Big Data Systems (S1-23_DSEOGZG522) Assignment

Group Submission

- **GAGAN GUPTA (2022OG04044)**
- **VINAY SINGLA (2022OG04045)**

Part 2 - OLTP Queries

- ***GITHUB URL : Execution Video***

https://github.com/sinvin-dse/spotify/blob/main/BDS_Assignment_Part_2_OLTP_Queries_GG_VS.mp4

- ***DESCRIBE bds_spotify_recomm_vg.tracks;***

This will display the columns, data types, clustering keys, and other properties of our DB table within **Keyspace**.

```
CREATE TABLE bds_spotify_recomm_vg.tracks (
    id int PRIMARY KEY,
    acousticness float,
    album_name text,
    artists text,
    danceability float,
    duration_ms int,
    energy float,
    explicit boolean,
    instrumentalness float,
    key int,
    liveness float,
    loudness float,
    mode int,
    popularity int,
    speechiness float,
    tempo float,
    time_signature int,
    track_genre text,
    track_id text,
    track_name text,
    valence float
) WITH bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
CREATE INDEX idx_popularity ON bds_spotify_recomm_vg.tracks (popularity);
CREATE INDEX idx_artists ON bds_spotify_recomm_vg.tracks (artists);
CREATE INDEX idx_track_name ON bds_spotify_recomm_vg.tracks (track_name);
CREATE INDEX idx_track_genre ON bds_spotify_recomm_vg.tracks (track_genre);
```

- ***CONSISTENCY;***

```
cqlsh> CONSISTENCY;
Current consistency level is ONE.
```

- ***CONSISTENCY LOCAL_QUORUM;***

```
cqlsh> CONSISTENCY LOCAL_QUORUM;
Consistency level set to LOCAL_QUORUM.
```

As we have configured single node cluster of Cassandra on our local system, so consistency level is ONE. In real time system we will be having multi node & multi cluster system so we can set Consistency level

(**ANY, ONE, QUORUM,LOCAL_QUORUM, ALL**) according to our project needs. Consistency levels determine how and when the database system ensures that a read or write operation is reflected consistently across all nodes in a cluster.

CRUD OPERATIONS

- ***INSERT INTO bds_spotify_recomm_vg.tracks (id, track_id, artists, album_name, track_name, popularity, duration_ms, explicit, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, time_signature, track_genre) VALUES (114004, '2hETkH7cOfqmz3LqZDHZf9', 'Diljit', 'Animal','Arjan Valley', 90, 200000, FALSE, 0.8, 0.7, 2, -5.5, 1, 0.1, 0.2, 0.1, 0.5, 0.6, 111.0, 5, 'r-n-b');***

This query inserts the new record in the DB table with the above specified details. As we have used the Consistency LOCAL_QUORUM before insert query, it will make sure to have the acknowledgement from quorum within the local datacentre.

- ***SELECT * FROM bds_spotify_recomm_vg.tracks WHERE id=114004;***

```
cqlsh> SELECT * FROM bds_spotify_recomm_vg.tracks WHERE id=114004;

 id     | acousticness | album_name | artists | danceability | duration_ms | energy | explicit | instrumentalness | key | liveness | loudness | mode | popul
arity | speechiness | tempo | time_signature | track_genre | track_id               | track_name   | valence
--------+--------------+------------+---------+--------------+-------------+--------+----------+------------------+-----+----------+----------+------+------
--------+-------------+-------+----------------+-------------+------------------------+--------------+---------
 114004 |          0.2 |     Animal |  Diljit |          0.8 |      200000 |    0.7 |    False |              0.1 |   2 |      0.5 |     -5.5 |    1 |
   90 |         0.1 |   111 |              5 |       r-n-b | 2hETkH7cOfqmz3LqZDHZf9 | Arjan Valley |     0.6

(1 rows)
```

This query will retrieve the newly created record using the id 114004.

- ***UPDATE bds_spotify_recomm_vg.tracks SET popularity = 99 WHERE id = 114004;***

This will update the record with ID 114004 with the new popularity value.

- ***SELECT * FROM bds_spotify_recomm_vg.tracks WHERE popularity > 96 ALLOW FILTERING;***

```
cqlsh> SELECT * FROM bds_spotify_recomm_vg.tracks WHERE popularity > 96 ALLOW FILTERING;

 id     | acousticness | album_name                          | artists                  | danceability | duration_ms | energy | explicit | instrumentaln
ess | key | liveness | loudness | mode | popularity | speechiness | tempo     | time_signature | track_genre | track_id               | track_name
                | valence
--------+--------------+-------------------------------------+--------------------------+--------------+-------------+--------+----------+--------------
----+-----+----------+----------+------+------------+-------------+-----------+----------------+-------------+------------------------+-----------------------
-----------------+---------
 114000 |          0.2 |                             Deewana |                Sonu Nigam |          0.8 |      200000 |    0.7 |    False |
0.1 |   2 |      0.5 |     -5.5 |    1 |         99 |         0.1 |       111 |              5 |         pop | 2hETkH7cOfqmz3LqZDHZf6 |
        Deewana |     0.6
  20001 |        0.013 |                Unholy (feat. Kim Petras) |       Sam Smith;Kim Petras |        0.714 |      156943 |  0.472 |    False |         4.51e
-06 |   2 |    0.266 |   -7.375 |    1 |        100 |      0.0864 |   131.121 |              4 |       dance | 3nqQXoyQOWXiESFLlDF1hG |              Unholy (
feat. Kim Petras) |   0.238
  30003 |      0.00383 |                       I'm Good (Blue) |    David Guetta;Bebe Rexha |        0.561 |      175238 |  0.965 |     True |         7.07e
-06 |   7 |    0.371 |   -3.673 |    0 |         98 |      0.0343 | 128.03999 |              4 |         edm | 4uUG5RXrOk84mYEfFvj3cK |
  I'm Good (Blue) |   0.304
 114004 |          0.2 |                              Animal |                   Diljit |          0.8 |      200000 |    0.7 |    False |
0.1 |   2 |      0.5 |     -5.5 |    1 |         99 |         0.1 |       111 |              5 |       r-n-b | 2hETkH7cOfqmz3LqZDHZf9 |
     Arjan Valley |     0.6
  88410 |        0.583 |                         La Bachata |                Manuel Turizo |        0.835 |      162637 |  0.679 |    False |         1.98e
-06 |   7 |    0.218 |   -5.329 |    0 |         98 |      0.0364 |    124.98 |              4 |      reggae | 5ww2BF9slyYgNOk37BlC4u |
       La Bachata |    0.85
  88405 |       0.0901 |                   Un Verano Sin Ti | Bad Bunny;Chencho Corleone |        0.911 |      178567 |  0.712 |     True |          2.7e
-05 |   1 |   0.0933 |   -5.105 |    0 |         97 |      0.0817 |    92.005 |              4 |      reggae | 6Sq7ltF9Qa7SNFBsV5Cogx |
  Me Porto Bonito |   0.425
  89411 |        0.583 |                         La Bachata |                Manuel Turizo |        0.835 |      162637 |  0.679 |    False |         1.98e
-06 |   7 |    0.218 |   -5.329 |    0 |         98 |      0.0364 |    124.98 |              4 |   reggaeton | 5ww2BF9slyYgNOk37BlC4u |
       La Bachata |    0.85
  89407 |       0.0901 |                   Un Verano Sin Ti | Bad Bunny;Chencho Corleone |        0.911 |      178567 |  0.712 |     True |          2.7e
-05 |   1 |   0.0933 |   -5.105 |    0 |         97 |      0.0817 |    92.005 |              4 |   reggaeton | 6Sq7ltF9Qa7SNFBsV5Cogx |
  Me Porto Bonito |   0.425
```

This will filter the records where popularity is greater than 96. ALLOW FILTERING is a directive used in queries that involve filtering on non-indexed columns. The use of ALLOW FILTERING should be avoided whenever possible, especially in production environments, as it can lead to inefficient queries and performance degradation.