Boston University
CS460/660
Introduction to Database Systems
Final Exam – Fall 2016

Name: _____

Instructions

1. This exam is closed book and notes. Only a 1-page crib-sheet is allowed.
2. You have **120 minutes** to complete it. There are 5 problems.
3. Return everything (crib-sheet, question-sheet and answer booklet.)

Problem 1 (20)

Consider the following database that stores information professors, departments, and committees:

Professor(pid, profname, did, building, officenumber)

Department(did, deptname, building)

Committee(commname, pid, meetingtime)

Write the SQL statements for the following queries:

1. Find all the professor names that are in any one of the committees that professor Snyder is in.

2. For each building, find the number of professor offices in this building.

3. Find all the professors who are in at least all those committees that professor Snyder is in.

You can use sub-queries and/or intermediate relations.

Problem 2 (20)

In the following schedules, Ri(A) stands for a Read(A) operation by transaction i; Wi(A) stands for a Write(A) operation by transaction i.

1. For the following schedule show if it is conflict-serializable and if it is, give a conflict-equivalent serial schedule. Show all the conflict operations.

    1. R1(A) W1(A) R2(A) R2(B) W3(B) W2(C) R4(A) R4(B) R4(C) R2(D) R3(E)

2. Consider the following schedule example where an item is missing:

    W3(B) R2(A) W1(A) R3(?) R2(B) W1(B)

and let:

I. **Producible using 2 Phase Locking**
II. **Conflict Serializable**

Choose one answer for the following cases:

a) If **? =** A, this schedule is which of the following:
a. I & II b. I only c. II only d. neither I nor II

b) If **?** = B, this schedule is which of the following:
a. I & II b. I only c. II only d. neither I nor II

c) If **?** = C, this schedule is which of the following:
a. I & II b. I only c. II only d. neither I nor II

Problem 3 (20 points)

Answer the following questions about Extendible Hash indexes. Assume that a bucket of the index can store up to 4 keys.

(1) Insert the following keys into an initially empty Extendible Hashing index:

3; 5; 7; 9; 10; 15; 25; 31; 44; 64

(2) Show the index after inserting a single key whose insertion causes a bucket split into the index that you created in (1).

(3) What is the maximum number of data entries that can be inserted into the index that you created in (1) before you have to split a bucket? Explain briefly.

(4) What is the minimum number of record insertions into the index that you created in (1) that will cause a split of all buckets? Explain briefly.

Problem 4 (20)

Consider the following schema:

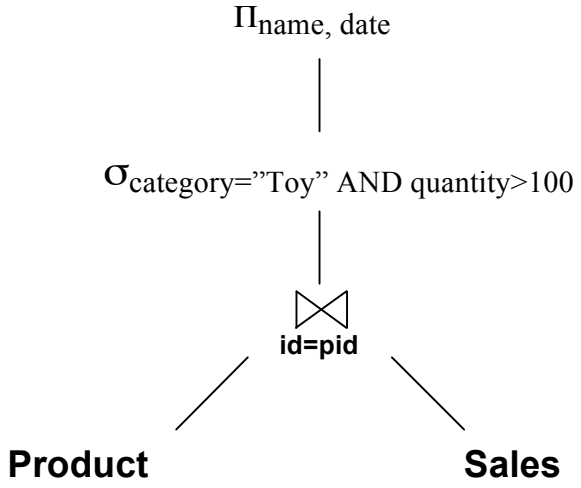
Product (id(4), name(16), manufacturer(20), category(10), color(10), webpage(40))
Sales (pid(4), quantity(4), shippingaddress(20), date(12), shippingmethod(10))

Each attribute has a fixed length, with the size (in bytes) indicated by the number in parentheses. The number of tuples in each relation is: T(Product) = 1,000,000 and T(Sales)=2,000,000.

The size of one disk block is 1000 bytes, and there are 101 buffer blocks available in main memory.

1. Compute the number of blocks taken by each table B(Product) and B(Sales).

2. Consider the logical plan below:



A) . Assume the following execution plan: the join is implemented as a **hash join**, its result is pipelined into the selection operator, and, from there, pipelined into the projection operators (i.e. on-the-fly). Compute the total cost of this evaluation plan.

B). Derive a new logical plan by pushing selections and projections down as far as possible (you have to draw a plan).

C). Consider an evaluation plan for your new logical plan in which all selections and projections are pipelined and the join is a **block-nested loop join**. Further assume that 1% of all Products are in category "Toy" and that 20% of all Sales have a quantity over 100. Compute the cost of your plan.

Problem 5 (20 points)

Consider a database system that uses the ARIES recovery protocol (Algorithm for Recovery and Isolation Exploiting Semantics.)
Next, we show the tables that were stored at the checkpoint and the state of the database on the disk at the time of the crash.

Database: (PageNo: pageLSN)		
A: 30	B: 3	C: 7
D: 10	E: 8	

Database state on disk

Transaction Table		
Xact ID	status	lastLSN
2	running	30

Dirty Page Table	
page	recLSN
A	30

Tables stored at the checkpoint.

Master Log Record:
LSN of last checkpoint = 35

Log

LSN	TransID	Page	PrevLSN	Type	UndoNextLSN
10	1	D	Null	Update	NA
20	1	NA	10	Commit	NA
25	1	NA	20	End	NA
30	2	A	Null	Update	NA
35	NA	NA	NA	Begin checkpoint	NA
37	See tables above			End checkpoint	NA
40	2	B	30	Update	NA
50	3	D	Null	Update	NA
60	4	C	Null	Update	NA
70	3	E	50	Update	NA
80	3	NA	70	Commit	NA
90					
100					
110					
120					
130					
140					

Your job will be to perform recovery according to ARIES. Follow the procedures next:

a. Analysis. Do analysis to correctly update the Transaction and Dirty Page Tables. Write these tables on your booklet. Which transactions need to be undone?
b. Redo. Perform redo to repeat history. Write the LSNs of the log records that you have to redo in the database during this phase. State which redo operations are not needed and why. Add any additional log record if needed.
c. Undo. Perform undo. State which operations will be undone and on which pages. Also, add the appropriate log entries while undo is being performed.