

## Reto 2.3: Consultas con subconsultas II

### [Bases de Datos] Unidad Didáctica 2: DML - Consultas Avanzadas

Para este reto, volveremos a usar la base de datos Chinook (más información en el Reto 2.1).

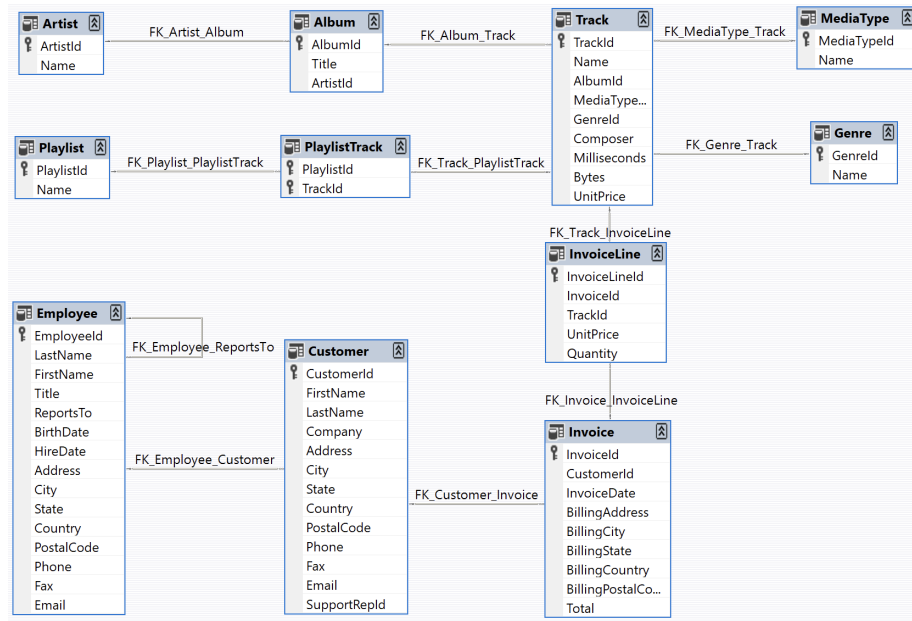


Figure 1: Diagrama relacional de Chinook (fuente: [github.com/lerocha/chinook-database](https://github.com/lerocha/chinook-database)).

Tras cargar esta base de datos en tu SGBD, realiza las siguientes consultas:

### Subconsultas Escalares (Scalar Subqueries)

Estas subconsultas devuelven un solo valor, por lo general, se utilizan en contextos donde se espera un solo valor, como parte de una condición **WHERE**, **SELECT**, **HAVING**, etc. Ejemplo:

*Obtener una lista de empleados que ganan más que el salario medio de la empresa.*  
[1]

```
SELECT
    lastname,
    firstname,
    salary
FROM employee
```

```
WHERE salary > (SELECT avg(salary)
                FROM employee)
```

### Consulta 1

Obtener las canciones con una duración superior a la media.

### Consulta 2

Listar las 5 últimas facturas del cliente cuyo email es “emma\_jones@hotmail.com”.

## Subconsultas de varias filas

Diferenciamos dos tipos:

1. Subconsultas que devuelven una columna con múltiples filas (es decir, una lista de valores). Suelen incluirse en la cláusula **WHERE** para filtrar los resultados de la consulta principal. En este caso, suelen utilizarse con operadores como **IN**, **NOT IN**, **ANY**, **ALL**, **EXISTS** o **NOT EXISTS**.
2. Subconsultas que devuelven múltiples columnas con múltiples filas (es decir, tablas). Se comportan como una tabla temporal y se utilizan en lugares donde se espera una tabla, como en una cláusula **FROM**. [2]

### Consulta 3

Mostrar las listas de reproducción en las que hay canciones de reggae.

### Consulta 4

Obtener la información de los clientes que han realizado compras superiores a 20€.

### Consulta 5

Álbumes que tienen más de 15 canciones, junto a su artista.

### Consulta 6

Obtener los álbumes con un número de canciones superiores a la media.

### Consulta 7

Obtener los álbumes con una duración total superior a la media.

### Consulta 8

Canciones del género con más canciones.

### Consulta 9

Canciones de la *playlist* con más canciones.

### Subconsultas Correlacionadas (Correlated Subqueries):

Son subconsultas en las que la subconsulta interna depende de la consulta externa. Esto significa que la subconsulta se ejecuta una vez por cada fila devuelta por la consulta externa, suponiendo una gran carga computacional. Ejemplo:

*Supongamos que queremos encontrar a todos los empleados con un salario superior al promedio de su departamento. [1]*

```
SELECT
    lastname,
    firstname,
    salary
FROM employee e1
WHERE e1.salary > (SELECT avg(salary)
                   FROM employee e2
                   WHERE e2.dept_id = e1.dept_id)
```

La principal diferencia entre una subconsulta correlacionada en SQL y una subconsulta simple es que las subconsultas correlacionadas hacen referencia a columnas de la tabla externa. En el ejemplo anterior, `e1.dept_id` es una referencia a la tabla de la subconsulta externa. [1]

### Consulta 10

Muestra los clientes junto con la cantidad total de dinero gastado por cada uno en compras.

### Consulta 11

Obtener empleados y el número de clientes a los que sirve cada uno de ellos.

### Consulta 12

Ventas totales de cada empleado.

### Consulta 13

Álbumes junto al número de canciones en cada uno.

### Consulta 14

Obtener el nombre del álbum más reciente de cada artista. Pista: los álbumes más antiguos tienen un ID más bajo.

## Referencias

- [1] <https://learnsql.es/blog/subconsulta-correlacionada-en-sql-una-guia-para-principiantes/>
- [2] <https://learnsql.es/blog/cuales-son-los-diferentes-tipos-de-subconsultas-sql/>

## Query 1

Obtener las canciones con una duración superior a la media.

### Resolución

Obtengo la duración media de las canciones

AVG(Milliseconds)
393599

*Número de filas en la tabla:* 1

A continuación, selecciono todas las canciones con una duración superior a esta media (obtengo la media con una subconsulta)

Ordeno por duración para verificar que la duración de la más corta es superior a la media obtenida anteriormente

TrackId	Name	AlbumId	MediaTypeId	5 columnas más
806	Wicked Ways	64	1	...
3413	Concerto for Clarine...	282	2	...
1897	The Shortest Straw	156	1	...
1864	The Unforgiven II	153	1	...
1387	22 Acacia Avenue	112	1	...
959	King For A Day	76	1	...
1855	The Thing That Shoul...	152	1	...
2711	Save Me (Remix)	219	1	...
1191	Deep Waters	93	1	...
1880	Creeping Death	154	1	...
...	...	...	...	...

*Número de filas en la tabla:* 494

## Query 2

Listar las facturas del cliente cuyo email es “emma\_jones@hotmail.com”.

### Resolución

Obtengo el ID del cliente

CustomerId
52

*Número de filas en la tabla: 1*

Obtengo las facturas asociadas a este ID

InvoiceId	CustomerId	InvoiceDate	BillingAddress	5 columnas más
11	52	2021-02-06 00:00:00	202 Hoxton Street	...
140	52	2022-09-13 00:00:00	202 Hoxton Street	...
163	52	2022-12-16 00:00:00	202 Hoxton Street	...
185	52	2023-03-20 00:00:00	202 Hoxton Street	...
237	52	2023-11-08 00:00:00	202 Hoxton Street	...
358	52	2025-05-01 00:00:00	202 Hoxton Street	...
369	52	2025-06-11 00:00:00	202 Hoxton Street	...

*Número de filas en la tabla: 7*

Ordeno por fecha y me quedo con las 5 últimas

InvoiceId	CustomerId	InvoiceDate	BillingAddress	5 columnas más
369	52	2025-06-11 00:00:00	202 Hoxton Street	...
358	52	2025-05-01 00:00:00	202 Hoxton Street	...
237	52	2023-11-08 00:00:00	202 Hoxton Street	...
185	52	2023-03-20 00:00:00	202 Hoxton Street	...
163	52	2022-12-16 00:00:00	202 Hoxton Street	...

*Número de filas en la tabla: 5*

### Query 3

Mostrar las listas de reproducción en las que hay canciones de reggae.

### Resolución

Obtengo el ID del género reggae

GenreId
8

*Número de filas en la tabla:* 1

Obtengo los IDs de las canciones de este género

TrackId
282
283
284
285
286
287
288
289
290
291
...

*Número de filas en la tabla:* 58

Obtengo las listas de reproducción en las que aparecen estas canciones

PlaylistId	Name
1	Music
8	Music
5	90's Music

*Número de filas en la tabla:* 3

## Query 4

Obtener la información de los clientes que han realizado compras superiores a 20€.

## Resolución

Podemos sacar los ID de los clientes que realizaron compras por importe superior a 20€ en la tabla 'Invoice'.

CustomerId
6
26
45
46

*Número de filas en la tabla: 4*

Ahora podemos usar la tabla ‘Customer’ para mostrar la información de los clientes de esta lista.

CustomerId	FirstName	LastName
6	Helena	Holý
26	Richard	Cunningham
45	Ladislav	Kovács
46	Hugh	O’Reilly

*Número de filas en la tabla: 4*

## Query 5

Álbumes que tienen más de 15 canciones, junto a su artista.

### Resolución

En primer lugar, podemos contar el número de canciones de cada álbum

**Error al ejecutar la consulta:** (1054, “Unknown column ‘Title’ in ‘field list’ ”)

Ahora es fácil filtrar aquellos que tienen más de 15 canciones

**Error al ejecutar la consulta:** (1054, “Unknown column ‘Title’ in ‘field list’ ”)

Si nos quedamos únicamente con la columna ‘AlbumId’, podemos realizar fácilmente una subconsulta y mostrar la información de la tabla ‘Album’

AlbumId	Title	ArtistId
18	Body Count	13
21	Prenda Minha	16
23	Minha Historia	17
24	Afrociberdelia	18
26	Acústico MTV [Live]	19
33	Chill: Brazil (Disc ...	24
34	Chill: Brazil (Disc ...	6
36	Greatest Hits II	51

AlbumId	Title	ArtistId
37	Greatest Kiss	52
39	International Superh...	54
...	...	...

*Número de filas en la tabla:* 61

Mostraremos la información del artista con un JOIN

AlbumId	Title	ArtistId	Name
18	Body Count	13	Body Count
21	Prenda Minha	16	Caetano Veloso
23	Minha Historia	17	Chico Buarque
24	Afrociberdelia	18	Chico Science & Naçã...
26	Acústico MTV [Live]	19	Cidade Negra
33	Chill: Brazil (Disc ...	24	Marcos Valle
34	Chill: Brazil (Disc ...	6	Antônio Carlos Jobim
36	Greatest Hits II	51	Queen
37	Greatest Kiss	52	Kiss
39	International Superh...	54	Green Day
...	...	...	...

*Número de filas en la tabla:* 61

También podemos obtener el mismo resultado con una subconsulta escalar correlacionada en el SELECT

AlbumId	Title	ArtistId	(SELECT Name FROM Artist WHERE ArtistId = Album.ArtistId)
18	Body Count	13	Body Count
21	Prenda Minha	16	Caetano Veloso
23	Minha Historia	17	Chico Buarque
24	Afrociberdelia	18	Chico Science & Naçã...
26	Acústico MTV [Live]	19	Cidade Negra
33	Chill: Brazil (Disc ...	24	Marcos Valle
34	Chill: Brazil (Disc ...	6	Antônio Carlos Jobim
36	Greatest Hits II	51	Queen
37	Greatest Kiss	52	Kiss
39	International Superh...	54	Green Day



AlbumId	Title	ArtistId	(SELECT Name FROM Artist WHERE ArtistId = Album.ArtistId)
...	...	...	...

*Número de filas en la tabla:* 61

## Query 6

Obtener los álbumes con un número de canciones superiores a la media.

### Resolución

Obtengo el número de canciones de cada álbum

AlbumId	COUNT(*)
1	10
2	1
3	3
4	8
5	15
6	13
7	12
8	14
9	8
10	14
...	...

*Número de filas en la tabla:* 347

A partir de aquí, obtengo el número medio de canciones de cada álbum

AVG(n_tracks)
10.0951

*Número de filas en la tabla:* 1

Ya estoy en disposición de obtener aquellos álbumes con más pistas que la media

AlbumId	n_tracks
19	11
20	11
35	11

AlbumId	n_tracks
61	11
77	11
81	11
94	11
96	11
98	11
113	11
...	...

*Número de filas en la tabla:* 183

## Query 7

Obtener los álbumes con una duración total superior a la media.

## Resolución

Para obtener la duración de un álbum, tendremos que sumar la duración de sus canciones

AlbumId	Duracion
1	2400415
2	342562
3	858088
4	2453259
5	4411709
6	3450925
7	3249365
8	2906926
9	2671407
10	3927713
...	...

*Número de filas en la tabla:* 347

Ahora podemos sacar la media de la duración de los álbumes

AVG(Duracion)
3.97342e+06

*Número de filas en la tabla:* 1

Saco todos los álbumes con una duración superior a esta media

AlbumId	Title	Duracion
5	Big Ones	4411709
14	Alcohol Fueled Brewt...	4059919
23	Minha Historia	7875643
24	Afrociberdelia	4238776
30	BBC Sessions [Disc 1...	4489920
34	Chill: Brazil (Disc ...	4221459
36	Greatest Hits II	4558541
37	Greatest Kiss	4575025
45	Sambas De Enredo 200...	4388174
46	Supernatural	4500551
...	...	...

*Número de filas en la tabla:* 72

## Query 8

Canciones del género con más canciones

### Resolución

Canciones de cada género

GenreId	COUNT(*)
1	1297
2	130
3	374
4	332
5	12
6	81
7	579
8	58
9	48
10	43
...	...

*Número de filas en la tabla:* 25

Género con más canciones

GenreId	n_canciones
1	1297

*Número de filas en la tabla:* 1

Canciones del género con más canciones (como si fuera una subconsulta escalar)

TrackId	Name	AlbumId	MediaTypeId	5 columnas más
1	For Those About To R...	1	1	...
2	Balls to the Wall	2	2	...
3	Fast As a Shark	3	2	...
4	Restless and Wild	3	2	...
5	Princess of the Dawn	3	2	...
6	Put The Finger On Yo...	1	1	...
7	Let's Get It Up	1	1	...
8	Inject The Venom	1	1	...
9	Snowballed	1	1	...
10	Evil Walks	1	1	...
...	...	...	...	...

*Número de filas en la tabla:* 1297

Puesto que MySQL en ocasiones no permite usar LIMIT en subconsultas situadas en el WHERE, haremos esta triquiñuela usando JOIN (workaround)

GenreId	TrackId	Name	AlbumId	6 columnas más
1	1	For Those About To R...	1	...
1	2	Balls to the Wall	2	...
1	3	Fast As a Shark	3	...
1	4	Restless and Wild	3	...
1	5	Princess of the Dawn	3	...
1	6	Put The Finger On Yo...	1	...
1	7	Let's Get It Up	1	...
1	8	Inject The Venom	1	...
1	9	Snowballed	1	...
1	10	Evil Walks	1	...
...	...	...	...	...

*Número de filas en la tabla:* 1297

Vemos que el número de filas devuelto es idéntico al número de canciones en el género con más canciones, por lo que estamos en lo correcto

## Query 9

Canciones de la playlist con más canciones

### Resolución

Canciones de cada playlist

PlaylistId	COUNT(*)
1	3290
3	213
5	1477
8	3290
9	1
10	213
11	39
12	75
13	25
14	25
...	...

*Número de filas en la tabla:* 14

Playlist con más canciones

PlaylistId	n_canciones
1	3290

*Número de filas en la tabla:* 1

Canciones de la playlist con más canciones

PlaylistId	TrackId	Name	AlbumId	7 columnas más
1	1	For Those About To R...	1	...
1	2	Balls to the Wall	2	...
1	3	Fast As a Shark	3	...
1	4	Restless and Wild	3	...
1	5	Princess of the Dawn	3	...
1	6	Put The Finger On Yo...	1	...
1	7	Let's Get It Up	1	...
1	8	Inject The Venom	1	...
1	9	Snowballed	1	...
1	10	Evil Walks	1	...

PlaylistId	TrackId	Name	AlbumId	7 columnas más
...	...	...	...	...

*Número de filas en la tabla: 3290*

## Query 10

Muestra los clientes junto con la cantidad total de dinero gastado por cada uno en compras.

### Resolución

De la tabla ‘Invoice’, podemos obtener las compras de un cliente concreto (por ejemplo, el 1)

InvoiceId	CustomerId	Total
98	1	3.98
121	1	3.96
143	1	5.94
195	1	0.99
316	1	1.98
327	1	13.86
382	1	8.91

*Número de filas en la tabla: 7*

Del mismo modo, es fácil obtener la suma total gastada por este cliente

SUM(Total)
39.62

*Número de filas en la tabla: 1*

Si iteramos sobre la tabla ‘Customer’, podemos realizar una subconsulta correlacionada y obtener cada uno de estos totales. Para ello, será necesario igualar el ‘CustomerId’ fuera y dentro de la subconsulta

CustomerId	FirstName	
1	0	39.62
2	0	37.62
3	0	39.62
4	0	39.62

CustomerId	FirstName	
5	0	40.62
6	0	49.62
7	0	42.62
8	0	37.62
9	0	37.62
10	0	37.62
...	...	...

*Número de filas en la tabla: 59*

## Query 11

Obtener empleados y el número de clientes a los que sirve cada uno de ellos.

### Resolución

Usando la tabla ‘Customer’, es fácil contar cuántos clientes tiene el empleado 3 (por poner un ejemplo)

NumberOfCustomers (SupportRepId = 3)
21

*Número de filas en la tabla: 1*

La información de los empleados está en la tabla ‘Employee’. Podemos sacar la información de esta tabla poniendo como campo el conteo que acabamos de sacar. En lugar de usar un empleado concreto, usamos el ‘EmployeeId’, en una subconsulta correlacionada como la anterior.

EmployeeId	FirstName	LastName	NumberOfCustomers
1	Andrew	Adams	0
2	Nancy	Edwards	0
3	Jane	Peacock	21
4	Margaret	Park	20
5	Steve	Johnson	18
6	Michael	Mitchell	0
7	Robert	King	0
8	Laura	Callahan	0

*Número de filas en la tabla: 8*

## Query 12

Ventas totales de cada empleado

### Resolución

Obtenemos los clientes cuyo representante de ventas es el empleado 3 (por ejemplo)

SupportRepId	CustomerId
3	1
3	3
3	12
3	15
3	18
3	19
3	24
3	29
3	30
3	33
...	...

*Número de filas en la tabla:* 21

Obtenemos las facturas para estos clientes, que corresponderán con las ventas del empleado

InvoiceId	CustomerId	Total
98	1	3.98
121	1	3.96
143	1	5.94
195	1	0.99
316	1	1.98
327	1	13.86
382	1	8.91
99	3	3.98
110	3	13.86
165	3	8.91
...	...	...

*Número de filas en la tabla:* 146

Sumando los totales, podemos sacar las ventas de este empleado



SUM(Total)
833.04

*Número de filas en la tabla: 1*

Ahora, podemos iterar sobre los empleados con una subconsulta correlacionada. Ya que estamos, ordenamos por totales de ventas.

EmployeeId	FirstName	LastName	TotalSales
3	Jane	Peacock	833.04
4	Margaret	Park	775.4
5	Steve	Johnson	720.16
1	Andrew	Adams	
2	Nancy	Edwards	
6	Michael	Mitchell	
7	Robert	King	
8	Laura	Callahan	

*Número de filas en la tabla: 8*

## Query 13

Obtener empleados y el número de clientes a los que sirve cada uno de ellos.

## Resolución

El número de canciones de cada álbum sale de la tabla ‘Track’, mientras que la información de cada álbum está en ‘Album’

AlbumId	Title	NumberOfSongs
1	For Those About To R...	10
2	Balls to the Wall	1
3	Restless and Wild	3
4	Let There Be Rock	8
5	Big Ones	15
6	Jagged Little Pill	13
7	Facelift	12
8	Warner 25 Anos	14
9	Plays Metallica By F...	8
10	Audioslave	14
...	...	...

*Número de filas en la tabla: 347*

## Query 14

Obtener el nombre del álbum más reciente de cada artista. Pista: los álbumes más antiguos tienen un ID más bajo.

### Resolución

ArtistName	LatestAlbum
AC/DC	Let There Be Rock
Accept	Restless and Wild
Aerosmith	Big Ones
Alanis Morissette	Jagged Little Pill
Alice In Chains	Facelift
Antônio Carlos Jobim	Chill: Brazil (Disc ...
Apocalyptica	Plays Metallica By F...
Audioslave	Revelations
BackBeat	BackBeat Soundtrack
Billy Cobham	The Best Of Billy Co...
...	...

*Número de filas en la tabla:* 204