CSE 562 Final - Spring 2017

May 11, 2017
Start Time: 12:40 PM
End Time: 1:50 PM

Name:

UBIT:

Person #:

Maximum number of points possible: 100. Questions vary in difficulty. Do not get stuck on one question, and make sure to write down any assumptions you make.

In addition to the cover page, reference materials, and scratch pages in the back, there are 11 pages in this exam. Page 2 includes some reference materials that may or may not be useful.

Be sure to write your UBIT at least once on every page.

Question	Points Possible	Points Earned
A	15	
B.1	5	
B.2	10	
B.3	5	
B.4	10	
B.5	10	
B	40	
C.1	15	
C.2	8	
C.3	2	
C	25	
D.1	10	
D.2	5	
D.3	5	
D	20	
Total	100	

Extended Relational Algebra Operator Reference

Select	$\sigma_c(R)$	c : The selection condition
Extended Project	$\pi_{e_1,e_2,\ldots}(R)$	e_i : The column or expression to project
Product	$R_1 \times R_2$	
Join	$R_1 \bowtie_c R_2$	c : the join condition
Distinct	$\delta(R)$	
Group	$\gamma_{gb_1,gb_2,\ldots,\texttt{AGG}(e_1),\ldots}(R)$	gb_i : group by columns, e_i : expression
Set Difference	$R_1 - R_2$	
Union	$R_1 \cup R_2$	
Sort	τ_A	A one or more attributes to sort on

Relational Algebra Equivalences

Rule	**Notes**
$\sigma_{C_1 \wedge C_2}(R) \equiv \sigma_{C_1}(\sigma_{C_2}(R))$	
$\sigma_{C_1 \vee C_2}(R) \equiv \sigma_{C_1}(R) \cup \sigma_{C_2}(R)$	Note, this is only true for set, not bag union
$\sigma_C(R \times S) \equiv R \bowtie_C S$	
$\sigma_C(R \times S) \equiv \sigma_C(R) \times S$	If C references only R's attributes, also works for joins
$\pi_A(\pi_{A \cup B}(R)) \equiv \pi_A(R)$	
$\sigma_C(\pi_A(R)) \equiv \pi_A(\sigma_C(R))$	If A contains all of the attributes referenced by C
$\pi_{A \cup B}(R \times S) \equiv \pi_A(R) \times \pi_B(S)$	Where A (resp., B) contains attributes in R (resp., S)
$R \times (S \times T) \equiv (R \times S) \times T$	Also works for joins
$R \times S \equiv S \times R$	Also works for joins
$R \cup (S \cup T) \equiv (R \cup S) \cup T$	Also works for intersection and bag-union
$R \cup S \equiv S \cup R$	Also works for intersections and bag-union
$\sigma_C(R \cup S) \equiv \sigma_C(R) \cup \sigma_C(S)$	Also works for intersections and bag-union
$\pi_A(R \cup S) \equiv \pi_A(R) \cup \pi_A(S)$	Also works for intersections and bag-union
$\sigma_C(\gamma_{A,AGG}(R)) \equiv \gamma_{A,AGG}(\sigma_C(R))$	If A contains **all** of the attributes referenced by C

Part A: Transactions
(15 points)

This question pertains to the schedules below, with entries in temporal order the form TransactionID: Operation. For 2-phase locking assume that locks are released on commit/abort. For OCC, assume that timestamps are assigned in the order in which the transactions commit/abort. For both timestamp concurrency control protocols, use the transaction ID as its timestamp (i.e., T1 has a timestamp of 1). *Note that the questions ask about the **specific** schedules presented, and not the possibility of an equivalent schedule meeting the desired constraints.*

- **Schedule 1:** T1: W(A), T1: COMMIT, T2: R(A), T3: W(B), T3: W(C), T3: ABORT, T2: W(A), T2: W(B), T2: COMMIT
- **Schedule 2:** T1: R(A), T1: COMMIT, T2: R(A), T3: R(B), T2: W(B), T2: W(A), T2: COMMIT, T3: R(C), T3: ABORT
- **Schedule 3:** T2: R(A), T1: W(A), T1: COMMIT, T3: R(B), T2: R(B), T2: W(A), T3: W(A), T3: COMMIT, T2: COMMIT
- **Schedule 4:** T1: W(B), T2: R(A), T2: W(B), T1: W(A), T3: R(C), T1: COMMIT, T2: W(A), T2: COMMIT, T3: W(B), T3: W(A) T3: COMMIT
- **Schedule 5:** T1: R(A), T3: R(B), T3: W(A), T2: R(A), T2: W(A), T2: COMMIT, T1: R(B), T1: W(A), T3: W(A), T3: COMMIT, T1: COMMIT

Mark Y/N in each grid square below

The schedule ...	**S1**	**S2**	**S3**	**S4**	**S5**
... is conflict serializable					
... is view serializable					
... could be realized through 2-phase locking					
... could be realized through optimistic concurrency control					
... could be realized through timestamp concurrency control					
... could be realized through multi-version timestamp concurrency control					

Part B: Logging & ARIES
(40 points)

This question pertains to the following log, disk-state, and most-recent checkpoint. Log entries are written in the form: $Object(OriginalValue \rightarrow NewValue)$.

Undo/Redo Log

Seq #	Xact	Prev #	Operation
#20	T1	n/a	$A(5 \rightarrow 10)$
#21	T2	n/a	$B(3 \rightarrow 7)$
#22	T1	#20	$C(1 \rightarrow 2)$
#23	n/a	n/a	`begin_checkpoint`
#24	T3	n/a	$D(5 \rightarrow 3)$
#25	T2	#21	`COMMIT`
#26	T5	n/a	$B(7 \rightarrow 6)$
#27	n/a	n/a	`end_checkpoint`
#28	T5	#26	$C(2 \rightarrow 3)$
#29	T5	#28	`ABORT`
#30	T5	#26	`CLR` $C \rightarrow 2$
#31	T3	#24	$A(10 \rightarrow 1)$
#32	T1	#22	$D(3 \rightarrow 5)$

Transaction Table Checkpoint

Xact	Last Seq #	State
T1	#22	`ACTIVE`
T2	#21	`ACTIVE`

Disk State

Object	Last Seq #	Value
A	#12	5
B	#21	7
C	#22	2
D	#4	5

1. **(5pt)** What is the first log entry read by the **ANALYZE** phase of ARIES?

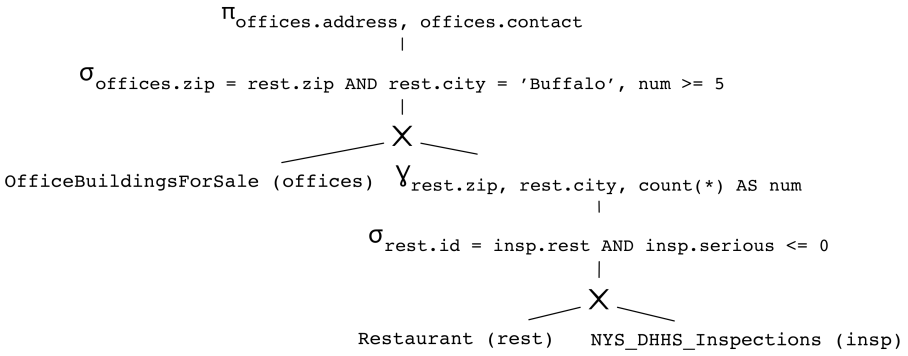
2. **(10pt)** What is the state of the transaction table after the **ANALYZE** phase ends?

3. **(5pt)** What is the first log entry read by the **REDO** phase of ARIES?

4. **(10pt)** What is the state of the buffer manager (i.e., the in-memory copy of the disk state) and the transaction table after the **REDO** phase ends? You may assume that nothing gets paged out, and that log entries before #20 are not relevant.

5. **(10pt)** List *and number* the full set of compensation log records (CLRs) created during the **UNDO** phase of ARIES.

Part C: Relational Algebra
(25 points)



1. **(15pt)** Construct the most optimal form of the query above, using only the rules on page 2. Show your work (including which rules and their role on the equivalence).

2. **(8pt)** Using only the rules on 2, derive the delta of this query with respect to an insertion into `offices` (δ_+`offices`).

3. **(2pt)** How would you approach computing the delta with respect to an insertion into `insp`?

Part D: Probabilistic Databases
(20 points)

Questions in this section pertain to the following set of possible worlds:

R_1	A	B
	1	2
	3	4

R_2	A	B
	1	4
	3	4

R_3	A	B
	1	4
	3	2

1. **(10 pt)** Construct a C-Table for the possible worlds.

2. **(5 pt)** What is the set of possible answer tuples (that appear in at least one possible result) to the query $\pi_A \sigma_{B \geq 3} \mathcal{R}$?

3. **(5 pt)** What is the set of certain answer tuples (that appear in all possible results) to the query $\pi_A \sigma_{B \geq 3} \mathcal{R}$?

[[**Scratch Page**]]

[[**Scratch Page**]]