

Tarea Chica 2

Bases de datos

Pregunta 1:

Para esta pregunta cruzamos “Artistas” con “Canciones” bajo la primary key “aid”. A esto le agregamos la condición de que el nombre del artista sea ‘J Balvin’.

Finalmente nos interesa simplemente el nombre del artista y el nombre de sus canciones por lo que elegimos que se muestre solo el “anombre” y “cnombre” obteniendo la siguiente respuesta:

```
1 SELECT anombre, cnombre
2 FROM Artistas, Canciones
3 WHERE Artistas.aid=Canciones.aid AND Artistas.anombre=="J Balvin"
4
5
6
```

| ! anombre | cnombre |
|-----------|---------|
| J Balvin | Safari |
| J Balvin | Bobo |
| J Balvin | Ginza |

Pregunta 2:

Para responder esta pregunta nuevamente necesitamos solo conocer el nombre y artista de las canciones, por lo que cruzamos “Artistas” con “Canciones” bajo la primary key “aid”. Ahora a esto le agregamos la condición de que el nombre de la canción incluya la palabra “lover”. Esto lo hacemos mediante el comando *LIKE* (usamos los % para que antes y después de la palabra pueda haber cualquier cosa). Finalmente queremos saber solo el nombre de la canción y sus artistas, obteniendo:

```
1 SELECT anombre, cnombre
2 FROM Artistas, Canciones
3 WHERE Artistas.aid=Canciones.aid AND cnombre LIKE "%lover%"
4
5
6
```

| ! anombre | cnombre |
|---------------|----------------------------------|
| Petit Biscuit | Sunset Lover |
| Adele | Send My Love (To Your New Lover) |

Pregunta 3:

Para encontrar los distintos países en que se escucha 'Daddy Yankee' es necesario cruzar todas las tablas (Artistas, Canciones, SeEscucha, Países). Cruzamos "Artistas" con "Canciones" bajo la primary key "aid", luego cruzamos "Canciones" con "SeEscucha" bajo la primary key "cid" y por último cruzamos la tabla "SeEscucha" con "Países" con la primary key "pid". Así obtenemos todo lo que hay en la base de datos. Estos los agrupamos según el nombre del artista (anombre) y el país (pnombre) usando *GROUP BY*, bajo la condición de que el nombre del artista sea 'Daddy Yankee' usando *HAVING*, finalmente buscamos solo el nombre y el país obteniendo lo siguiente:

```
1 SELECT anombre, pnombre
2 FROM Artistas, Canciones, SeEscucha, Países
3 WHERE Canciones.cid=SeEscucha.cid AND Artistas.aid=Canciones.aid AND Países.pid=SeEscucha.pid
4 GROUP BY anombre, pnombre
5 HAVING anombre="Daddy Yankee"
```

| ! anombre | pnombre |
|--------------|----------|
| Daddy Yankee | Bolivia |
| Daddy Yankee | Honduras |

Pregunta 4

En esta pregunta debemos encontrar el identificador de la canción "cid" con su promedio de streams. Para esto cruzamos simplemente la tabla "Canciones" y "SeEscucha" bajo la primary key "cid". Lo obtenido lo agrupamos con el comando *GROUP BY* por nombre de la canción, a esto le agregamos la función *AVG(n_streams)* para obtener el promedio de número de streams (nombramos la columna con el comando *AS*).

Cabe destacar que no fue necesario utilizar el comando *ORDER BY* ya que al hacer el *GROUP BY* predeterminadamente se ordenaron por orden alfabético. Además, debido a cómo agrupa el sistema, los números van antes que las letras, explicando el resultado obtenido. Finalmente, se buscan solo las columnas cid, cnombre y el promedio de steams:

```
1 SELECT Canciones.cid, cnombre, AVG(n_streams) AS "Promedio_Streams"
2 FROM Canciones, SeEscucha
3 WHERE Canciones.cid=SeEscucha.cid
4 GROUP BY cnombre
5 LIMIT 5
```

| ! cid | cnombre | Promedio_Streams |
|-------|-----------------------|------------------|
| 165 | 2 Heads | 22141 |
| 23 | 24K Magic | 16343.2 |
| 169 | 679 (feat. Remy Boyz) | 21142 |
| 181 | 7 Years | 10755.5 |
| 249 | Ain't My Fault | 1213 |

Pregunta 5:

Aquí al pedirnos el promedio de número de streams de la artista "Shakira" es necesario cruzar todas las tablas (Artistas, Canciones, SeEscucha, Países). Cruzamos "Artistas" con "Canciones" bajo la primary key "aid", luego cruzamos "Canciones" con "SeEscucha" bajo la primary key "cid" y por último cruzamos la tabla "SeEscucha" con "Países" con la primary key "pid".

Tras esto, agrupamos mediante el comando *GROUP BY* por nombre de artista y el nombre del país (anombre y pnombre). Además agregamos función *AVG(n_streams)* para obtener el promedio de número de streams (nombramos la columna con el comando *AS*). Luego, usamos con el comando *HAVING* la condición de que el nombre del artista sea "Shakira" y que el promedio de número de streams en cada país sea mayor a 5.000.

Finalmente, ordenamos de mayor a menor los resultados con el comando *ORDER BY _DESC* y mostramos las columnas del identificador del país (pid), el nombre del país (pnombre) y el promedio calculado con el nombre "Promedio_Shakira" obteniendo lo siguiente:

```
1 SELECT Países.pid, pnombre, AVG(n_streams) AS "Promedio_Shakira"
2 FROM Artistas, Canciones, SeEscucha, Países
3 WHERE Canciones.cid=SeEscucha.cid AND Artistas.aid=Canciones.aid AND Países.pid=SeEscucha.pid
4 GROUP BY anombre, pnombre
5 HAVING anombre="Shakira" AND AVG(n_streams)>5000
6 ORDER BY "Promedio_Shakira" DESC
```

| pid | pnombre | Promedio_Shakira |
|-----|----------|------------------|
| ca | Canada | 27607 |
| hn | Honduras | 6467 |
| bo | Bolivia | 5506 |

Bonus:

En internet encontré que el join era este símbolo Θ (solo colocho el símbolo ya que es un *natural Join*).

Quedando así:

$$\pi_{anombre, cnombre}(\sigma_{anombre="J Balvin"}(Artistas \Theta Canciones))$$