Student Name: **Chuqi Wang** UCInetID: **chuqiw4**

1.

a) CQL Query: DESCRIBE hoofers;

b) Result:

```
token@cqlsh> DESCRIBE hoofers;

CREATE KEYSPACE hoofers WITH replication = {'class': 'NetworkTopologyStrategy', 'us-east1': '3'}  AND durable_writes = true;

CREATE TABLE hoofers.boats (
    bid int PRIMARY KEY,
    bname text,
    color text
) WITH additional_write_policy = '99p'
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.UnifiedCompactionStrategy'}
    AND compression = {'chunk_length_in_kb': '16', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    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 = 'BLOCKING'
    AND speculative_retry = '99p';
```

c) Answers:

The Hoofers keyspace maintain 3 copies of data, it resides in us-east1 cloud region. The read quorum size (R) and write quorum size (W) for consistent reads and writes are both 2 since read quorum = write quorum = 3/2 + 1 = 2.

2.

a) CQL CREATE Statements:

```
CREATE TABLE Users (
        user_id text,
        email text,
        joined_date date,
        nickname text,
        street text,
        city text,
        state text,
        zip text,
        genres text,
        PRIMARY KEY (user_id)
);
```

```
CREATE TABLE Records (
        record_id text,
        artist_user_id text,
        title text,
        genre text,
        release_date date,
        PRIMARY KEY (record_id)
);

CREATE TABLE Reviews (
        review_id text,
        user_id text,
        record_id text,
        rating int,
        body text,
        posted_at timestamp,
        PRIMARY KEY (review_id)
);

CREATE TABLE Sessions (
        session_id text,
        user_id text,
        record_id text,
        track_number int,
        initiate_at timestamp,
        leave_at timestamp,
        music_quality text,
        device text,
        remaining_time int,
        replay_count int,
        PRIMARY KEY (session_id)
);
```

3.

a) PostgreSQL COPY commands:

\COPY zotmusic.Users to 'Users/chuqiwang/Desktop/UCI/CS224P/assignments/HW2/setup/users.csv' DELIMITER ',' CSV HEADER;

\COPY zotmusic.Records to 'Users/chuqiwang/Desktop/UCI/CS224P/assignments/HW2/setup/records.csv' DELIMITER ',' CSV HEADER;

\COPY zotmusic.Reviews to 'Users/chuqiwang/Desktop/UCI/CS224P/assignments/HW2/setup/reviews.csv' DELIMITER ',' CSV HEADER;

\COPY zotmusic.Sessions to 'Users/chuqiwang/Desktop/UCI/CS224P/assignments/HW2/setup/sessions.csv' DELIMITER ',' CSV HEADER;

4.

a) First CQL Query:

SELECT record_id, title, genre

FROM Records

WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3';

b) Result:

```
token@cqlsh:zotmusic> SELECT record_id, title, genre
           ... FROM Records
           ... WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3';                                    Region: us-east1
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you wa
nt to execute this query despite the performance unpredictability, use ALLOW FILTERING"
token@cqlsh:zotmusic>
```

c) Modified CQL Query:

SELECT record_id, title, genre FROM Records

WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3' ALLOW FILTERING;

b) Result:

```
token@cqlsh:zotmusic> SELECT record_id, title, genre FROM Records
            ... WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3' ALLOW FILTERING;

 record_id                                    | title                        | genre
----------------------------------------------+------------------------------+---------
 record_62389a63-e95f-43d1-acea-aa1bac0e0050 |            Result guess for |  Gospel
 record_116fbdd6-e706-41f7-9809-12e174e48e8f |               Discover rate |    Jazz
 record_2406e933-23e3-4db1-acf9-3c863d48bff6 |            General job heavy | Country
 record_3e4ed054-cf1a-4a04-8e97-e0177c6d3575 | Summer civil political beat |    Folk
 record_91c6325d-b17f-4f4c-be6d-3517b2173a9f |             Statement matter | Country
 record_57061d35-de20-4bf1-9aac-a689f0db7e16 |              Would determine |    Soul
 record_822961a3-946a-49ff-8173-74d4035286b9 |                   Apply size |  Gospel
 record_cbf93efd-2deb-48ae-ad73-83aa088c6f13 |              Democratic what |    Soul
 record_5cbf14c7-7b54-4e32-bfce-cba507c7277f |                Bar talk long |    Jazz

(9 rows)
token@cqlsh:zotmusic>
```

5.

a) CQL Create Statement:

```
CREATE TABLE Records_q5 (
        record_id text,
        artist_user_id text,
        title text,
        genre text,
        release_date date,
        PRIMARY KEY (artist_user_id, record_id)
);
```

b) CQL Query:

```
SELECT record_id, title, genre
FROM Records_q5
WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3';
```

c) Result:

```
token@cqlsh:zotmusic> SELECT record_id, title, genre
           ... FROM Records_q5
           ... WHERE artist_user_id = 'user_6ac27408-a0a6-4c57-a025-7b6854f7a8e3';

 record_id                                    | title                        | genre
----------------------------------------------+------------------------------+---------
 record_116fbdd6-e706-41f7-9809-12e174e48e8f |                 Discover rate |    Jazz
 record_2406e933-23e3-4db1-acf9-3c863d48bff6 |              General job heavy | Country
 record_3e4ed054-cf1a-4a04-8e97-e0177c6d3575 | Summer civil political beat |    Folk
 record_57061d35-de20-4bf1-9aac-a689f0db7e16 |                Would determine |    Soul
 record_5cbf14c7-7b54-4e32-bfce-cba507c7277f |                  Bar talk long |    Jazz
 record_62389a63-e95f-43d1-acea-aa1bac0e0050 |              Result guess for |  Gospel
 record_822961a3-946a-49ff-8173-74d4035286b9 |                    Apply size |  Gospel
 record_91c6325d-b17f-4f4c-be6d-3517b2173a9f |              Statement matter | Country
 record_cbf93efd-2deb-48ae-ad73-83aa088c6f13 |               Democratic what |    Soul

(9 rows)
token@cqlsh:zotmusic>
```

d) Explanation:

Changing the partitioning key to artist_user_id did altered Cassandra's behavior because Cassandra is set to use partition key to find the location of data. In question 4, artist_user_id is not partition key, so Cassandra needs to scan multiple partitions to locate the data and it need allow filtering. And why we also need to include record_id in the primary key is because PRIMARY KEY (record_id, artist_user_id) ensures each record is unique for a given artist. If we don't include record_id in the primary key, then record_id will not be unique, but record_id must be unique.

6.

a) CQL Query:

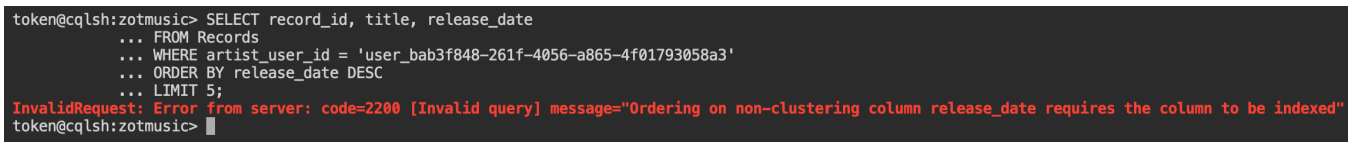
```
SELECT record_id, title, release_date
FROM Records
WHERE artist_user_id = 'user_bab3f848-261f-4056-a865-4f01793058a3'
ORDER BY release_date DESC
LIMIT 5;
```

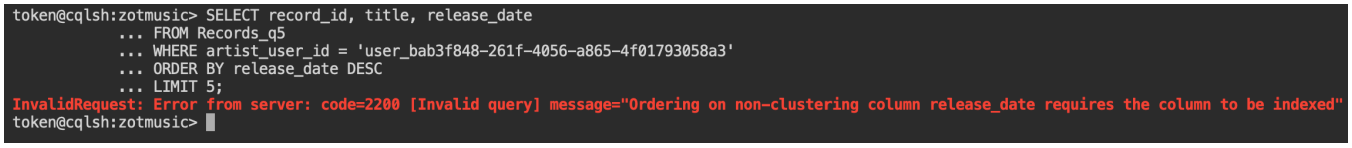
```
SELECT record_id, title, release_date
FROM Records_q5
WHERE artist_user_id = 'user_bab3f848-261f-4056-a865-4f01793058a3'
ORDER BY release_date DESC
LIMIT 5;

SELECT record_id, title, release_date
FROM Records_q6
WHERE artist_user_id = 'user_bab3f848-261f-4056-a865-4f01793058a3'
LIMIT 5;
```

b) CQL CREATE Statement:

```
CREATE TABLE Records_q6 (
        record_id text,
        artist_user_id text,
        title text,
        genre text,
        release_date date,
        PRIMARY KEY (artist_user_id, release_date, record_id)
) WITH CLUSTERING ORDER BY (release_date DESC);
```

c) Results:





```
token@cqlsh:zotmusic> SELECT record_id, title, release_date
           ... FROM Records_q6
           ... WHERE artist_user_id = 'user_bab3f848-261f-4056-a865-4f01793058a3'
           ... LIMIT 5;

 record_id                                    | title                  | release_date
----------------------------------------------+------------------------+--------------
 record_ff3420fe-cf7f-43f9-9131-1965883acc51 |    Plant worker doctor |   2022-10-07
 record_4f4f27a7-03f8-4adb-b96e-93adf9eb4e62 | Under total throughout |   2022-09-08
 record_2c83cd72-4450-4936-bb9d-1732ced5a166 |         Money material |   2022-02-18
 record_8c389ed6-2101-489d-a301-b66ab43ff51c |          Discover fast |   2021-08-22
 record_25778e66-e835-4347-9d4e-98a48f8424a1 |               Way real |   2021-04-03

(5 rows)
token@cqlsh:zotmusic>
```

d) Explanation:

Adding release_date as a clustering key and specifying a clustering order by (release_date DESC) ensures Cassandra fill in the data for each artist and sorted by their release_date. In PRIMARY KEY (artist_user_id, release_date, record_id), release_date is the clustering key and record_id is the second clustering key. Including record_id ensures each records is unique within the same partition.

7.

a) CQL Create Statement:

```
CREATE TABLE Reviews_q7a (
        review_id text,
        user_id text,
        record_id text,
        rating int,
        body text,
        posted_at timestamp,
        PRIMARY KEY (user_id, rating, review_id)
) WITH CLUSTERING ORDER BY (rating DESC);
```

b) CQL Create Statement:

```
CREATE TABLE Records_q7b (
   artist_user_id text,
   record_id text,
   title text,
```

```
    genre text,
    release_date date,
    PRIMARY KEY (genre, record_id)
);
```

c) CQL Create Statement:

```
CREATE TABLE Reviews_q7c (
    artist_user_id text,
    posted_at timestamp,
    review_id text,
    record_id text,
    title text,
    rating int,
    PRIMARY KEY (artist_user_id, posted_at, review_id)
) WITH CLUSTERING ORDER BY (posted_at DESC);
```

d) CQL Create Statement:

```
CREATE TABLE Sessions_q7d (
    user_id text,
    record_id text,
    initiate_at timestamp,
    leave_at timestamp,
    session_id text,
    track_number int,
    replay_count int,
    music_quality text,
    device text,
    remaining_time int,
    PRIMARY KEY (user_id, initiate_at, session_id)
) WITH CLUSTERING ORDER BY (initiate_at ASC);
```

8.

a)

- CQL Query:

```
SELECT review_id, record_id
FROM Reviews_q7a
WHERE user_id = 'user_9e48cbb4-0bf9-43fc-a578-213fae51068b' LIMIT 10;
```

- Result:

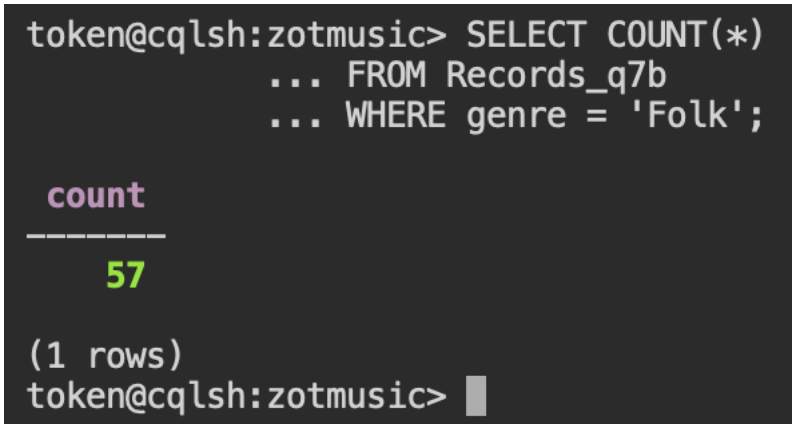
```
token@cqlsh:zotmusic> SELECT review_id, record_id
           ... FROM Reviews_q7a
           ... WHERE user_id = 'user_9e48cbb4-0bf9-43fc-a578-213fae51068b' LIMIT 10;

 review_id                                    | record_id
----------------------------------------------+----------------------------------------------
 review_23801293-5f92-4159-805f-d2a4809027f1 | record_6c4b5698-4cb2-47a3-acb2-5161e51e2b47
 review_23e212cf-70cf-4f5e-a151-5650e585074e | record_04b59df9-af69-423b-90c9-35bf9148835c
 review_405b9a5f-22fd-433f-bf97-4fbbe7e1dd79 | record_591478b0-1fa1-4f90-9094-76586f44066a
 review_785f16cb-1d01-44b7-9ef1-13c4de101196 | record_99f7b9ad-7969-4faf-9285-162f50c1578c
 review_c0215ff1-50ca-4b07-a722-50d4d998498b | record_e4b025e5-69f2-4ef7-a2b1-8b786794c009
 review_c9b4a223-5b83-4a77-897b-c35d8ffd467b | record_2c9d0def-b9e4-4577-a023-3aad8ce11806
 review_fc195a51-6fff-472e-86ae-d749dad03389 | record_c3f90c67-bf41-4f04-89cc-afbfa663a4cc
 review_2a193df6-6a49-461e-b347-916663d656f3 | record_92179aa1-4d00-4d1e-8ffb-fbbc96c196e9
 review_2c28ef7e-b611-42e4-ae49-0278c5f303ac | record_c3504b5f-2504-44a7-af79-87968eaceaac
 review_37df2d42-83b4-4fbe-9b30-15ee98878f8e | record_8c389ed6-2101-489d-a301-b66ab43ff51c

(10 rows)
token@cqlsh:zotmusic>
```

b)

- CQL Query:

```
SELECT COUNT(*)
FROM Records_q7b
WHERE genre = 'Folk';
```

- Result:



```
token@cqlsh:zotmusic> SELECT COUNT(*)
           ... FROM Records_q7b
           ... WHERE genre = 'Folk';

 count
-------
    57

(1 rows)
token@cqlsh:zotmusic>
```

c)

- CQL Query:

```
SELECT review_id, record_id, title, rating
FROM Reviews_q7c
WHERE artist_user_id = 'user_6f33f39e-7659-4673-bd80-ca11394424b0' LIMIT 10;
```

- Result:

```
token@cqlsh:zotmusic> SELECT review_id, record_id, title, rating
           ... FROM Reviews_q7c
           ... WHERE artist_user_id = 'user_6f33f39e-7659-4673-bd80-ca11394424b0' LIMIT 10;

 review_id                                | record_id                                | title                | rating
------------------------------------------+------------------------------------------+----------------------+--------
 review_658cf2aa-b73f-4da6-b5f0-a9cf9edb65b5 | record_1e0d5cc2-2593-4718-94a4-a87a2bc73878 |            Free say |      3
 review_346885af-b687-4d5a-82dc-e21dd6458390 | record_f9220687-10b1-469c-888b-e306ed7a3376 |   Successful but up |      2
 review_3b74edf5-dc86-48ca-afe6-cd12c3677aa1 | record_324b326c-9cd7-4cb7-8064-b431d3d96f4c |   Police whose goal |      1
 review_86418842-4609-47b4-891c-c7a5c1d1ccb4 | record_f274db0f-3a28-4781-9a00-dbacf5b902bf | Pick administration |      4
 review_85696c26-1a1b-44bc-bd61-50e5ed9a28d1 | record_324b326c-9cd7-4cb7-8064-b431d3d96f4c |   Police whose goal |      3
 review_39517c83-a010-49a4-b276-d58e423e655b | record_f0fcf759-2df7-437b-b497-b6cc4a3d726e |    Hour myself seat |      2
 review_d828453b-13b1-420b-9921-534a8a508936 | record_e793d30f-958f-499d-a8de-170732db99cd |         They rather |      5
 review_e6a3afab-f7bc-40fd-8b71-1883e8595baa | record_f9220687-10b1-469c-888b-e306ed7a3376 |   Successful but up |      1
 review_8361944a-193c-465e-a0aa-c13349eb29e5 | record_f274db0f-3a28-4781-9a00-dbacf5b902bf | Pick administration |      5
 review_ef32d5f6-f5cf-4816-baa5-208d3ebfd966 | record_1e0d5cc2-2593-4718-94a4-a87a2bc73878 |            Free say |      1

(10 rows)
token@cqlsh:zotmusic>
```

d)

- CQL Query:

```
SELECT MAX(replay_count)
FROM Sessions_q7d
WHERE user_id = 'user_05f9132b-47fb-4d2b-992c-17b3c4afb2df'
AND initiate_at >= '2024-08-01 00:00:00'
AND initiate_at <= '2024-09-01 00:00:00';
```

- Result:

```
token@cqlsh:zotmusic> SELECT MAX(replay_count)
           ... FROM Sessions_q7d
           ... WHERE user_id = 'user_05f9132b-47fb-4d2b-992c-17b3c4afb2df'
           ... AND initiate_at >= '2024-08-01 00:00:00'
           ... AND initiate_at <= '2024-09-01 00:00:00';

 system.max(replay_count)
--------------------------
                        4

(1 rows)
token@cqlsh:zotmusic>
```

9.

- CQL INSERT statements:

INSERT INTO Records (record_id, artist_user_id, title, genre, release_date)

VALUES ('record_d2f498f8-d7ff-4f1c-a967-7090417751f5', 'user_38eaa9f8-e8fc-4ce4-a8ae-ffb882c1786c', 'Blue By You', 'Rock', '2024-10-07');

INSERT INTO Records_q7b (genre, record_id, artist_user_id, title, release_date)

VALUES ('Rock', 'record_d2f498f8-d7ff-4f1c-a967-7090417751f5', 'user_38eaa9f8-e8fc-4ce4-a8ae-ffb882c1786c', 'Blue By You', '2024-10-07');

- Verification queries:

SELECT *

FROM Records

WHERE record_id = 'record_d2f498f8-d7ff-4f1c-a967-7090417751f5';

SELECT *

FROM Records_q7b

WHERE record_id = 'record_d2f498f8-d7ff-4f1c-a967-7090417751f5'

AND genre = 'Rock';

- Result:

```
token@cqlsh:zotmusic> SELECT *
           ... FROM Records
           ... WHERE record_id = 'record_d2f498f8-d7ff-4f1c-a967-7090417751f5';

 record_id                                | artist_user_id                           | genre | release_date | title
------------------------------------------+------------------------------------------+-------+--------------+-------------
 record_d2f498f8-d7ff-4f1c-a967-7090417751f5 | user_38eaa9f8-e8fc-4ce4-a8ae-ffb882c1786c |  Rock |   2024-10-07 | Blue By You

(1 rows)
token@cqlsh:zotmusic>
```

```
token@cqlsh:zotmusic> SELECT *
           ... FROM Records_q7b
           ... WHERE record_id = 'record_d2f498f8-d7ff-4f1c-a967-7090417751f5'
           ... AND genre = 'Rock';

 genre | record_id                                | artist_user_id                           | release_date | title
-------+------------------------------------------+------------------------------------------+--------------+-------------
  Rock | record_d2f498f8-d7ff-4f1c-a967-7090417751f5 | user_38eaa9f8-e8fc-4ce4-a8ae-ffb882c1786c |   2024-10-07 | Blue By You

(1 rows)
token@cqlsh:zotmusic>
```

10.

Python script:

```
from cassandra.cluster import Cluster
from cassandra.auth import PlainTextAuthProvider

def connect_astra():
    cloud_config = {
        'secure_connect_bundle': '/Users/chuqiwang/Desktop/UCI/CS224P/assignments/HW2/setup/secure-connect-cs224p-fall.zip'
    }
    auth_provider = PlainTextAuthProvider('SeNvcKCScXavolPBtUFOnUXX',
                                          'Zk0,qmc9uITdUNs7X1sZrIt-StdX1lloZyrzxy6fFxwUhnJc84foe+lmksNRzfzd4kr_tFSN_RBlhaD5kCSl,Z0fZznU9OUqfpaJw3JQ,13XkMCPSDN9jrmj7JuzvLyz')
    cluster = Cluster(cloud=cloud_config, auth_provider=auth_provider)
    session = cluster.connect('zotmusic')
    return session


def insert_record(session):
    record_id = 'record_632fe768-eecb-4596-9780-cc21734feec5'
    artist_user_id = 'user_b91cf915-487b-42fc-b6b8-6c17935bb755'
    title = 'One Sour Day'
    genre = 'R&B'
    release_date = '2024-10-07'

    session.execute(
        """
        INSERT INTO Records (record_id, artist_user_id, title, genre, release_date)
        VALUES (%s, %s, %s, %s, %s)
        """,
        (record_id, artist_user_id, title, genre, release_date)
    )

    session.execute(
        """
        INSERT INTO RecordsByGenre (genre, record_id, artist_user_id, title, release_date)
        VALUES (%s, %s, %s, %s, %s)
        """,
        (genre, record_id, artist_user_id, title, release_date)
    )
```

```
def main():
    session = connect_astra()
    insert_record(session)
    session.shutdown()

if __name__ == "__main__":
  main()
```

- Result:

```
token@cqlsh:zotmusic> SELECT *
           ... FROM Records
           ... WHERE record_id = 'record_632fe768-eecb-4596-9780-cc21734feec5';

 record_id                                | artist_user_id                           | genre | release_date | title
------------------------------------------+------------------------------------------+-------+--------------+-------------
 record_632fe768-eecb-4596-9780-cc21734feec5 | user_b91cf915-487b-42fc-b6b8-6c17935bb755 |   R&B |   2024-10-07 | One Sour Day

(1 rows)
token@cqlsh:zotmusic>
```

```
token@cqlsh:zotmusic> SELECT *
           ... FROM Records_q7b
           ... WHERE record_id = 'record_632fe768-eecb-4596-9780-cc21734feec5'
           ... AND genre = 'R&B';

 genre | record_id                                | artist_user_id                           | release_date | title
-------+------------------------------------------+------------------------------------------+--------------+-------------
   R&B | record_632fe768-eecb-4596-9780-cc21734feec5 | user_b91cf915-487b-42fc-b6b8-6c17935bb755 |   2024-10-07 | One Sour Day

(1 rows)
token@cqlsh:zotmusic>
```