table analyst.ratings 5 minutes ago Drop table ratings 10 minutes ago Creating table analyst.ratings Load data in analyst.ratings INFO : Loading data to table analyst.ratings from hdfs://quickstart.cloudera:8 020/user/cloudera/analyst/dualcore/ratings\_2013.txt INFO : Starting task [Stage-1:STATS] in serial mode INFO : Table analyst.ratings stats: [numFiles=2, totalSize=1267251] INFO : Completed executing command(queryId=hive\_20200306162222\_85a31c50-9e66-4 7f4-8cf1-258bc86f6f02); Time taken: 0.536 seconds Load data in analyst.ratings a minute ago Load data in analyst.ratings 3 minutes ago Creating table analyst.ratings 6 minutes ago Drop table ratings

10.

•	posted	cust_id	prod_id	rating	message
1	2012-05-21 12:52:48	1043182	1274362	5	This is truly fantastic!
2	2012-10-14 01:36:07	1242853	1273879	2	The product quality was OK
3	2012-10-14 02:41:50	1047430	1273799	2	Shoddy quality
4	2012-10-14 10:10:05	1087455	1274476	4	Quality was passable
5	2012-10-14 10:42:41	1170230	1273964	2	It was OK
6	2012-10-14 19:12:33	1063130	1274734	4	It was OK
7	2012-10-14 22:00:56	1031378	1274616	4	Quality was passable
8	2012-10-15 00:27:47	1203215	1273850	5	Awesome product
9	2012-10-15 01:14:26	1135616	1274218	4	Value of product was just alright

#### 11 y 14. \*El resultado ya incluye ratings 2012 y 2013

count(\*)





21997

1<u>5.</u>

```
Os default * text * * ?

Os default * text * ?

Os default * ?

Os default
```

SELECT job\_title, COUNT(\*) AS num
FROM employees
GROUP BY job\_title

hive> sqoop import --connect jdbc:mysql://localhost/analyst\_dualcore --username root --pa ssword cloudera --fields-terminated-by '\t' --warehouse-dir analyst\dualcore --table supp liers--hive-import--create-hive-table--hive-table analyst.suppliers

0K

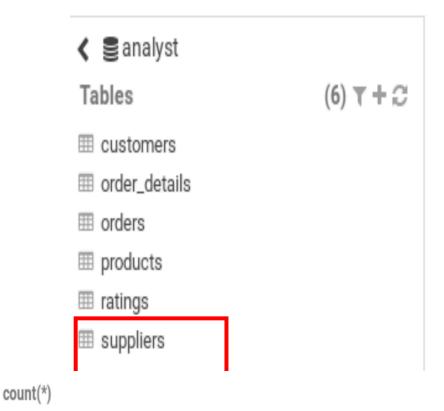
Time taken: 4.76 seconds

Loading data to table analyst.suppliers

Table analyst.suppliers stats: [numFiles=4, totalSize=6060]

0K

Time taken: 0.544 seconds\_



### 19. ALTER TABLE analyst.suppliers CHANGE company na

■ suppliers supp\_id (int) name (string)

20.

1	supp_id	int
2	name	string
3	contact	string
4	address	string
5	city	string
6	state	string
7	zipcode	string
8	phone	string

21. ALTER TABLE suppliers RENAME TO vendors

## SELECT supp\_id, name FROM vendors LIMIT 10;

	supp_id	name
1	1000	Smithco Equipment Manufacturing
2	1001	Simian Tech Systems, Inc.
3	1002	Division Zero Corp.
4	1003	General Diode Co.
5	1004	JPD Companies
6	1005	Harvey Mettle Electrical Supply Co.
7	1006	The Connector Company