



# Proyecto Integrador Base de Datos

★ José Luis Benítez ★

# Importación del Archivo

	4.800 rows						Tx: Auto	DDL	
	WHERE		ORDER BY						

production_companies	:
[{"name": "Metakwon Filmworks", "id": 65815}]	
genres	:
Action Adventure Fantasy Science Fiction	
Adventure Fantasy Action	
homepage	1
<null>	
<null>	
<null>	

status
Released
Released
Released
Released
Released
Released
Released
Released
Released
Released
Released
Released
Released
Released



# Normalización de Columnas con Datos Repetitivos



job

1	24 Frame Playback
2	2D Artist
3	2D Supervisor
4	3D Animator
5	3D Artist
6	3D Coordinator
7	3D Modeller
8	3D Supervisor

status

1	Post Production
2	Released
3	Rumored

department

1	Actors
2	Art
3	Camera
4	Costume & Make-Up
5	Crew
6	Directing
7	Editing
8	Lighting
9	Production
10	Sound
11	Visual Effects
12	Writing

gender

1	0
2	1
3	2



```
9  ----- Normalización de Status -----
10
11  DROP TABLE IF EXISTS status;
12
13  CREATE TABLE status (status VARCHAR(100)) AS SELECT DISTINCT status
14  FROM movie_dataset.movie_dataset;
15
16
17  ALTER TABLE status
18  ADD PRIMARY KEY (status);
19
20  -- Creación de la relación movie-status
21  CREATE TABLE status_movie (id INT NOT NULL,
22  status VARCHAR(100)) AS SELECT id, st.status AS status
23  FROM movie_dataset.movie_dataset movie_status, status st
24  WHERE movie_status.status = st.status;
25
26  ALTER TABLE status_movie
27  ADD FOREIGN KEY (id) REFERENCES movie(id),
28  ADD FOREIGN KEY (status) REFERENCES status(status);
29
```



## Columnas con valores no atómicos

```
341 SELECT id, cast, length(cast) as longitud,  
342       length(REPLACE(cast, ' ', '')) as longitudSinEspacios,  
343       length(cast) - length(REPLACE(cast, ' ', '')) as CantidadEspacios,  
344       length(cast) - length(REPLACE(cast, ' ', '')) + 1 as CantidaDePalabras  
345 FROM cast ;
```

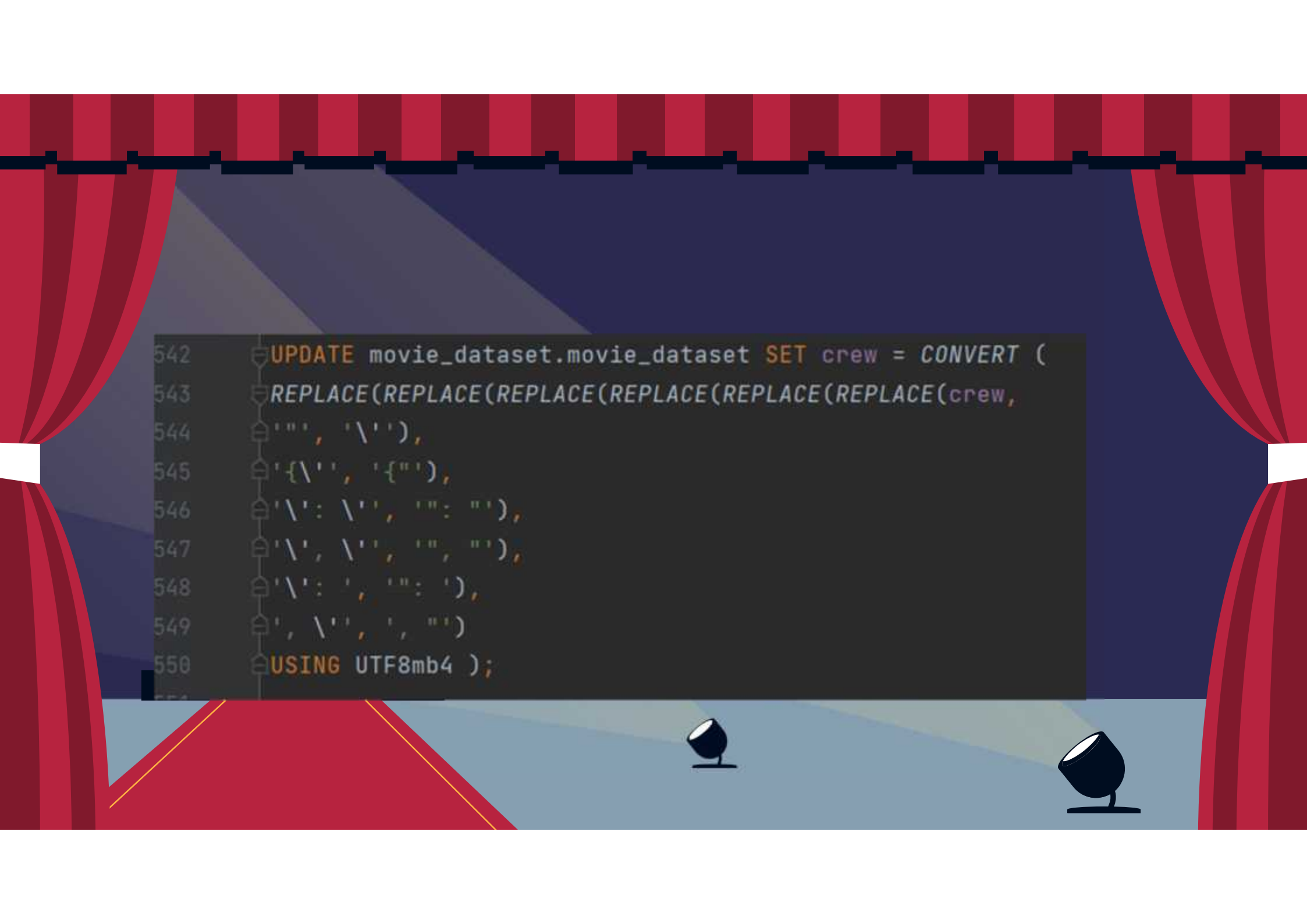
cast

Michael Madsen Iman Crosson Lazar Ristovski

```
66  
67 UPDATE cast  
68 SET cast = REPLACE(cast, 'Anh Le Guy Pearce', 'Anh Le_Guy Pearce');  
69
```

# Columnas Formato Json

```
401 ..... Normalización de spoken_languages .....
402
403 * (('xss_APS_XT', 'en', 'name': 'English'))
404
405 DROP PROCEDURE IF EXISTS spoken_languages;
406 DELIMITER $$
407 CREATE PROCEDURE spoken_languages()
408 BEGIN
409     DECLARE done INT DEFAULT FALSE;
410     DECLARE jsonData JSON;
411     DECLARE jsonId VARCHAR(255);
412     DECLARE jsonLabel VARCHAR(255);
413     DECLARE movieId INT;
414     DECLARE i INT;
415
416     -- Declarar el cursor
417     DECLARE myCursor
418     CURSOR FOR
419     SELECT id, JSON_EXTRACT(CONVERT(spoken_languages USING UTF8MB4), '$[*]')
420     FROM movie_dataset.movie_dataset;
421     -- Declarar el handler para NOT FOUND
422
423     DECLARE CONTINUE HANDLER
424     FOR NOT FOUND SET done = TRUE;
425
426     -- Abrir el cursor
427     OPEN myCursor;
428     cursorLoop: LOOP
429
430     FETCH myCursor INTO movieId, jsonData;
431     SET i = 0;
432     IF done THEN
433     LEAVE cursorLoop;
434     END IF;
435
436     WHILE (JSON_EXTRACT(jsonData, CONCAT('$[', i, ']')) IS NOT NULL) DO
437     SET jsonId = IFNULL(JSON_EXTRACT(jsonData, CONCAT('$[', i, '].id')), '');
438     SET jsonLabel = IFNULL(JSON_EXTRACT(jsonData, CONCAT('$[', i, '].name')), '');
439     SET @sql_text = CONCAT('INSERT INTO spoken_languages VALUES (', movieId, ', ', REPLACE(jsonId, '\\', '\\\\'), ', ', jsonLabel, ');');
440
441     PREPARE stmt FROM @sql_text;
442     EXECUTE stmt;
443     DEALLOCATE PREPARE stmt;
444     SET i = i + 1;
445     END WHILE;
446
447     END LOOP;
448
449     CLOSE myCursor;
450
451 END$$
452 DELIMITER ;
```

The image depicts a stage with red curtains on the left and right sides. A dark blue backdrop covers the center. Two spotlights on the floor cast beams of light onto the stage. Overlaid on the stage is a dark gray rectangular area containing SQL code. The code is color-coded: keywords like UPDATE, SET, CONVERT, REPLACE, and USING are in orange, while identifiers like movie\_dataset and crew are in purple. The code is presented as if it's being typed into a terminal or editor, with line numbers 542 through 550 visible on the left side of the code block.

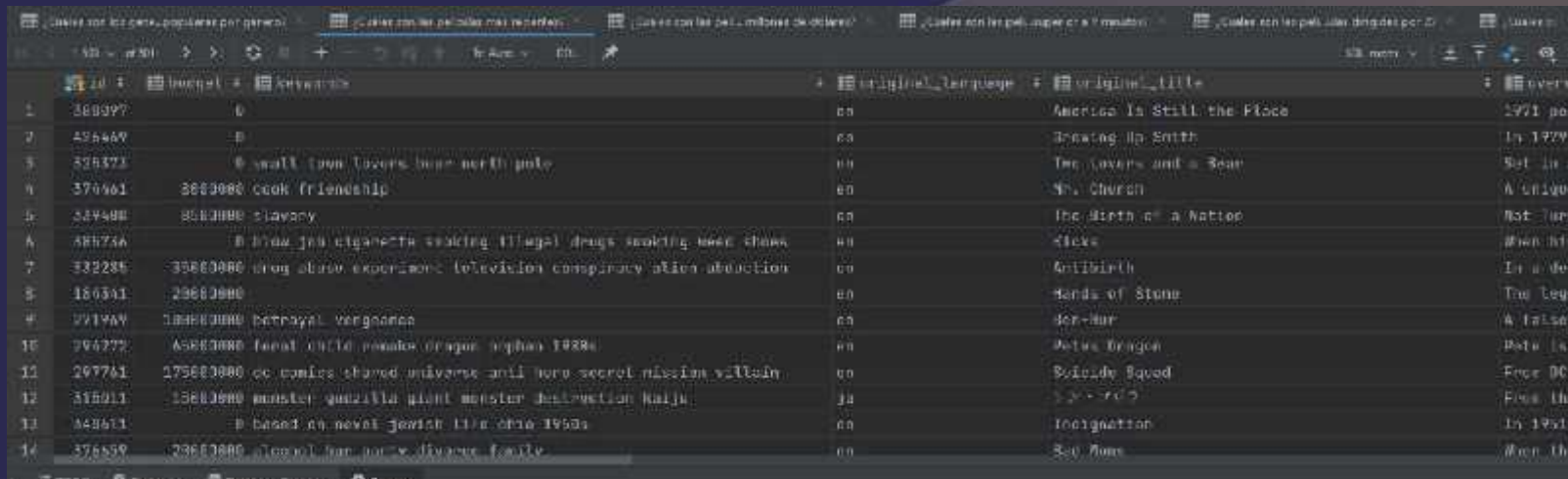
```
542 UPDATE movie_dataset.movie_dataset SET crew = CONVERT (
543 REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(crew,
544 '"', '\'),
545 '{\ ', '{'),
546 '\\': '\', '"': '"'),
547 '\\', '\', '"', '"'),
548 '\\': ', '"': '),
549 ', \', ', '"')
550 USING UTF8mb4 );
```



-- ¿Cuáles son las películas más recientes?

SELECT \* FROM movie ORDER BY release\_date DESC;

1



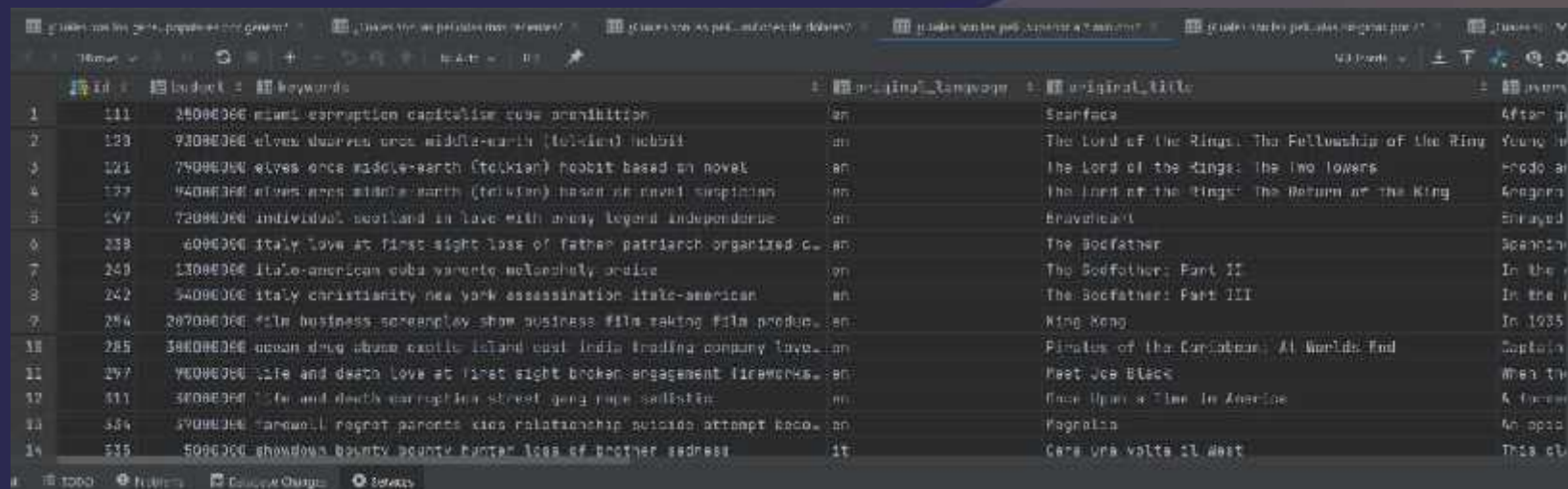
	id	budget	keywords	original_language	original_title	overview
1	369397	0		en	America Is Still the Place	In 1971, peo
2	454449	0		en	Braving Up Smith	In 1979,
3	328373	0	small town towners bear north pole	en	Two Lovers and a Bear	Set in a
4	374461	3663000	cook friendship	en	Mr. Church	A unique
5	349488	8563000	slavery	en	The Birth of a Nation	Not ju
6	486736	0	how you cigarette smoking illegal drugs smoking weed chaw	en	Sticks	When his
7	322285	35663000	drug abuse experiment television conspiracy alien abduction	en	Gridlock	In a des
8	166341	23663000		en	Hands of Stone	The legs
9	271949	18663000	betrayal vengeance	en	Don-Hur	A false
10	294272	45663000	fatal child predator dragon alpha 1888	en	Alpha Dragon	Pete is
11	297761	27663000	cc comics shared universe anti hero secret mission villain	en	Solo: A Star Wars Story	From DC
12	316911	15663000	monster Godzilla giant monster destruction kaiju	ja	ゴジラ	From the
13	443411	0	based on novel jewish life Ohio 1950s	en	Indignation	In 1951,
14	376559	23663000	fatal bar port divorce family	en	See How	When the



-- ¿Cuáles son las películas con una duración superior a Y minutos?

SELECT \* FROM movie WHERE runtime > 160;

2



	id	budget	keywords	original_language	original_title	genre
1	111	24000000	miami corruption capitalism crime prohibition	en	Scarface	Crime
2	123	93000000	elves dwarves orcs middle-earth (tolkien) hobbit	en	The Lord of the Rings: The Fellowship of the Ring	Young
3	121	79000000	elves orcs middle-earth (tolkien) hobbit based on novel	en	The Lord of the Rings: The Two Towers	Adventure
4	127	94000000	elves orcs middle-earth (tolkien) based on novel	en	The Lord of the Rings: The Return of the King	Adventure
5	197	72000000	individual scotland in love with army legend independence	en	Braveheart	War
6	239	60000000	italy love at first sight loss of father patriarch organized crime	en	The Godfather	Drama
7	240	130000000	italo-american cuba varmeto mobbeduly crime	en	The Godfather: Part II	Drama
8	242	54000000	italy christianity new york assassination italo-american	en	The Godfather: Part III	Drama
9	254	307000000	film business screenplay show business film making film production	en	King Kong	Adventure
10	285	380000000	ocean drug abuse exotic island east india trading company love	en	Pirates of the Caribbean: At World's End	Adventure
11	297	90000000	life and death love at first sight broken engagement fireworks	en	Meet Joe Black	Drama
12	311	30000000	life and death corruption street gang rape sadistic	en	Once Upon a Time in America	Drama
13	354	19000000	farewell regret paranoia kids relationship suicide attempt kidnapping	en	Magnolia	Drama
14	335	50000000	showdown bounty bounty hunter loss of brother sadness	it	Cera era volta al West	Western

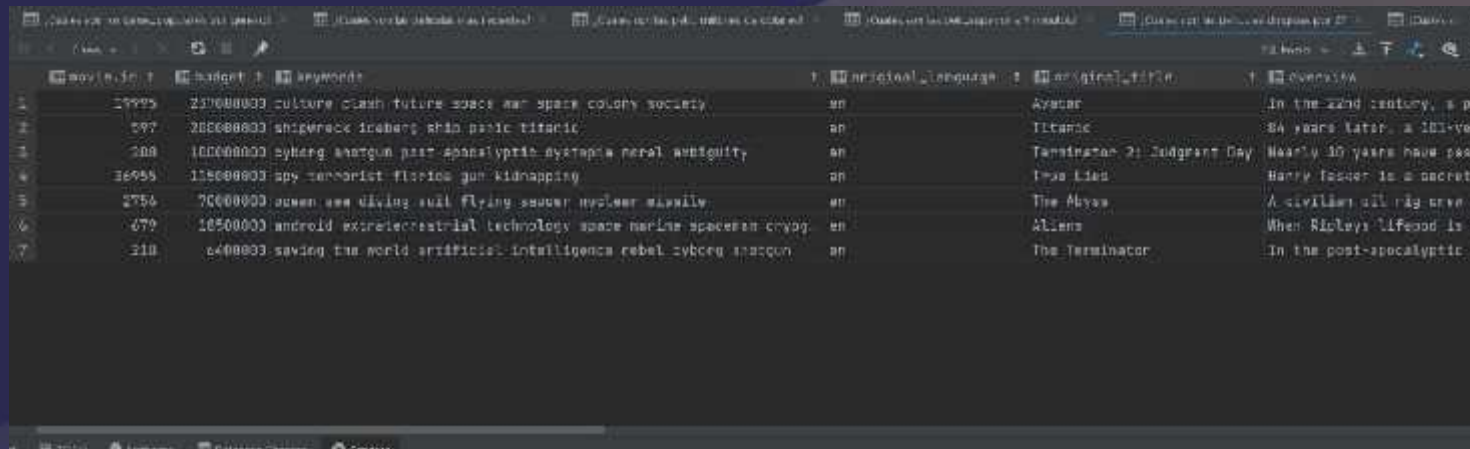
-- ¿Cuáles son las películas dirigidas por Z?

SELECT \* FROM movie, director

WHERE director.id = movie.id

AND director = 'James Cameron';

3



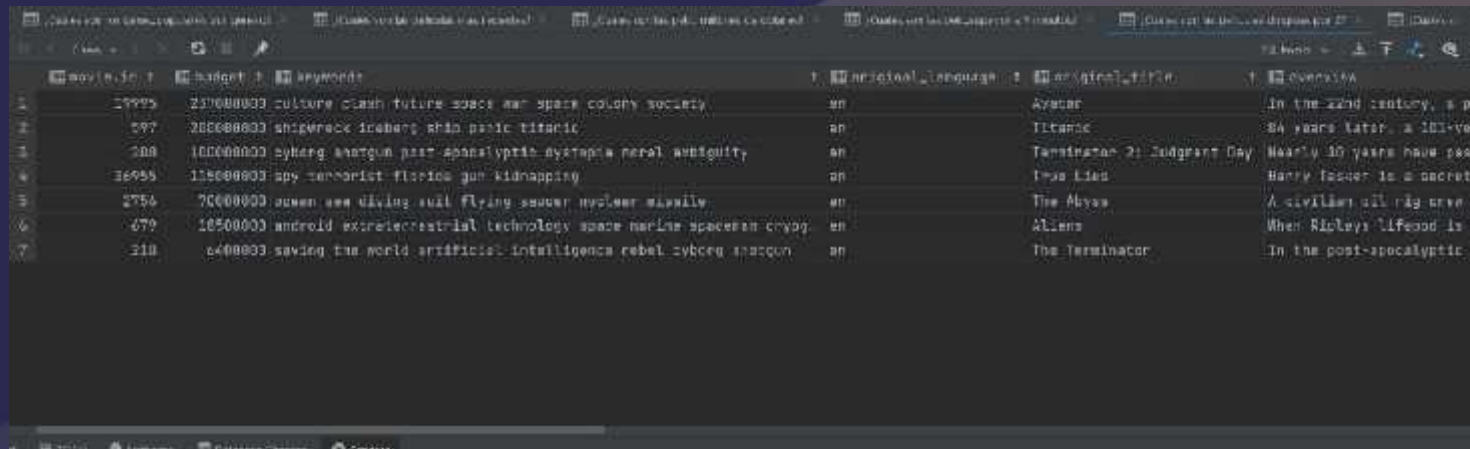
The screenshot shows a database query result with columns: movie.id, budget, keywords, original\_language, original\_title, and overview. The results are filtered for movies directed by James Cameron.

movie.id	budget	keywords	original_language	original_title	overview
19995	217000000	culture clash future space war space colony society	en	Ayestar	In the 22nd century, a pa
597	200000000	shipwreck iceberg ship panic titanic	en	Titanic	84 years later, a 101-yea
100	100000000	cyberg anagun past apocalyptic dystopia moral ambiguity	en	Terminator 2: Judgment Day	Nearly 10 years have pass
16956	115000000	spy terrorist florida gun kidnapping	en	True Lies	Harry Tasker is a secret
2756	70000000	women war diving sub flying saucer nuclear missile	en	The Abyss	A civiliser all rig time in
679	10500000	android extraterrestrial technology space marine spaceship cryog	en	Aliens	When Ripley's lifedood is f
110	6400000	saving the world artificial intelligence rebel cyborg spacegun	en	The Terminator	In the post-apocalyptic f

-- ¿Cuáles son las películas con un actor/actriz específica en el reparto?

SELECT \* FROM movie WHERE cast LIKE '%Roberth Downey Jr%';

4



The screenshot shows a database query result with columns: movie\_id, budget, keywords, original\_language, original\_title, and overview. The results are as follows:

movie_id	budget	keywords	original_language	original_title	overview
19995	217000000	culture clash future space war space colony society	en	Ayestar	In the 22nd century, a pa
597	200000000	shipwreck iceberg ship panic titanic	en	Titanic	84 years later, a 101-year
100	100000000	cyberg anagun past apocalyptic dystopia moral ambiguity	en	Terminator 2: Judgment Day	Nearly 10 years have pass
16956	115000000	spy terrorist florida gun kidnapping	en	True Lies	Harry Tasker is a secret
2756	70000000	women war diving sub flying saucer nuclear missile	en	The Abyss	A civiliser all rig time in
679	10500000	android extraterrestrial technology space marine spaceship cryog	en	Aliens	When Ripley's lifecood is f
110	6400000	saving the world artificial intelligence rebel cyborg apocyon	en	The Terminator	In the post-apocalyptic f

-- ¿Cuáles son los géneros más populares por género?

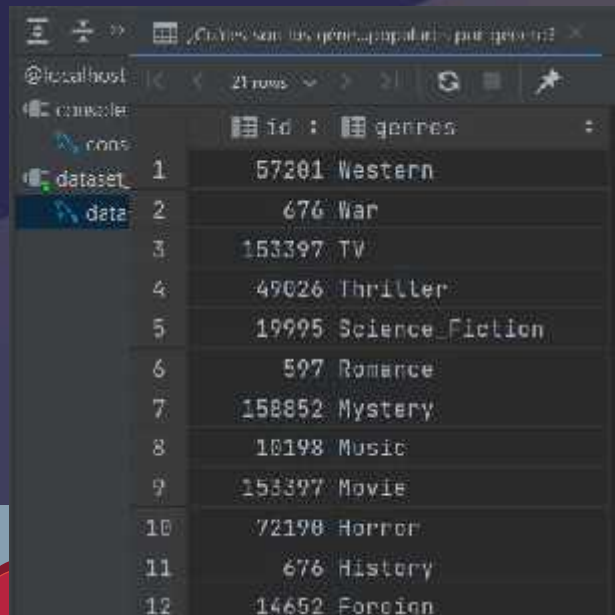
```
SELECT id, genres
```

```
FROM movie, genres
```

```
WHERE movie.id = genres.id_movie
```

```
GROUP BY genres ORDER BY genres DESC;
```

5



A screenshot of a database query result window. The window title is "¿Cuáles son los géneros más populares por género?". The left sidebar shows a tree view with "console" selected. The main area displays a table with 21 rows. The table has two columns: "id" and "genres". The rows are ordered by the number of movies in each genre, descending. The first five rows are: 1. 57201 Western, 2. 676 War, 3. 153397 TV, 4. 49026 Thriller, 5. 19995 Science\_Fiction. The remaining rows are: 6. 597 Romance, 7. 158852 Mystery, 8. 10198 Music, 9. 153397 Movie, 10. 72198 Horror, 11. 676 History, 12. 14652 Foreign.

	id	genres
1	57201	Western
2	676	War
3	153397	TV
4	49026	Thriller
5	19995	Science_Fiction
6	597	Romance
7	158852	Mystery
8	10198	Music
9	153397	Movie
10	72198	Horror
11	676	History
12	14652	Foreign



Gracias