



CURSO DE ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL Y BIGDATA

BIGDATA APLICADO

TAREA EVALUABLE 4.1

Autor: Carlos Sánchez Recio.
06 / 01 / 2025

Índice

Apartado 2: Bibliotecas		1
1	Recupera el título de todos los libros publicados en el siglo XX.	3
2	Recupera el título de los libros del tema "Historia".	4
3	Recupera el total de ejemplares disponibles en la biblioteca de Levante. .	5
4	Recupera el título de los libros de Ficción disponibles en la biblioteca del Centro.	6

Apartado 2: Bibliotecas

Tenemos los siguientes datos sobre libros y bibliotecas:

ID	Info_Libro	Temas	Ejemplares_Biblioteca
1	{"Título": "1984", "Autor": "George Orwell", "Año": 1949}	"Ficción", "Distopía", "Sociedad"	{"Centro": 5, "Levante": 2}
2	{"Título": "Sapiens", "Autor": "Yuval Noah Harari", "Año": 2011}	"Ensayo", "Historia", "Antropología", "Sociedad"	{"Levante": 4, "Ponente": 3}
3	{"Título": "Dune", "Autor": "Frank Herbert", "Año": 1965}	"Ficción", "Aventura", "Ciencia-ficción"	{"Centro": 7, "Ponente": 2}
4	{"Título": "El Señor de los anillos", "Autor": "JRR Tolkien", "Año": 1954}	"Ficción", "Aventura", "Fantasía"	{"Centro": 8, "Levante": 3}
5	{"Título": "Historia de dos ciudades", "Autor": "Charles Dickens", "Año": 1859}	"Ficción", "Historia", "Drama"	{"Levante": 2}

Cada libro tiene una información (título, autor y año de publicación) un listado de temas y una información sobre los ejemplares disponibles en cada una de las tres bibliotecas de la ciudad (Centre, Llevant y Ponent).

Crea una tabla en el almacén de datos de Hive, utilizando los tipos complejos STRUCT, ARRAY y MAP. Carga los datos ejecutando sentencias INSERT . Ejecuta las siguientes consultas en Hive (con HiveQL) e Impala (con SQL):

Como en anteriores apartados, el primer paso será crear una base de datos y las tablas necesarias para almacenar los datos, además de insertarlos. Para este apartado de la práctica, no se trabaja con archivos por lo que no es necesario usar comandos de Hadoop.

```

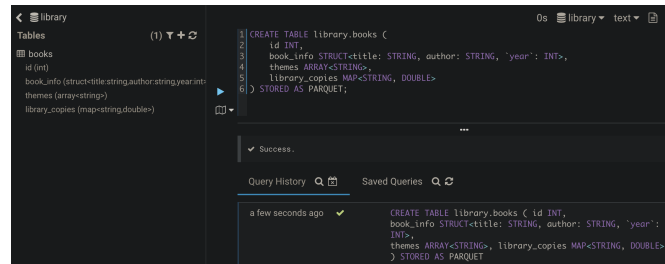
1 CREATE DATABASE library;
2
3 USE library;
4
5 CREATE TABLE library.books (
6     id INT,
7     book_info STRUCT<title: STRING, author: STRING, 'year': INT>,
8     themes ARRAY<STRING>,
9     library_copies MAP<STRING, DOUBLE>
10 ) STORED AS PARQUET;
11
12 INSERT INTO library.books
13 SELECT 2,
14     NAMED_STRUCT('title', 'Sapiens', 'author', 'Yuval Noah Harari', 'year', 2011),
15     ARRAY('Assaig', 'Hist ria', 'Antropologia', 'Societat'),
16     MAP('Llevant', 4.0, 'Ponent', 3.0);

```

```

17
18 INSERT INTO library.books
19     SELECT 2,
20     NAMED_STRUCT('title', 'Sapiens', 'author', 'Yuval Noah Harari', 'year', 2011),
21     ARRAY('Assaig', 'Història', 'Antropologia', 'Societat'),
22     MAP('Llevant', 4.0, 'Ponent', 3.0);
23
24 INSERT INTO library.books
25     SELECT 3,
26     NAMED_STRUCT('title', 'Dune', 'author', 'Frank Herbert', 'year', 1965),
27     ARRAY('Ficció', 'Aventura', 'Ciència-ficció'),
28     MAP('Centre', 7.0, 'Ponent', 2.0);
29
30 INSERT INTO library.books
31     SELECT 4,
32     NAMED_STRUCT('title', 'El Senyor dels anells', 'author', 'J.R.R. Tolkien',
33     'year', 1954),
34     ARRAY('Ficció', 'Aventura', 'Fantasia'),
35     MAP('Centre', 8.0, 'Llevant', 3.0);
36
37 INSERT INTO library.books
38     SELECT 5,
39     NAMED_STRUCT('title', 'Història de dues ciutats', 'author', 'Charles Dickens',
40     'year', 1859),
41     ARRAY('Ficció', 'Història', 'Drama'),
42     MAP('Llevant', 2.0);

```



a few seconds ago	✓	SELECT * FROM books
a minute ago	✓	INSERT INTO library.books SELECT 5, NAMED_STRUCT('title', 'Història de dues ciutats', 'author', 'Charles Dickens', 'year', 1859), ARRAY('Ficció', 'Història', 'Drama'), MAP('Llevant', 2.0)
2 minutes ago	✓	INSERT INTO library.books SELECT 4, NAMED_STRUCT('title', 'El Senyor dels anells', 'author', 'J.R.R. Tolkien', 'year', 1954), ARRAY('Ficció', 'Aventura', 'Fantasia'), MAP('Centre', 8.0, 'Llevant', 3.0)
2 minutes ago	✓	INSERT INTO library.books SELECT 3, NAMED_STRUCT('title', 'Dune', 'author', 'Frank Herbert', 'year', 1965), ARRAY('Ficció', 'Aventura', 'Ciència-ficció'), MAP('Centre', 7.0, 'Ponent', 2.0)
3 minutes ago	✓	INSERT INTO library.books SELECT 2, NAMED_STRUCT('title', 'Sapiens', 'author', 'Yuval Noah Harari', 'year', 2011), ARRAY('Assaig', 'Història', 'Antropologia', 'Societat'), MAP('Llevant', 4.0, 'Ponent', 3.0)
11 minutes ago	✓	INSERT INTO library.books SELECT 1, NAMED_STRUCT('title', '1984', 'author', 'George Orwell', 'year', 1949), ARRAY('Ficció', 'Distopia', 'Societat'), MAP('Centre', 5.0, 'Llevant', 2.0)
12 minutes ago	!	INSERT INTO library.books SELECT 1, NAMED_STRUCT('title', '1984', 'author', 'George Orwell', 'year', 1949), ARRAY('Ficció', 'Distopia', 'Societat'), MAP('Centre', 5, 'Llevant', 2);
17 minutes ago	✓	CREATE TABLE library.books (id INT, book_info STRUCT<title: STRING, author: STRING, 'year': INT>, themes ARRAY<STRING>, library_copies MAP<STRING, DOUBLE>) STORED AS PARQUET

books.id	books.book_info	books.themes
1	["title":"1984","author":"George Orwell","year":1949]	["Ficció","Distopia","Societat"]
2	["title":"Sapiens","author":"Yuval Noah Harari","year":2011]	["Assaig","Història","Antropologia","Societat"]
3	["title":"Dune","author":"Frank Herbert","year":1965]	["Ficció","Aventura","Ciència-Ficció"]
4	["title":"El Senyor dels anells","author":"J.R.R. Tolkien","year":1954]	["Ficció","Aventura","Fantasia"]
5	["title":"Història de dues ciutats","author":"Charles Dickens","year":1859]	["Ficció","Història","Drama"]

NOTA: Para el número de copias de los ejemplares he utilizado un tipo DOUBLE por equivocación, ya que debería haber sido un tipo INT. No afecta en absoluto al funcionamiento pero a la hora de insertar datos he tenido que escribir los números en formato de coma flotante.

En los siguientes comandos es importante mencionar que primero ejecuté los de Hive y luego los de Impala, por lo que tras finalizar los primeros, ejecuté el siguiente comando¹

```
1 INVALIDATE METADATA library.books;
```

1 Recupera el título de todos los libros publicados en el siglo XX.

- Hive:

```
1 SELECT book_info.title
2 FROM library.books
3 WHERE book_info.`year` >= 1900 AND book_info.`year` < 2000;
```

title
1 1984
2 Dune
3 El Senyor dels anells

- Impala:

```
1 SELECT book_info.title
2 FROM library.books
3 WHERE book_info.year >= 1900 AND book_info.`year` < 2000;
```

¹No estoy seguro de que fuera necesario pero lo hice por si acaso.

```
1 SELECT book_info.title
2 FROM library.books
3 WHERE book_info.year >= 1900 AND book_info.year < 2000;
```

Query History Saved Queries Results (3)

	book_info.title
1	1984
2	Dune
3	El Senyor dels anells

2 Recupera el título de los libros del tema "Historia".

- Hive:

```
1 SELECT book_info.title
2 FROM library.books
3 WHERE array_contains(themes, 'Hist ria');
```

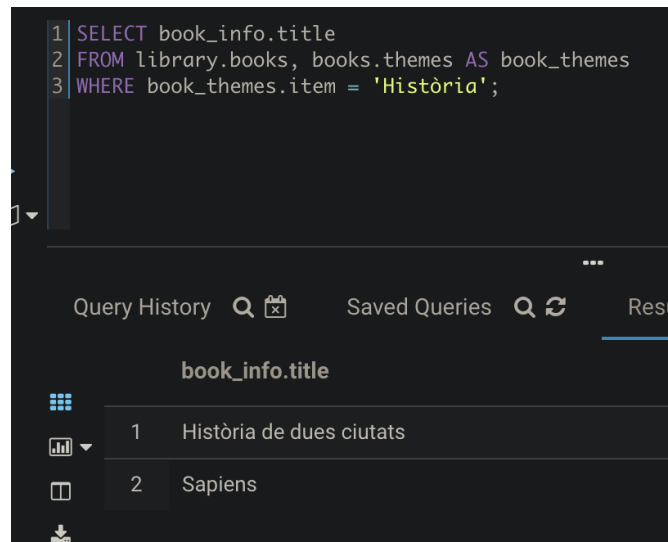
```
1 SELECT book_info.title
2 FROM library.books
3 WHERE array_contains(themes, 'Història');
```

Query History Saved Queries Results (2)

	title
1	Sapiens
2	Història de dues ciutats

- Impala:

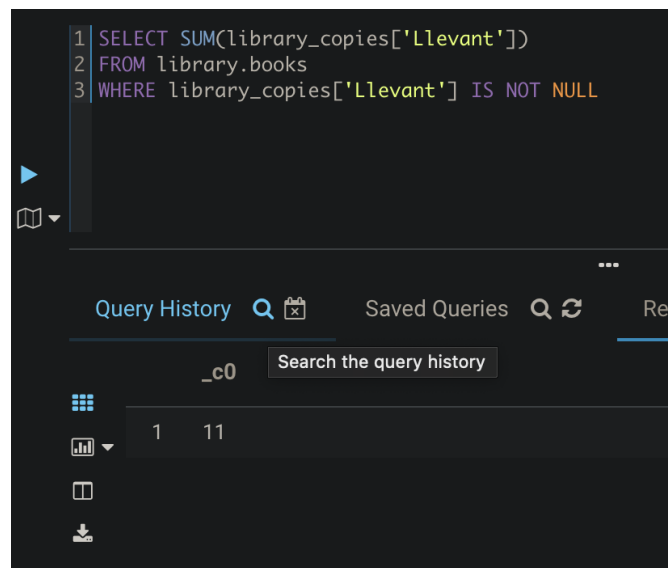
```
1 SELECT book_info.title
2 FROM library.books, books.themes AS book_themes
3 WHERE book_themes.item = 'Hist ria';
```



3 Recupera el total de ejemplares disponibles en la biblioteca de Levante.

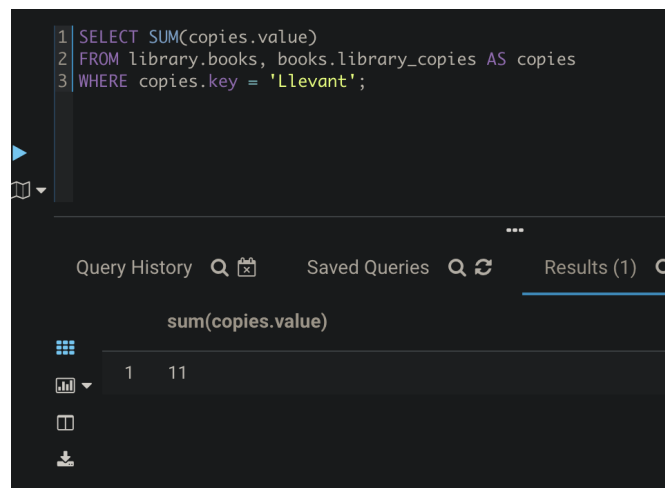
- Hive:

```
1 SELECT SUM(library_copies['Llevant'])
2 FROM library.books
3 WHERE library_copies['Llevant'] IS NOT NULL
```



- Impala:

```
1 SELECT SUM(copies.value)
2 FROM library.books, books.library_copies AS copies
3 WHERE copies.key = 'Llevant';
```



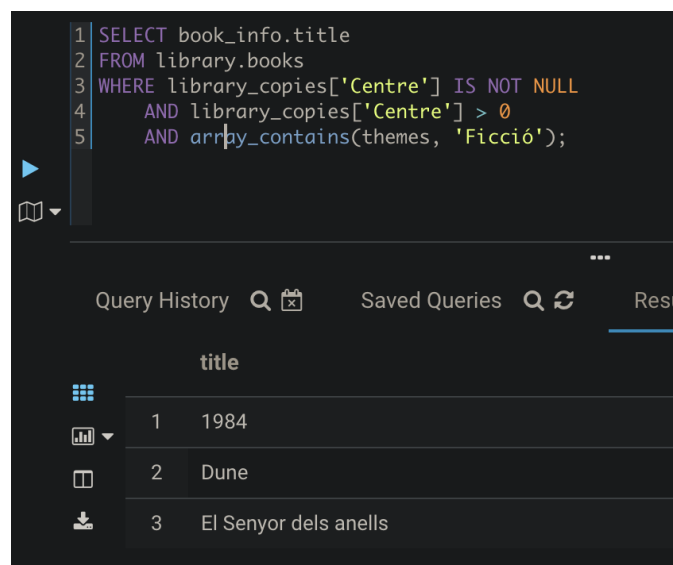
4 Recupera el título de los libros de Ficción disponibles en la biblioteca del Centro.

- Hive:

```

1 SELECT book_info.title
2 FROM library.books
3 WHERE library_copies['Centre'] IS NOT NULL
4       AND library_copies['Centre'] > 0
5       AND array_contains(themes, 'Ficci ');

```



- Impala:

```

1 SELECT book_info.title
2 FROM library.books, books.library_copies AS copies, books.themes AS book_themes
3 WHERE copies.key = 'Centre'
4       AND copies.value > 0
5       AND book_themes.item = 'Ficci ';

```



```
1 SELECT book_info.title
2 FROM library.books, books.library_copies AS copies, books.themes AS book_themes
3 WHERE copies.key = 'Centre'
4     AND copies.value > 0
5     AND book_themes.item = 'Ficció';
```

Query History 🔍 📄 Saved Queries 🔍 🔄 Results (3) 🔍 📄

	book_info.title
1	El Senyor dels anells
2	1984
3	Dune