

PRÁCTICA Nº 2.4

Hive

Apache Hive



- Nombre y apellidos: Alvaro Lucio-Villegas de Cea



Índice

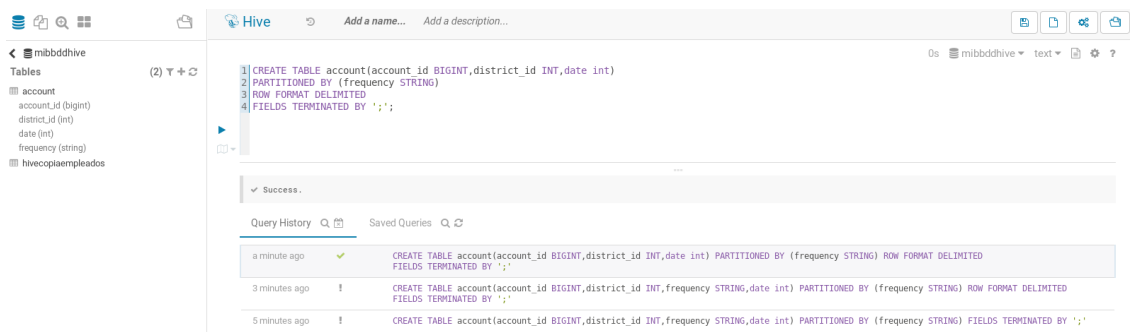
Ejercicio 1	3
Ejercicio 2	4
Ejercicio 3	7

Ejercicio 1

Carga una tabla interna con los datos account.asc de forma que queden fragmentados por medio del campo “frequency”. Revisa en el sistema de archivos como queda esto físicamente y coméntalo.

Creamos la tabla particionada:

```
CREATE TABLE account(account_id BIGINT , district_id INT,date INT)
PARTICIONED BY (frequency STRING)
FIELDS TERMINATED BY “;”;
```



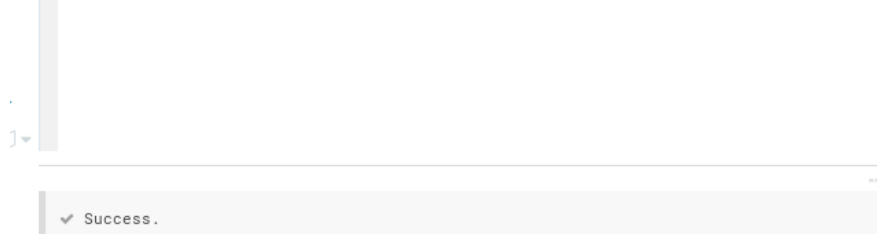
Para poder cargar la tabla particionada necesitamos hacer una tabla temporal donde cargaremos todos los datos y luego la cargaremos en la tabla final.

Creamos la tabla

```
1 CREATE TABLE account_tmp(account_id INT,district_id INT,frequency STRING,date INT)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ';'
3 tblproperties("skip.header.line.count"="1");
4
```

Cargamos los datos desde el fichero.

```
1 LOAD DATA INPATH '/account.asc.tmp' INTO TABLE account_tmp;
```



Una vez tenemos las tablas sobreescribimos la original con los datos de la temporal.

```
1|INSERT OVERWRITE TABLE account PARTITION(frequency) SELECT account_id,district_id,date,frequency FROM account_tmp
```

Success.

Ejercicio 2

Carga como tablas externas los ficheros clients.asc y disp.asc.

clients.asc

Creamos la tabla.

```
1|CREATE EXTERNAL TABLE clients(client_id INT,birth_number STRING,district_id INT)
2|ROW FORMAT DELIMITED FIELDS TERMINATED BY ';'
3|tblproperties("skip.header.line.count"="1");
4|
```

Success.

Cargamos los datos.

```
1|LOAD DATA INPATH '/client.asc' INTO TABLE clients;
```

Success.

Comprobamos que se encuentran los datos correctamente en la tabla.

mibbddhive.clients			
Columns Details Sample			
	clients.client_id	clients.birth_number	clients.district_id
1	1	"706213"	18
2	2	"450204"	1
3	3	"406009"	1
4	4	"561201"	5
5	5	"605703"	5
6	6	"190922"	12
7	7	"290125"	15
8	8	"385221"	51

Table Browser

disp.asc.

Creamos la tabla.

```

1 CREATE EXTERNAL TABLE disp(dispatch_id INT,client_id INT,account_id INT,type STRING)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
3 tblproperties("skip.header.line.count"="1");
4

```

Success.

Cargamos los datos en la tabla.

```

1 LOAD DATA INPATH '/disp.asc' INTO TABLE disp;

```

Success.



Comprobamos que se encuentran los datos correctamente en la tabla.

mibbddhive.disp

Columns Details Sample

	disp.disp_id	disp.client_id	disp.account_id	disp.type
1	1	1	1	"OWNER"
2	2	2	2	"OWNER"
3	3	3	2	"DISPONENT"
4	4	4	3	"OWNER"
5	5	5	3	"DISPONENT"
6	6	6	4	"OWNER"
7	7	7	5	"OWNER"
8	8	8	6	"OWNER"

Ejercicio 3

Haz una consulta que devuelva el código de cliente y su cuenta asociada. Estas tablas se conectan a través de la tabla N:M disp.asc

clients.asc: "client_id";"birth_number";"district_id"

disp.asc: "disp_id";"client_id";"account_id";"type"

accounts.asc: "account_id";"district_id";"frequency";"date"

Guarda el resultado de esta consulta como un nuevo archivo de texto

La consulta sería así:

```
SELECT clients_asc.client_id, accounts_asc.account_id
FROM clients
INNER JOIN disp ON clients_asc.client_id = disp.asc.client_id
INNER JOIN accounts ON disp.account_id = accounts.account_id;
```

The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo and options to add a name or description. Below this, the 'Query File' section shows the path '/consulta_CodCliente-CodCuenta.asc'. The query editor contains the following SQL code:

```
1 SELECT clients.client_id,account.account_id FROM clients
2 INNER JOIN disp ON clients.client_id = disp.client_id
3 INNER JOIN account ON disp.account_id = account.account_id;
```

Below the query editor, the 'Results (100+)' tab is selected, displaying a table with two columns: 'clients.client_id' and 'account.account_id'. The table contains 12 rows of data, showing a one-to-one relationship between client IDs and account IDs.

clients.client_id	account.account_id
1	1
2	2
3	2
4	3
5	3
6	4
7	5
8	6
9	7
10	8
11	8
12	9