

DSGA1004 - BIG DATA

Lab 1: SQLite

Question 1

Query:

```
sqlite> SELECT track_id, tracks.artist_id, artists.artist_name,  
artist_term.term FROM tracks  
...> INNER JOIN artists ON artists.artist_id = tracks.artist_id  
...> INNER JOIN artist_term ON artist_term.artist_id =  
artists.artist_id  
...> WHERE track_id='TRMMWLD128F9301BF2';
```

Output:

| | | | | |
|--------------------|--------------------|-------|---|------------------------|
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | salsa |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | religious music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | smooth jazz |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | hip hop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | funk |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | alternative rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | r&b |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | techno |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | united states |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | dance |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | jazz |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | urban |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | electronic |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | latin |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | pop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | house |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | american |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | male vocalist |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | contemporary christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | black gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | black music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | funk soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | contemporary gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | san diego |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | gospel jams |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | gospel soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tone | X | syntax records |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | | salsa |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | | religious music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | | smooth jazz |

| | | | |
|--------------------|--------------------|-------|------------------------|
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | hip hop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | funk |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | alternative rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | r&b |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | techno |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | united states |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | dance |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | jazz |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | urban |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | electronic |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | latin |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | pop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | house |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | american |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | male vocalist |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | contemporary christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | black gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | black music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | funk soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | contemporary gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | san diego |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | gospel jams |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | gospel soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonex | syntax records |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | salsa |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | religious music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | smooth jazz |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | hip hop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | funk |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | alternative rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | r&b |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | techno |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | united states |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | dance |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | rock |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | jazz |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | urban |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | electronic |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | latin |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | pop |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | house |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | american |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | male vocalist |

| | | | |
|--------------------|--------------------|-------|------------------------|
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | contemporary christian |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | black gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | black music |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | funk soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | contemporary gospel |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | san diego |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | gospel jams |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | gospel soul |
| TRMMWLD128F9301BF2 | ARE592P1187B99F3A7 | Tonéx | syntax records |

Question 2

Query:

```
sqlite> SELECT DISTINCT * FROM tracks WHERE duration > 3020;
```

Output:

| | | |
|--|--|----------------------|
| TRB0GZY128F4269C26 | Hauptteil | Endlich |
| Schlank 0 3020.87791 | ARRXCAY1242077F342 | |
| TRDZTT012903CF1A2E | Grounation | Carried |
| Beyond 0 3034.90567 | ARSSDRQ1241B9CB1A5 | |
| TRGASQU128F92D762F | Rag: Gaud Malhar & Rag: Mian | |
| Malhar Nidhi 0 3030.17751 | ARG62WR1187FB461BC | |
| TRGXALK12903D15504 | Don Carlos: Act III | Verdi: Don |
| Carlos 0 3024.97914 | ARFULUG1272BA849E4 | |
| TRHWGWD12903CD132B | Germans are people too (Alive & Kickin) | Germans Are |
| People too (Alive & Kickin') 0 3027.17342 | ARFIKPQ1272BCD1FD4 | |
| TRJ0ADI12903CC00F9 | Monolithe II | Monolithe |
| II 2005 3024.66567 | AR8XMK11187B9AF8C7 | |
| TRL0KWE12903CF205E | Hud | Le Saint Coran (avec |
| Doae) 0 3026.75546 | ARHDNFW11F4C83BECD | |
| TRMTJMT128F93221B0 | Galactic Odyssey | Galactic |
| Odyssey 0 3029.08036 | ARNHB3M1187B98B512 | |
| TROVFHS128F934D39D | Live Acid Wandererz | Live Acid |
| Wandererz 0 3026.57261 | ARIAXFE11F50C50923 | |
| TROZPG0128F426FD26 | Kunverbainu Mameru – Aakhyaana | Kunverbainu |
| Mameru 0 3022.05342 | ARPLGTP12454A46FA1 | |
| TRPIWVS128F4289D7F | Chapitre Un (a): Toutes Les Histoires | Histoire(s) |
| du cinéma (Complete Soundtrack) 0 3032.76363 | ARIV4271187B9B824F | |
| TRPJBMF128F9355E4C | Members Select Special Bonus Mix (Free Bonus mix | |
| with full album download) | Private Beach Club – By | |
| Afterlife 0 3030.62159 | ARGDNIE1269FCD12AF | |
| TRPWIUP128F426B47B | Der Geist des Llano Estacado Ein Spion | Karl May: |
| Unter Geiern 0 3032.58077 | ARJRKGN12420781485 | |
| TRQCLSM128F42944D9 | Delirium Tremens | Delirium |
| Tremens 0 3024.61342 | ARBNOH41187FB5B059 | |
| TRQRHPV12903CB345A | Galactic Odyssey | Galactic |
| Odyssey 0 3030.90893 | ARKIGPF1187B98BD79 | |
| TRSMLIB128F934C0A8 | Discussion 2 | Inner |
| Voices 0 3033.44281 | ARCRQDC1272BA831EB | |

TRTBHVMV128F4245478|Kamalapadamalar – Harikambhoji – Adi|Sanjay
Subrahmanyam – December Season 2003|0|3022.28853|ARBTWFL122988F01AF
TRTBTAV128F4246BB2|Openin' Doors|Running 4 President 2K8 (2 Disc
Set)|0|3030.02077|AREBBGV1187FB523D2
TRUMEHU128F932D272|Raga: Kaushi Kanada (Tabla: Kumar Bose)|Lingering
Emotions|0|3032.5024|ARI4ARP1187FB50847
TRZSLOD12903CF6FB5|Iraqi Prisoner Abuse|George W. Bush – Taking
America|0|3027.722|ARMDCV21187B9B13AD

Question 3

Query:

```
sqlite> SELECT track_id, year, duration FROM tracks  
...> WHERE year BETWEEN 2010 AND 2014  
...> ORDER BY duration ASC  
...> LIMIT 10;
```

Output:

| | | |
|--------------------|------|----------|
| TRUJRNE12903CCE156 | 2010 | 3.70893 |
| TRLMWPC12903D0EB6D | 2010 | 5.0673 |
| TRPMUA012903CBBEA3 | 2010 | 6.66077 |
| TRTKXQG12903D06EA3 | 2010 | 8.48934 |
| TRCERDU12903D03CF7 | 2010 | 10.86649 |
| TRGVKMD12903CEB193 | 2010 | 11.4673 |
| TRXWWHJ12903CAB222 | 2010 | 11.85914 |
| TRVJQOT12903CEB184 | 2010 | 12.56444 |
| TRADDAQ12903CBBEA4 | 2010 | 14.13179 |
| TRJUJCN12903D06E9F | 2010 | 14.39302 |

Question 4

```
Query: sqlite> SELECT term, COUNT(term) as cnt FROM artist_term  
...> GROUP BY term  
...> ORDER BY cnt DESC  
...> LIMIT 20;
```

Output:

| | |
|------------------|-------|
| rock | 23631 |
| electronic | 20797 |
| pop | 17025 |
| alternative rock | 13206 |
| hip hop | 11980 |
| united states | 11735 |
| house | 10352 |
| jazz | 10100 |
| indie | 9582 |

```
alternative|9545
electro|9488
experimental|8658
indie rock|8613
pop rock|8420
punk|8394
techno|8295
folk|7849
downtempo|7750
ambient|7130
soul|7127
```

Question 5

Query:

```
sqlite> SELECT artist_name, tracks.duration FROM artists
...> INNER JOIN tracks ON artists.artist_id = tracks.artist_id
...> WHERE tracks.duration = (SELECT MAX(tracks.duration) FROM
tracks);
```

Output:

```
Mystic Revelation of Rastafari|3034.90567
```

Question 6

Query:

```
sqlite> SELECT AVG(duration) FROM tracks;
```

Output:

```
249.713383942139
```

Question 7

Query:

```
sqlite> SELECT count(track_id) FROM tracks
...> LEFT JOIN artist_term ON artist_term.artist_id =
tracks.artist_id
...> WHERE artist_term.term IS NULL;
```

Output:

```
3710
```

Question 8

Query:

```
sqlite> SELECT track_id, tracks.artist_id, artists.artist_name,  
artist_term.term FROM tracks  
...> INNER JOIN artists ON artists.artist_id = tracks.artist_id  
...> INNER JOIN artist_term ON artist_term.artist_id =  
artists.artist_id  
...> WHERE track_id='TRMMWLD128F9301BF2';
```

```
sqlite> CREATE INDEX idx_tracks ON tracks (artist_id);  
sqlite> CREATE INDEX idx_artists ON artists (artist_id);  
sqlite> CREATE INDEX idx_term ON artist_term (artist_id);
```

```
sqlite> SELECT track_id, tracks.artist_id, artists.artist_name,  
artist_term.term FROM tracks  
...> INNER JOIN artists ON artists.artist_id = tracks.artist_id  
...> INNER JOIN artist_term ON artist_term.artist_id =  
artists.artist_id  
...> WHERE track_id='TRMMWLD128F9301BF2';
```

Comparison:

The minimum running time of a round without index is 0.8668s while the minimum running time of a round with index is 1.7881e-05s. Thus, an index can help accelerate queries significantly.

Question 9

Query:

```
sqlite> BEGIN;  
sqlite> SELECT COUNT(*) FROM tracks;
```

Output:

866475

Query:

```
sqlite> DELETE FROM tracks  
...> WHERE artist_id IN (SELECT artist_id FROM artist_term WHERE  
term='eurovision winner');
```

```
sqlite> SELECT COUNT(*) FROM tracks;
```

Output:

866291

Query:

```
sqlite> ROLLBACK;  
sqlite> SELECT COUNT(*) FROM tracks;
```

Output:
866475