CMPSCI 645: Homework 4 - Part B

Due: Monday, April 9 2018, 11:59 pm

Prof. Yanlei Diao

Name: Mili Shah

Question	Points	Score
5	16	
6	24	
Total	40	

Please turn in this homework electronically, as a PDF, through Gradescope. You may handwrite your solutions, and then scan the document, or type directly into the PDF form. Make sure the PDF you upload includes all pages (including this front page).

Question 5 [3 parts, 16 points] Query Optimization
Consider the following schema

Sailors(sid, sname, rating, age)
Boats(bid, bname, size)
Reserves(sid, bid, day)

Reserves.sid is a foreign key to Sailors and Reserves.bid is a foreign key to Boats.bid.

We are given the following information about the database:
Sailors contains 50 pages with 20 records per page, so 1000 records in total.
Boats contains 10 pages with 10 records per page, so 100 records in total.
Reserves contains 250 pages with 40 records per page, so 10,000 records in total.
There are 100 values for Reserves.sid.
There are 50 values for Reserves.bid.
There are 1000 values for Reserves.day

In the following queries, assume that a System R style optimizer is used.

Consider Query 1:

```
SELECT S.sid, S.sname, B.bname
FROM    Sailors S, Reserves R, Boats B
WHERE S.sid = R.sid AND R.bid = B.bid AND R.day = 'July 4, 2003';
```

(a) [4 points] Assuming uniform distribution of values and column independence, estimate the number of tuples returned by this query.

Result cardinality = max-num-tuples $\times \prod_i RF_i$

max-num-tuples = $1000 \times 100 \times 10{,}000 = 10^9$

RF for R.day = 'July 4, 2003' : $\frac{1}{1000}$

RF for S.sid = R.sid : $\frac{1}{1000}$

RF for R.bid = B.bid : $\frac{1}{100}$

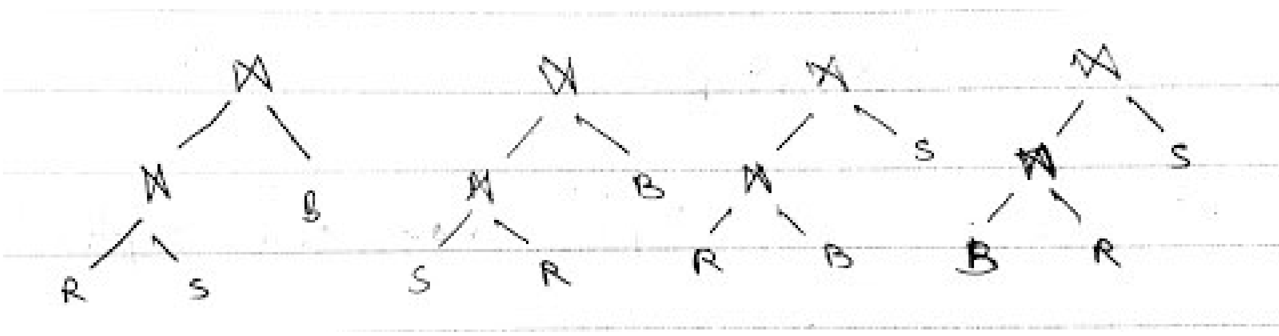
$\therefore$ Number of tuples returned by the query $= 10^9 \times \frac{1}{1000} \times \frac{1}{1000} \times \frac{1}{100}$

$= \underline{10}$

Consider Query 2:

```
SELECT      S.sid, S.sname, B.bname
FROM        Sailors S, Reserves R, Boats B
WHERE       S.sid = R.sid AND R.bid = B.bid
```

(b) [4 points] Draw all possible left-deep query plans for this query:



(c) [8 points] List all the one-relation, two-relation, and three-relation subplans that a System-R optimizer will consider in optimizing Query 2 based on the dynamic programming algorithm that we learned in class. There is no need to consider the particular join methods in this question.

Pass 1 : 1-relation : Find the best plans to access R, S, B individually
Pass 2 : 2-relation : Find the best plans for : join S-R, B-R, R-S, R-B
and store the best for (R-S), (R-B)
Pass 3 : 3-relation : Find the best plans for : join (R-S)-B, (R-B)-S
and store the best for (R-S-B)

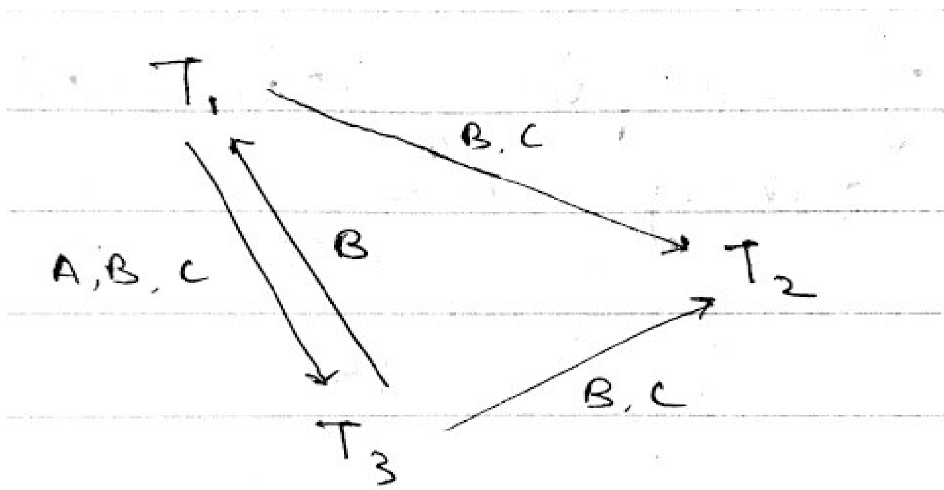
Question 6 [24 points] Transaction Management

(a) [6 points] Serializable Schedules

Consider the following schedule involving three transactions T1, T2 and T3:

	1	2	3	4	5	6	7	8	9	10
T1:	W(A)	R(B)	R(C)		W(B)					
T2:							R(C)	W(B)	W(C)	
T3:				W(B)		W(C)				R(A)

(i) Draw the precedence graph for this schedule.



(ii) Judge if this schedule is conflict serializable or not. If it is conflict serializable, give the equivalent serial schedule (just write the order of the transactions). Otherwise, briefly explain why the schedule is not conflict serializable.

The precedence graph is cyclic (with cycle T1 -> T3 -> T1), so the schedule is not conflict serializable.

(b) [6 points] Concurrency Control Protocol

Consider three transactions, T1, T2, and T3, with timestamps 1, 2, 3, respectively. Now consider the following sequences of actions, listed in the order that they are produced from user queries:

T1:	**R(A)**				**W(B)**	**(Commit)**
T2:		**W(A)**	**W(B)**			**(Commit)**
T3:				**W(B)**		**(Commit)**

Explain how this sequence of actions will be executed using the **Strict Two Phase Locking** protocol (2PL). The following table is designed for you to specify in time order the activities that take place in the lock manager and in access to the database. The activities include:

- S(O) or X(O): *the lock requested for access to a database object, O*;
- R(O) or W(O): *the read or write access to a database object, O*;
- Ti blocked: *the action to block a transaction under this protocol*;
- Ti resumed: *the action to resume a blocked transaction with its lock request granted*;
- Ti commits and releases locks: *the moment a transaction commits and releases all locks*.

The first few actions are already given below. Please complete the other actions for this schedule. List only one activity in each row, and use as many rows as needed for this schedule.

Time stamp	**Action:** please list one action at a time
1	T1: S(A)
2	T1: R(A)
3	T2: X(A)
4	T2: blocked
5	T3 : X(B)
6	T3 : W(B)
7	T1 : X(B)
8	T1 : blocked
9	T3 : commits and releases block
10	T1 : resumed
11	T1 : W(B)
12	T1 : commits and releases block
13	T2 : resumed
14	T2 : W(A)
15	T2 : X(B)
16	T2 : W(B)
17	T2 : commits and releases block
18	
19	
20	

(c) [12 points] The ARIES Recovery Algorithm
Consider the execution shown in the following figure.

LSN	Log Record
00	Update: T1 writes P2
10	Update: T1 writes P1
20	Update: T2 writes P5
30	Update: T3 writes P3
40	T3 commits
50	Update: T2 writes P5
60	Update: T2 writes P5
70	T2 aborts

(i) [3 points] Describe the actions taken to rollback transaction T2.

80 | CLR : Undo T2 LSN 60, undonextLSN=50
90 | CLR : Undo T2 LSN 50, undonextLSN=20
100 | CLR : Undo T2 LSN 20, undonextLSN=NULL
110 | T2 ends

(ii) [9 points] Assume that the system crashed right after generating the log record 70. Now the system wakes up and runs the ARIES recovery algorithm. Please show all new log entries generated in the recovery process, where each log record includes the LSN, type of log record, transaction ID, page ID, prevLSN, and undonextLSN.

LSN	Type	Xact ID	Page ID	prevLSN	undonextLSN
80	CLR	T2	P5	70	50
90	CLR	T2	P5	80	20
100	CLR	T2	P5	90	NULL
110	END	T2		100	
120	ABORT	T1		10	
130	CLR	T1	P1	120	00
140	CLR	T1	P2	130	NULL
	END	T1		140	