

Tarea evaluable

CE_5075 4.1

Big data aplicado



Índice

APARTADO 2 - Bibliotecas	2
1 IMPORTAR DATOS	3
2 CONSULTAS EN HIVE	5
1. Recupera el título de todos los libros publicados en el siglo XX.	5
2. Recupera el título de los libros del tema “Historia”.	5
3. Recupera el total de ejemplares disponibles en la biblioteca de Llevant.	6
4. Recupera el título de los libros de Ficción disponibles en la biblioteca del Centre.	6
3 CONSULTAS EN IMPALA	7
1. Recupera el título de todos los libros publicados en el siglo XX.	7
2. Recupera el título de los libros del tema “Historia”.	7
3. Recupera el total de ejemplares disponibles en la biblioteca de Llevant.	8
4. Recupera el título de los libros de Ficción disponibles en la biblioteca del Centre.	8

APARTADO 2 - Bibliotecas

Tenemos los siguientes datos sobre librerías y bibliotecas:

ID	Info_Llibre	Temes	Exemplars_Biblioteca
1	{"Títol": "1984", "Autor": "George Orwell", "Any": 1949}	["Ficció", "Distopia", "Societat"]	{"Centre": 5, "Llevant": 2}
2	{"Títol": "Sapiens", "Autor": "Yuval Noah Harari", "Any": 2011}	["Assaig", "Història", "Antropologia", "Societat"]	{"Llevant": 4, "Ponent": 3}
3	{"Títol": "Dune", "Autor": "Frank Herbert", "Any": 1965}	["Ficció", "Aventura", "Ciència-ficció"]	{"Centre": 7, "Ponent": 2}
4	{"Títol": "El Senyor dels anells", "Autor": "J.R.R. Tolkien", "Any": 1954}	["Ficció", "Aventura", "Fantasia"]	{"Centre": 8, "Llevant": 3}
5	{"Títol": "Història de dues ciutats", "Autor": "Charles Dickens", "Any": 1859}	["Ficció", "Història", "Drama"]	{"Llevant": 2}

Cada libro tiene información (título, autor y año de publicación), una lista de temas y datos sobre los ejemplares disponibles en cada una de las tres bibliotecas de la ciudad (Centro, Levante y Poniente).

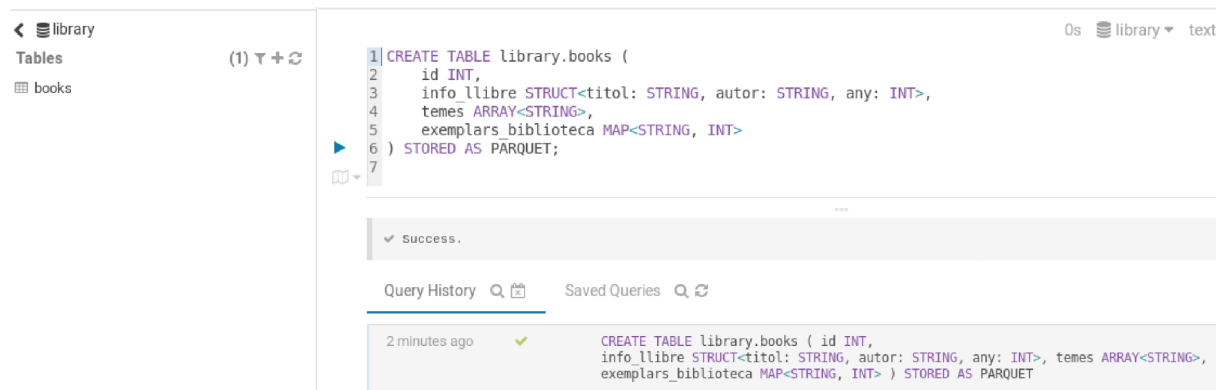
Crea una tabla en el almacén de datos de Hive utilizando los tipos complejos **STRUCT**, **ARRAY** y **MAP**. Carga los datos ejecutando sentencias **INSERT**. Ejecuta las siguientes consultas en Hive (con **HiveQL**) y en Impala (con **SQL**):

1 IMPORTAR DATOS

Se ha creado la base de datos *library* y seguidamente se ha creado la tabla e insertado los datos.:

```
CREATE DATABASE library;
```

```
CREATE TABLE library.books (  
  id INT,  
  info_llibre STRUCT<titol: STRING, autor: STRING, any: INT>,  
  temes ARRAY<STRING>,  
  exemplars_biblioteca MAP<STRING, INT>  
) STORED AS PARQUET;
```



```
INSERT INTO library.books  
  SELECT 1,  
  NAMED_STRUCT('titol', '1984', 'autor', 'George Orwell', 'any',  
1949),  
  ARRAY('Ficció', 'Distopia', 'Societat'),  
  MAP('Centre', 5, 'Llevant', 2);
```

```
INSERT INTO library.books  
  SELECT 2,  
  NAMED_STRUCT('titol', 'Sapiens', 'autor', 'Yuval Noah Harari',  
'any', 2011),  
  ARRAY('Assaig', 'Història', 'Antropologia', 'Societat'),  
  MAP('Llevant', 4, 'Ponent', 3);
```

```
INSERT INTO library.books  
  SELECT 3,  
  NAMED_STRUCT('titol', 'Dune', 'autor', 'Frank Herbert', 'any',  
1965),  
  ARRAY('Ficció', 'Aventura', 'Ciència-ficció'),  
  MAP('Centre', 7, 'Ponent', 2);
```

```
INSERT INTO library.books
  SELECT 4,
  NAMED_STRUCT('titol', 'El Senyor dels anells', 'autor',
'J.R.R. Tolkien', 'any', 1954),
  ARRAY('Ficció', 'Aventura', 'Fantasia'),
  MAP('Centre', 8, 'Llevant', 3);
```

```
INSERT INTO library.books
  SELECT 5,
  NAMED_STRUCT('titol', 'Història de dues ciutats', 'autor',
'Charles Dickens', 'any', 1859),
  ARRAY('Ficció', 'Història', 'Drama'),
  MAP('Llevant', 2);
```

3 minutes ago	✓	INSERT INTO library.books SELECT 5, NAMED_STRUCT('titol', 'Història de dues ciutats', 'autor', 'Charles Dickens', 'any', 1859), ARRAY('Ficció', 'Història', 'Drama'), MAP('Llevant', 2)
9 minutes ago	✓	INSERT INTO library.books SELECT 4, NAMED_STRUCT('titol', 'El Senyor dels anells', 'autor', 'J.R.R. Tolkien', 'any', 1954), ARRAY('Ficció', 'Aventura', 'Fantasia'), MAP('Centre', 8, 'Llevant', 3)
18 minutes ago	✓	INSERT INTO library.books SELECT 3, NAMED_STRUCT('titol', 'Dune', 'autor', 'Frank Herbert', 'any', 1965), ARRAY('Ficció', 'Aventura', 'Ciència-ficció'), MAP('Centre', 7, 'Ponent', 2)
22 minutes ago	✓	INSERT INTO library.books SELECT 2, NAMED_STRUCT('titol', 'Sapiens', 'autor', 'Yuval Noah Harari', 'any', 2011), ARRAY('Assaig', 'Història', 'Antropologia', 'Societat'), MAP('Llevant', 4, 'Ponent', 3)
4 minutes ago	✓	INSERT INTO library.books SELECT 1, NAMED_STRUCT('titol', '1984', 'autor', 'George Orwell', 'any', 1949), ARRAY('Ficció', 'Distòpia', 'Societat'), MAP('Centre', 5, 'Llevant', 2)
29 minutes ago	✓	CREATE TABLE library.books (id INT, info_llibre STRUCT<titol: STRING, autor: STRING, any: INT>, temes ARRAY<STRING>, exemplars_biblioteca MAP<STRING, INT>) STORED AS PARQUET

2 CONSULTAS EN HIVE

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

```
SELECT info_llibre.titol
FROM library.books
WHERE info_llibre.any >= 1900 AND info_llibre.any < 2000;
```

The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo and options to 'Add a name...' and 'Add a description...'. Below this is a text area containing the SQL query: `1 SELECT info_llibre.titol`, `2 FROM library.books`, `3 WHERE info_llibre.any >= 1900 AND info_llibre.any < 2000;`. To the left of the text area is a sidebar with icons for a factory, a book, and a download. Below the text area, there's a navigation bar with 'Query History', 'Saved Queries', and 'Results (3)'. The 'Results (3)' tab is active, showing a table with the column 'titol' and three rows of results: '1 Dune', '2 El Senyor dels anells', and '3 1984'.

	titol
1	Dune
2	El Senyor dels anells
3	1984

2. Recupera el título de los libros del tema “Historia”.

```
SELECT info_llibre.titol
FROM library.books
WHERE array_contains(temes, 'Història');
```

The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo and options to 'Add a name...' and 'Add a description...'. Below this is a text area containing the SQL query: `1 SELECT info_llibre.titol`, `2 FROM library.books`, `3 WHERE array_contains(temes, 'Història');`, `4`. To the left of the text area is a sidebar with icons for a factory, a book, and a download. Below the text area, there's a navigation bar with 'Query History', 'Saved Queries', and 'Results (2)'. The 'Results (2)' tab is active, showing a table with the column 'titol' and two rows of results: '1 Història de dues ciutats' and '2 Sapiens'.

	titol
1	Història de dues ciutats
2	Sapiens

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

```
SELECT SUM(exemplars_biblioteca['Llevant']) AS  
exemplars_llevant  
FROM library.books  
WHERE exemplars_biblioteca['Llevant'] IS NOT NULL;
```

The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo and options to 'Add a name...' and 'Add a description...'. Below this is a text area containing a SQL query. The query is as follows:

```
1  
2 SELECT SUM(exemplars_biblioteca['Llevant']) AS exemplars_llevant  
3 FROM library.books  
4 WHERE exemplars_biblioteca['Llevant'] IS NOT NULL;  
5
```

Below the query editor, there are tabs for 'Query History', 'Saved Queries', and 'Results (1)'. The 'Results (1)' tab is active, showing a table with the column 'exemplars_llevant' and a single row with the value '11'.

exemplars_llevant
11

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

```
SELECT info_llibre.titol  
FROM library.books  
WHERE exemplars_biblioteca['Centre'] IS NOT NULL  
AND exemplars_biblioteca['Centre'] > 0  
AND array_contains(temes, 'Ficció');
```

The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo and options to 'Add a name...' and 'Add a description...'. Below this is a text area containing a SQL query. The query is as follows:

```
1  
2 SELECT info_llibre.titol  
3 FROM library.books  
4 WHERE exemplars_biblioteca['Centre'] IS NOT NULL  
5 AND exemplars_biblioteca['Centre'] > 0  
6 AND array_contains(temes, 'Ficció');  
7
```

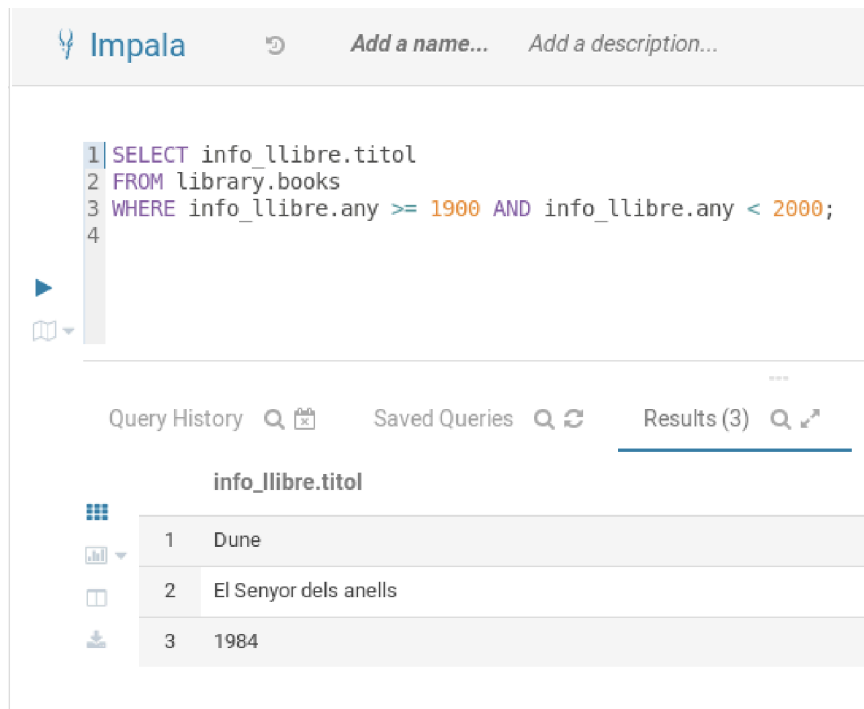
Below the query editor, there are tabs for 'Query History', 'Saved Queries', and 'Results (3)'. The 'Results (3)' tab is active, showing a table with the column 'titol' and three rows of results.

titol
1 Dune
2 El Senyor dels anells
3 1984

3 CONSULTAS EN IMPALA

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

```
SELECT info_llibre.titol
FROM library.books
WHERE info_llibre.any >= 1900 AND info_llibre.any < 2000;
```



The screenshot shows the Impala web interface. At the top, there's a header with the Impala logo, a refresh icon, and links to 'Add a name...' and 'Add a description...'. Below the header is a text area containing a SQL query:

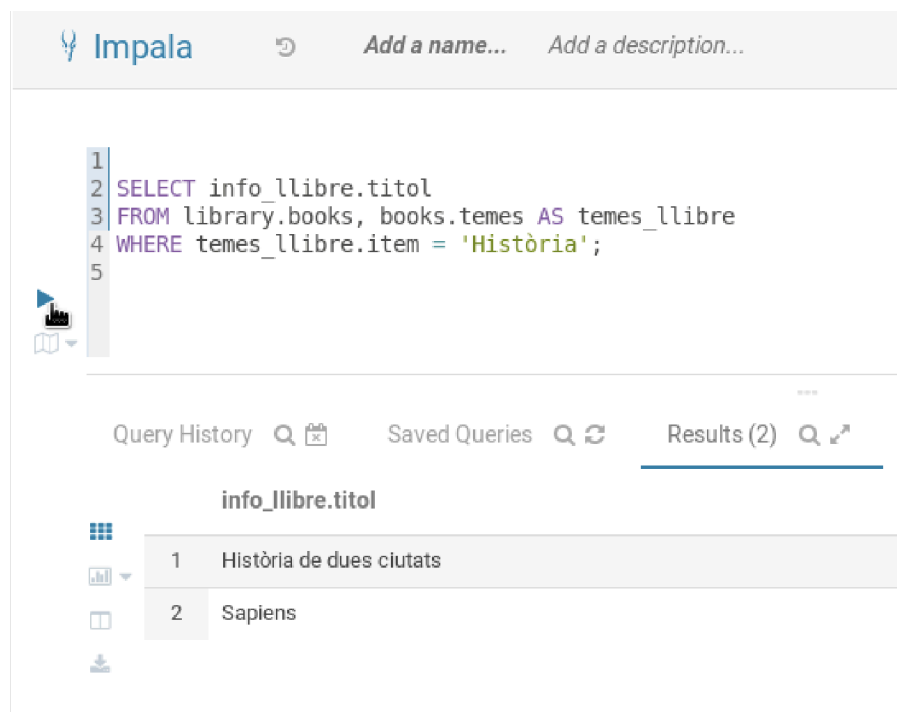
```
1 SELECT info_llibre.titol
2 FROM library.books
3 WHERE info_llibre.any >= 1900 AND info_llibre.any < 2000;
4
```

To the left of the query area is a vertical toolbar with icons for running the query, saving it, and other actions. Below the query area, there's a navigation bar with tabs for 'Query History', 'Saved Queries', and 'Results (3)'. The 'Results (3)' tab is active, showing a table with the following data:

	info_llibre.titol
1	Dune
2	El Senyor dels anells
3	1984

2. Recupera el título de los libros del tema “Historia”.

```
SELECT info_llibre.titol
FROM library.books, books.temes AS temes_llibre
WHERE temes_llibre.item = 'Història';
```



The screenshot shows the Impala web interface. At the top, there's a header with the Impala logo, a refresh icon, and links to 'Add a name...' and 'Add a description...'. Below the header is a text area containing a SQL query:

```
1
2 SELECT info_llibre.titol
3 FROM library.books, books.temes AS temes_llibre
4 WHERE temes_llibre.item = 'Història';
5
```

To the left of the query area is a vertical toolbar with icons for running the query, saving it, and other actions. Below the query area, there's a navigation bar with tabs for 'Query History', 'Saved Queries', and 'Results (2)'. The 'Results (2)' tab is active, showing a table with the following data:

	info_llibre.titol
1	Història de dues ciutats
2	Sapiens

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

```
SELECT SUM(exemplars.value) AS exemplars_llevant
FROM library.books, books.exemplars_biblioteca AS exemplars
WHERE exemplars.key = 'Llevant';
```

The screenshot shows the Impala web interface. At the top, there's a header with the Impala logo and options to 'Add a name...' and 'Add a description...'. Below the header, the SQL query from the previous block is entered into the text area. The query is: `1 SELECT SUM(exemplars.value) AS exemplars_llevant`, `2 FROM library.books, books.exemplars_biblioteca AS exemplars`, `3 WHERE exemplars.key = 'Llevant';`, `4`. Below the query, there's a 'Query History' tab, a 'Saved Queries' tab, and a 'Results (1)' tab. The 'Results (1)' tab is active, showing a table with the title 'exemplars_llevant' and one row with the value 11.

exemplars_llevant	
1	11

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

```
SELECT info_llibre.titol
FROM library.books, books.exemplars_biblioteca AS exemplars,
books.temes AS temes_llibre
WHERE exemplars.key = 'Centre'
AND exemplars.value > 0
AND temes_llibre.item = 'Ficció';
```

The screenshot shows the Impala web interface. At the top, there's a header with the Impala logo and options to 'Add a name...' and 'Add a description...'. Below the header, the SQL query from the previous block is entered into the text area. The query is: `1 SELECT info_llibre.titol`, `2 FROM library.books, books.exemplars_biblioteca AS exemplars, books.temes AS temes_llibre`, `3 WHERE exemplars.key = 'Centre'`, `4 AND exemplars.value > 0`, `5 AND temes_llibre.item = 'Ficció';`, `6`, `7`. Below the query, there's a 'Query History' tab, a 'Saved Queries' tab, and a 'Results (3)' tab. The 'Results (3)' tab is active, showing a table with the title 'info_llibre.titol' and three rows with the values 1984, Dune, and El Senyor dels anells.

info_llibre.titol	
1	1984
2	Dune
3	El Senyor dels anells