Question 1. In this question, you are asked the reverse of previous homework. Improve the expected run time of two queries from either Homework #4 or Homework #5 with respect to the mediumstreaming database.

Note: the method I choose is to create one or two indices and use my solution.

HW4 Q1

```
SELECT
        s.title,
        sd.director
FROM
        series s,
        seriesdirectors sd
WHERE
        s.seriesid=sd.seriesid
        and s.seasons>=15
        and s.imdbrating<5
ORDER BY
        s.title,
        sd.director;
```

- index creation commands

```
create index query1 on series(seriesid, imdbrating, seasons);
```

- a few lines of the query plan before you create the index and

```
Sort  (cost=286.07..286.08 rows=1 width=30)
  Sort Key: s.title, sd.director
  ->  Hash Join  (cost=283.42..286.06 rows=1 width=30)
        Hash Cond: (sd.seriesid = s.seriesid)
        ->  Seq Scan on seriesdirectors sd  (cost=0.00..2.30 rows=130 width=17)
        ->  Hash  (cost=283.39..283.39 rows=2 width=21)
              ->  Seq Scan on series s  (cost=0.00..283.39 rows=2 width=21)
                    Filter: ((seasons >= 15) AND (imdbrating < '5'::double precision))
```

- full query plan after you create the index.

```
Sort  (cost=97.87..97.88 rows=1 width=30)
Sort Key: s.title, sd.director
->  Hash Join  (cost=95.22..97.86 rows=1 width=30)
Hash Cond: (sd.seriesid = s.seriesid)
->  Seq Scan on seriesdirectors sd  (cost=0.00..2.30 rows=130 width=17)
->  Hash  (cost=95.19..95.19 rows=2 width=21)
 ->  Index Scan using query1 on series s  (cost=0.28..95.19 rows=2 width=21)
 Index Cond: ((imdbrating < '5'::double precision) AND (seasons >= 15))
```

HW4 Q2

```
SELECT
        count(*) as nummovies
FROM
        movies m
WHERE
        m.imdbrating IS NULL
        and m.rottentomatoes IS NULL
        and (m.year IS NULL or m.year>2015);
```

- index creation commands

```
create index query2 on movies(movieid, imdbrating, rottentomatoes, year);
```

- a few lines of the query plan before you create the index and

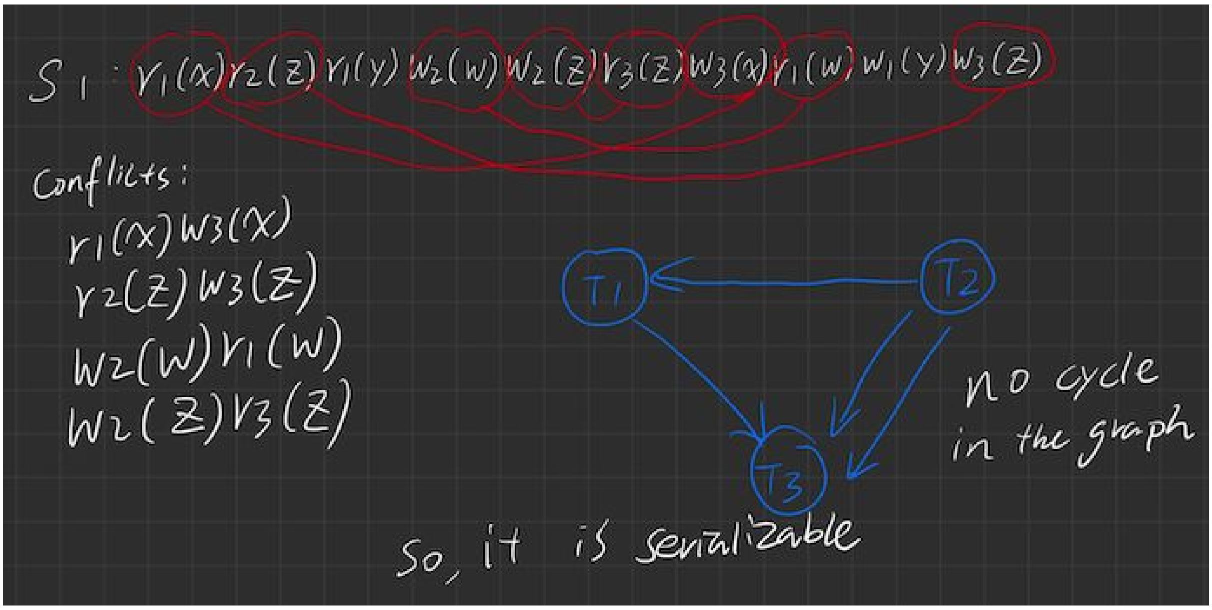
```
Aggregate  (cost=120.68..120.69 rows=1 width=8)
->  Seq Scan on movies m  (cost=0.00..120.61 rows=27 width=0)
    Filter: ((imdbrating IS NULL) AND (rottentomatoes IS NULL) AND ((year IS NULL)
```

- full query plan after you create the index.

```
        Aggregate  (cost=115.89..115.90 rows=1 width=8)
->  Index Only Scan using query2 on movies m  (cost=0.28..115.82 rows=27 width=0)
      Index Cond: ((imdbrating IS NULL) AND (rottentomatoes IS NULL))
      Filter: ((year IS NULL) OR (year > 2015))
```

Question 2. You are given the following schedules. For each schedule,

S1: r1(x) r2(z) r1(y) w2(w) w2(z) r3(z) w3(x) r1(w) w1(y) w3(z)
(a) check if it is serializable by drawing the conflict graph, and



S1: r1(x) r2(z) r1(y) w2(w) w2(z) r3(z) w3(x) r1(w) w1(y) w3(z)

Conflicts:
r1(x) w3(x)
r2(z) w3(z)
w2(w) r1(w)
w2(z) r3(z)

no cycle in the graph

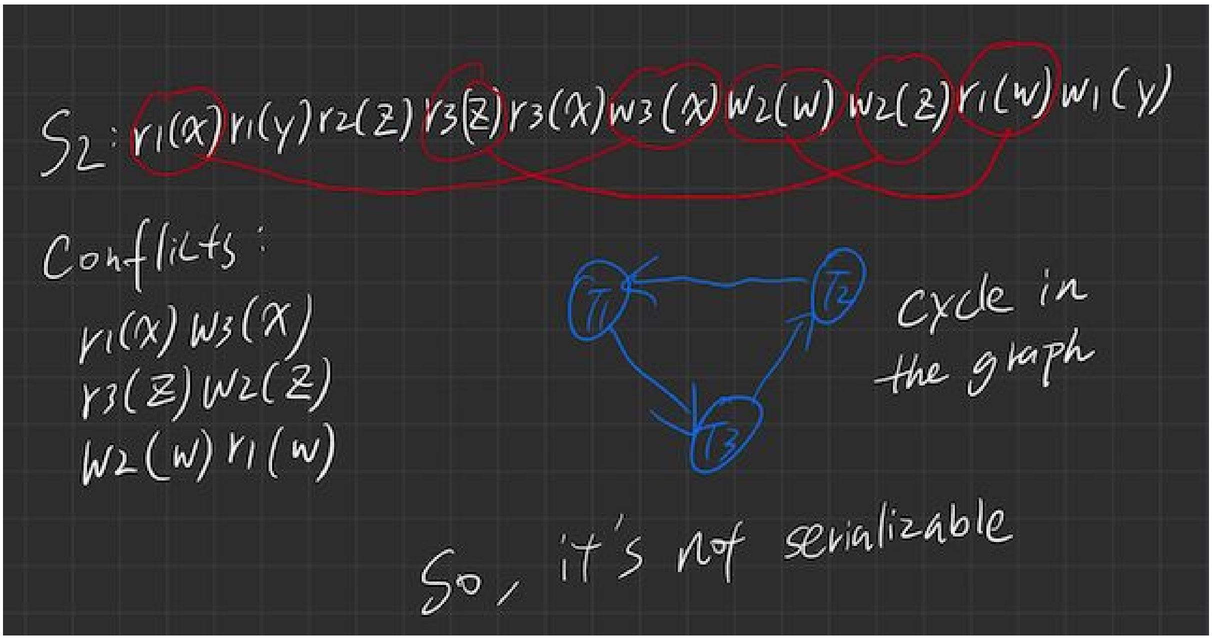
So, it is serializable

The topological order will be T2-T1-T3

(b) discuss if it is possible to obtain this schedule using Two-Phase Locking.

It is not possible to obtain this schedule using two-phase locking. Between r3(z) and w3(x) T3 will be shrinking and cannot acquire lock after w3(x)

S2: r1(x) r1(y) r2(z) r3(z) r3(x) w3(x) w2(w) w2(z) r1(w) w1(y)
(a) check if it is serializable by drawing the conflict graph, and



S2: r1(x) r1(y) r2(z) r3(z) r3(x) w3(x) w2(w) w2(z) r1(w) w1(y)

Conflicts:
r1(x) w3(x)
r3(z) w2(z)
w2(w) r1(w)

cycle in the graph

So, it's not serializable

(b) discuss if it is possible to obtain this schedule using Two-Phase Locking.

Not possible to obtain this schedule using two-phase locking because the schedule us not serializable.

Question 3. Suppose you are using REDO/UNDO recovery, and the following are the contents of the log and the disk after the crash.

(a) Which log entries should be redone, which should be undone, and in which order?

Redo:
LSN 103
LSN 105
Undo:
LSN 105
LSN 100

(b) Based on this information, can you conclude if FORCE or NO FORCE is used? Discuss.

NO FORCE is used. T3 was partially written to data page after committing which is not possible with FORCE.

(c) Based on this information, can you conclude if STEAL or NO STEAL is used? Discuss.

STEAL is used. T1 was not committed but the data page was changed.