Fundamentals of Database Systems

Assignment: 8

Due Date: 19th September, 2017

Instructions

This question paper contains 10 questions in 3 pages.

Q1: Consider the following schedule. All the locks are exclusive, and between the *lock l* and *unlock u* operations, the corresponding data item is first read and then written.

S: $l_1(A), l_3(D), l_1(B), u_1(A), l_2(C), l_2(B), l_1(D), l_2(B), l_3(C), l_4(A), l_4(C), l_5(A)$

The schedule will result in a deadlock.

A. True

B. False

Explanation: The *wait-for* graph of the schedule contains a cycle, hence it will cause deadlock.

Q2: Consider the following schedule S.

S: $l_1(A), l_3(D), l_1(B), u_1(A), l_2(C), l_2(B), l_1(D), l_2(B), l_3(C), l_4(A), l_4(C), l_5(A)$

Which of the following is a valid set of transactions that are potential *victims*?

A. $\{T_1, T_2, T_3, T_5\}$

B. $\{T_1, T_2\}$

C. $\{T_1, T_2, T_3, T_4\}$

D. $\{T_1, T_2, T_3\}$

Explanation: Transactions $\{T_1, T_2, T_3\}$ are involved in cycle of *wait-for* graph of the schedule.

Q3: Suppose a deadlock occurs in the schedule S given below.

S: $l_1(A), l_3(D), l_1(B), u_1(A), l_2(C), l_2(B), l_1(D), l_2(B), l_3(C), l_4(A), l_4(C), l_5(A)$

A transaction that causes the *least* number of cascading rollbacks is decided to be chosen as *victim*, then which of the following transaction can *not* be chosen as a victim?

A. T_3

B. T_2

C. T_1

D. Cannot be decided

Explanation: T_1 will cause cascading rollback, if aborted as T_4 is reading item A which was written by T_1.

Q4: Consider the following database

Col1	Col2	Col3
1	2	3
4	5	6
7	8	9

If Col1 and Col2 form a column family and Col3 is another column family, how is the data stored?

A. 1 2 3 4 5 6 7 8 9

B. 1 4 7 2 5 8 3 6 9

C. 1 2 4 5 7 8 3 6 9

D. 1 4 7 3 6 9 2 5 8

Explanation: Each columnar family is stored in a row major way, one after the other

Q5: Match the systems in column A with their best examples in column B.

A	B
a. Key Value Store	1. HBase
b. Big Table System	2. Titan
c. Document Database	3. Redis
d. Graph Database	4. MongoDB

A. a-3 b-1 c-4 d-2

B. a-2 b-1 c-3 d-4

C. a-4 b-3 c-1 d-2

D. a-4 b-2 c-3 d-1

Explanation: From the slides

Q6: Which of the following types of data can be associated with big data systems?
(1) Structured
(2) Semi-structured
(3) Un-structured

A. Only (2)

B. Only (3)

C. Only (2) and (3)

D. All of (1), (2) and (3)

Explanation: Big data can contain all types of data.

Q7: Which of the following can be described as a programming model used to develop applications processing massive amounts of data in a distributed and/or parallel manner?

A. Map Reduce

B. OLAP

C. Mahout

D. Cloud Computing

Explanation: Programming model is map-reduce.

Q8: Which statement is *true* about big data?

A. It must contain complex data.

B. Any data more than 1TB or so is big data.

C. It depends critically on the application.

D. Any data that does not fit into the main memory of a standard machine is big data.

Explanation: Big data depends critically on the application rather than the size of the data.

Q9: Which of the following statement(s) is/are *not* correct?
(1) 3 V's of big data are *volume*, *variety* and *velocity*.
(2) Big data can be structured as relational data.
(3) Hadoop is based on map-reduce framework.
(4) Data of size greater than a particular threshold is always treated as big data.

A. Only (2) and (4)

B. Only (3)

C. None of (1), (2), (3) and (4)

D. Only (4)

Explanation: Incorrect interpretation of big data.

Q10: Which of the following statement(s) is/are *not* correct?
(1) Eventual consistency may result in stale reads.
(2) NoSQL can handle schema less data.
(3) Columnar storage handles update operation as efficiently as RDBMS.
(4) In key value store, keys could be heterogeneous.

A. Only (3) and (4)

B. Only (1)

C. All of (1), (2), (3) and (4)

D. Only (3)

Explanation: Columnar storage stores tuple in *column-wise* manner and hence update requires multiple access to retrieve a tuple.