CS 348: Assignment 5 (Spring 2024)

Assigned on July 16 and due by 5:00pm EDT on July 29

Question 1.
Use Armstrong's axioms to prove the soundness of the union and decomposition rules (Module 9, slide 14).

Question 2.
Prove that if $A \notin ComputeX^+(X, F)$ then $X \to A \notin F^+$ (you must prove this without using/referring to the *Correctness theorem* for $ComputeX^+(X, F)$ introduced in Module 9, slide 17).

Question 3.
Translate each of the following SQL queries over the relational database schema for bibliography information given on Slide 10 of Module 2 to formulations in the relational algebra with multiset semantics defined in Module 10.

1. *All book titles.*

   ```
   select distinct title
   from publication, book
   where publication.pubid = book.pubid
   ```

2. *All authors with at least two publications.*

   ```
   select distinct r1.author
   from wrote r1, wrote r2
   where not r1.publication = r2.publication
   and r1.author = r2.author
   ```

3. *All author-publication ids for all publications except proceedings and journals.*

   ```
   select * from wrote
   where publication not in (
      (select pubid from proceedings)
      union
      (select pubid from journal))
   ```

Question 4.
Assume the DBA group has chosen the following standard physical design (as defined in Module 10) for the relational database schema for bibliography information given on Slide 10 of Module 2:

- A primary index called T-A is chosen for each table T, where A is the attribute name of the first column of T. The chosen data structure in each case is a Btree on A. For example, the primary index of table `WROTE` is called `WROTE-AUTHOR` and is a Btree on `author`, and, for querying, has arity 3:

  `WROTE-AUTHOR / (rid, author, publication).`

  Like all indices, recall that attribute `rid` is prepended to the list of attributes for `WROTE-AUTHOR`.

- A secondary index called `WROTE-PUBLICATION` is chosen for table `WROTE` that is a Btree on attribute `publication`, and, for querying, also has arity 3

  `WROTE-PUBLICATION / (rid, publication, wrote-author-rid).`

  Recall that there is a foreign key on attribute `wrote-author-rid` to attribute `rid` of primary key table `WROTE-AUTHOR`.

- A secondary index called `ARTICLE-APPEARS-IN` is chosen for table `ARTICLE` that is a Btree on attribute `appears-in`, and, for querying, has arity 3

  `ARTICLE-APPEARS-IN / (rid, appears-in, author-pubid-rid).`

Translate each of the following queries in SQL to an RA expression over this physical design, that is, in which all relation names are names of indices. Your RA expression should be as efficient as possible, assuming the implementation of operations and simple cost model given in Module 10. For example, you should consider using nested index scans. Recall from Slide 39 of Module 10 that these are expressed with the syntax

$$E \times \sigma_{\#i=\#j\ell}(R),$$

where $\#j$ refers to a column of a subexpression E and where R is an index for which $\#i$ is its `rid` column or its search key column (i.e., either column #1 or #2 for the primary index `WROTE-AUTHOR`).

1. *Names of all authors of books published in 2019 last year.*

```
select distinct name
from AUTHOR a
where exists (
   select * from WROTE w, BOOK b
   where w.publication = b.pubid and w.author = a.aid
   and b.year = 2019 )
```

2. *All publication identifiers of articles appearing in volume 1 number 1 of the journal ACM TODS.*

```
select distinct pubid
from ARTICLE a
where exists (
   select * from JOURNAL j, PUBLICATION p
   where j.pubid = p.pubid and p.title = 'ACM TODS'
   and j.volume = 1 and j.number = 1
   and a.appears-in = j.pubid )
```

Question 5.
Determine whether or not each of the following four transaction execution schedules is conflict serializable. If a schedule is conflict serializable, specify a serial order of transaction execution to which it is equivalent.

$$S_1 = r_1[x], r_2[y], w_2[x], r_1[z], r_3[z], w_3[z], w_1[z]$$
$$S_2 = w_1[x], w_1[y], r_2[u], w_2[x], r_2[y], w_2[y], w_1[z]$$
$$S_3 = w_1[x], w_1[y], r_2[u], w_1[z], w_2[x], r_2[y], w_1[u]$$
$$S_4 = w_1[x], w_2[u], w_2[y], w_1[y], w_3[x], w_3[u], w_1[z]$$

Question 6.
Suppose that after a system failure, the transaction log looks as shown below (the log tail is at the bottom). Describe what the database system must do to recover from the system failure. Indicate which objects must be modified, and in what order those modifications occur. Indicate which transactions are committed and which are aborted after the failure recovery is complete.

T_1, BEGIN
T_1, UNDO, x, 0
T_1, REDO, x, 10
T_2, BEGIN
T_2, UNDO, y, 10
T_2, REDO, y, 20
T_3, BEGIN
T_1, COMMIT
T_3, UNDO, x, 10
T_3, REDO, x, 400
T_4, BEGIN
T_3, UNDO, z, 0
T_3, REDO, z, 100
T_5, BEGIN
T_4, UNDO, a, 0
T_4, REDO, a, 1
T_6, BEGIN
T_4, ABORT
T_5, UNDO, a, 0
T_5, REDO, a, 2
T_3, COMMIT
T_6, UNDO, x, 400
T_6, REDO, x, 0
T_5, COMMIT
T_6, UNDO, a, 2
T_6, REDO, a, 3

Assignment Submission
Assignment submission should be a single file uploaded to Crowdmark with each question starting on a new page. The file should be a pdf file that may be computer generated or a scan/photo of a handwritten solution, as long as it is legible.