End-Semester Examination (Autumn'2018) IT 214 Database Management Systems	A

Student ID: _____ Name: _____

Time: 90 minutes Max Points: 115

IMPORTANT NOTE:
1. There are six pages make sure that you have them all in your set.
2. You need to answer all questions in question paper itself.
3. Write answers neat and clean. Over-writing is not allowed.

Marks

1	2	3	4	5	6	7	8	9	

1. Tick the correct option (there can be multiple options be correct; check them all) – [10]

   i. Heap files are roughly good for operations <INSERT/DELETE/SEACRH>

   ii. Secondary index can be <dense/clustered/b+-tree/~~sparse~~>

   iii. Participation of a weak entity with identifying relationship is always total [True/False]

   iv. Materialized view hold data <true/false> _____

   If yes, what data? __ result of query associated with view _____

   v. Unbound cursor can be bound with dynamic query <Yes/No>

   vi. TP_OP is a <implicit local variable/implicit parameter variable/~~global variable/ no such variable~~>

   vii. Embedded SQL provides an environment in which SQL statements are sent to DBMS for execution <Yes/No/both-depends on configuration>

   viii. Using "CREATE ASSERTION", we can create constraints that span to multiple tables <True/False/No such command>

   ix. In case of B+-tree based index, index scan mean "sequential scan of leaf nodes" <True/False/No such operation>

   x. Bloated index refers to index having <too much void spaces in blocks/ too much overflow/index corrupted /all of these>

2. Fill in the blanks-

[30]

i. Referential Integrity constraint requires that
__FK referring to existing tuple in referenced relation__

ii. Prepared statement helps in
__ Avoiding SQL Injection | faster execution | avoiding run-time errors __

iii. Entity Integrity constraint requires that __ PK cannot be NULL __

iv. **`CallableStatement`** object in JDBC is used for __ Calling Stored Procedure __

v. Which property of schedule ensures **Isolation** __ Serializability __

vi. Which property of schedule ensures **Atomicity** ___ Recoverability __

vii. Which property of schedule ensures **Durability** __ Recoverability ___

viii. Weak entity is the one that ___ Does not have its own key _____
_____Requires owner entity to be identified _____

ix. Which SQL Isolation level may cause Dirty Read ___Read Uncommited ____

x. Advantage of strict 2PL over standard 2PL ___Reduced Cascaded Rollback/aborts __

xi. System logs helps in __ Database Recovery __

xii. Write Ahead Logging Protocol is used for __ Database Recovery __

xiii. Noted problem in basic Snapshot Isolation is __ Skew Write __

xiv. Main problem in 2PL protocol is __ Deadlocks __

xv. Main problem with basic time stamp ordering techniques is
_Unnecessary aborts __

xvi. What is the name of technique that is used for avoiding cascaded rollbacks
__ (1) Do not read from uncommitted transaction__
__ (2) Rigorous 2PL __

xvii. What is highest SQL isolation level that has Phantom row problem
__ Repeatable Read _

xviii. One of the most important advantage of sql views is
__ (1) Abstraction __ (2) Hide Complexities __
__ (3) Reusability ___ (4) Data Hiding __
__ (5) logical data independence __

xix. Name one of the procedural data manipulation language
__ PL/PgSQL ____ PL/SQL _____

xx. Can you determine Normal form of relation R(AB), in absence of FD information. Note that, no information does not mean NO FD; if yes what is NF? __ BCNF____

3. Consider following keys; Primary Key(PK), Key(K), Candidate Key(CK), Super Key (SK), and choose which symbols is most appropriate to be placed in the blank space (⊆ or ⊇ or =) below. [5]

   i. PK ___=____ K

   ii. PK ___=____ CK

   iii. CK ___=___ K

   iv. CK ___⊆___ SK

   v. K ___⊆___ SK

4. Give short answers -

   i. Suppose Employee relation has B+-tree index on DNO. Compute approximately execution cost of query "SELECT * FROM EMPLOYEE WHERE DNO=5"? May express in terms number of block read/writes, and assume other parameters! [3]

H+ ⌈S/BF⌉ or H+1
Where H is index height, S is selectivity (no of tuples / distinct values of DNO), BF is blocking factor number of records in a block

   ii. Suppose you have following query to be executed
   $\sigma_{SALARY>30000 \wedge DNO=5}(EMPLOYEE)$
   What could be best strategy to execute this query. [3]

Strategy#1: if there is index on DNO, use this index and then sequential scan and apply salary selection criteria

Strategy#2: if there is index on DNO and SALARY both, select records on individual index and compute intersection

Strategy#3: if there is no index on any of attribute (i.e. DNO and SALARY), sequential scan is only option

   iii. Compute Join selectivity of join EMPLOYEE JOIN DEPARTMENT ON DNO. What is "Join Selectivity" of the JOIN? May assume other parameters [3]

(1) What is JS = |R JOIN S| / (|R| * |S|)
(2) EMPLOYEE JOIN DEPARTMENT ON DNO = |EMP| / (||EMP| * |DEP|)
= 1 / |DEP|

   iv. Translate following SQL query in terms of relational algebra: [3]

```
SELECT * FROM EMPLOYEE WHERE SSN IN
                (SELECT DISTINCT ESSN FROM DEPENDENT)
```

Solution-1

$$EMPLOYEE \underset{e.ssn = d.essn}{\overset{SEMI\ JOIN}{}} DEPENDENT$$

Solution-2

$$EMPLOYEE * (\pi_{SSN}(EMPLOYEE) \cap \pi_{ESSN}(DEPENDENT))$$

v. Consider relation R(A, B, C) and set of FDs { {A → B}, {B → C}}; can you find out a join dependency here. [3]

Yes: *{(AB), (BC)}

vi. Given, A → B and XB → C, prove that XA → C. [3]

A → B (FD1-given)
XB → C (FD2-given)
XA → XB (FD3- using Augmentation rule to FD1)
XA → C (using transitive rule on FD2 & FD3)

5. Suppose following schedule is executed by interleaving operations from transactions T1 and T2. What kind of concurrency problem do you see in following schedule? Give short reason. [5]

```
T1: Read X
T1: X = X + 50
T1: Write X
T2: Read X
T2: X = X + 100
T2: Write X
T1: Abort
T2: Commit
```

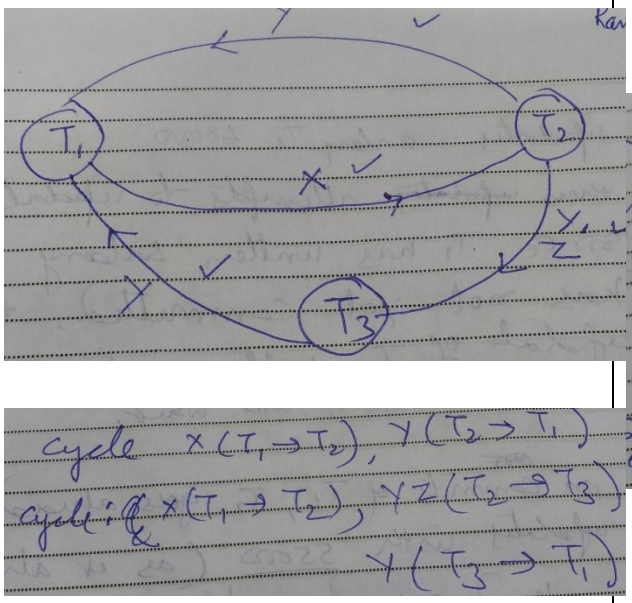
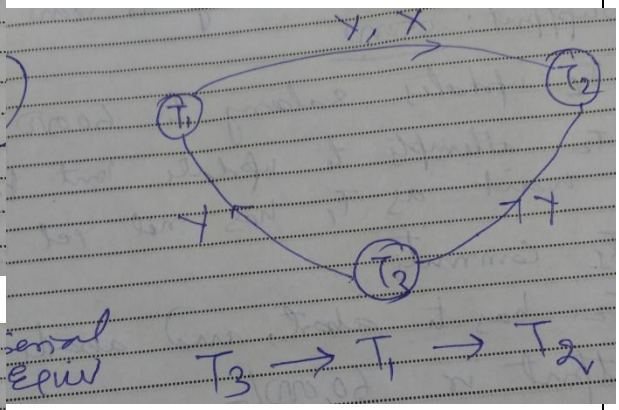
DIRTY READ: T2 reads write of T1, and T1 aborts.

6. Is the schedule given in previous question is recoverable? Give short reason. [5]

No, not recoverable; because T2 commits while T1 aborts.

7. Consider the three transactions T1, T2, and T3, and the schedules S1 and S2 given below. Draw the serializability (precedence) graphs for S1 and S2, and state whether each schedule is serializable or not. If a schedule is serializable, write down the equivalent serial schedule(s).

[10]

S1: r2(Z); r2(Y); w2(Y); r3(Y); r3(Z); r1(X); w1(X); w3(Y); w3(Z); r2(X); r1(Y); w1(Y); w2(X);	S2: r3(Y); r3(Z); r1(X); w1(X); w3(Y); w3(Z); r2(Z); r1(Y); w1(Y); r2(Y); w2(Y); r2(X); w2(X);
	

8. Consider following schedule executed on PostgreSQL. Assume that initial salary for employee with SSN 123 is 50000. [Here labels T1 and T2 against statements indicate Transaction that is issuing the statement.] What will be shown by SELECT statements at line numbers 3, 5, 7, 10 if transaction T2 is specified to execute at READ COMMITTED isolation level and SERIALIZABLE isolation level?

[6+6]

```
1  T1: begin;
2  T1: update employee set salary = salary+3000 where ssn = 123;
3  T1: select salary from employee where ssn = 123;
4  T2: begin;
5  T2: select salary from employee where ssn = 123;
6  T2: update employee set salary=salary+5000 where ssn = 123;
7  T2: select salary from employee where ssn = 123;
8  T1: commit;
9  T2: commit;
10 T1: select salary from employee where ssn = 123;
```

READ COMMITTED level	SERIALIZABLE level
Line 3: ___ 53000 _____	Line 3: ___ 53000 ___
Line 5: ___ 50000 _____	Line 5: ___ 50000 ___
Line 7: ___ 58000 ____	Line 7: _ ERROR _(53000_is also OK)
Line 10: ___ 58000 ___	Line 10: ___ 53000 __

Student ID: _____ Name: _____

9. Suppose following attributes are drawn from a sales/purchase system of trading company.

Sales_bill_no, sales_bill_date, customer_no, customer_name, item_no, item_name, quantity_in_stock, quantity_in_bill, item_rate, item_rate_in_bill, item_category, item_category_sales_tax_rate, supplier_id, supplier_name, purchase_bill_no, purchase_bill_date, quantity_in_purchase_bill, item_rate_in_purchase_bill, average_purchase_price

Assume that: (1) there is only one supplier for each purchase bill,
(2) an item comes only once in a bill
(3) there is only one customer for a sales bill (4) sales tax rate in a bill depends on item category
(5) a sales bill contains all items of same category.

Your tasks here are following - [10+10]

i. Identify Minimal FD Set
ii. Give normalized relations that are in BCNF. Specify Keys and FKs also.

FDs

sales_bill_no → {sales_bill_date, customer_no}

customer_no → customer_name

item_no → {item_name, qty_in_stock, item_rate, item_category, average_purc_price}

{sales_bill_no, item_no} → {quantity_in_bill, item_rate_in_bill}

item_category → item_category_sales_tax_rate,

supplier_id → supplier_name

purch_bill_no → {purch_bill_date, supplier_id}

{purch_bill_no, item_no}→{qty_in_purch_bill, item_rate_in_purch_bill}

Relations

sales_bill(sales bill no, sales_bill_date, customer_no)

customer(customer no, customer_name)

item(item no, item_name, qty_in_stock, item_rate, item_category, average_purc_price)

billitems(sales bill no, item no, quantity_in_bill, item_rate_in_bill)

category(item category, item_category_sales_tax_rate)

supplier(supplier id, supplier_name)

purchase(purch bill no, purch_bill_date, supplier_id)

purchasebill(purch bill no, item, qty_in_purch_bill, item_rate_in_purch_bill)