CS4224/CS5424 Quiz 6: Data Replication

Wk 7, Sem 1, 2023/24

1. Consider the execution of the following set of transactions

$$\begin{array}{ll} T_1: & W_1(x) \\ T_2: & R_2(x),\ R_2(y) \\ T_3: & W_3(y) \\ T_4: & R_4(x),\ R_4(y) \end{array}$$

on a distributed database with data replicated over two sites as follows:

$$\text{Site A} = \{x, y\},\ \ \text{Site B} = \{x, y\}$$

Suppose that the local schedules for the execution are as follows:

$$\begin{array}{ll} S_A: & W_1(x_A),\ W_3(y_A),\ R_2(x_A) \\ S_B: & R_2(y_B),\ W_3(y_B),\ W_1(x_B),\ R_4(x_B),\ R_4(y_B) \end{array}$$

Is the above schedule one-copy serializable?

Solution:

- Since T2 read x from T1, T1 must precede T2 in an equivalent serial 1C schedule.
- Since T4 read x from T1, T1 must precede T4 in an equivalent serial 1C schedule.
- Since T4 read y from T3, T3 must precede T4 in an equivalent serial 1C schedule.
- Since T2 read y from the initial database state, T2 must precede T3 in an equivalent serial 1C schedule.
- Hence, the RD schedule is equivalent to the serial 1C schedule (T_1, T_2, T_3, T_4) and it is 1SR.

2. Consider the execution of the following set of transactions

$$\begin{array}{ll} T_1: & W_1(x) \\ T_2: & R_2(x),\ R_2(y) \\ T_3: & W_3(y) \\ T_4: & R_4(x),\ R_4(y),\ W_4(x) \end{array}$$

on a distributed database with data replicated over two sites as follows:

$$\text{Site A} = \{x, y\},\ \ \text{Site B} = \{x, y\}$$

Suppose that the local schedules for the execution are as follows:

$$\begin{array}{ll} S_A: & W_1(x_A),\ W_3(y_A),\ R_2(x_A),\ W_4(x_A) \\ S_B: & R_2(y_B),\ R_4(x_B),\ W_1(x_B),\ R_4(y_B),\ W_4(x_B),\ W_3(y_B) \end{array}$$

Is the above schedule one-copy serializable?

Solution:

- Since T4 read x from the initial database state, T4 must precede T1 in an equivalent serial 1C schedule.
- However, if T_4 were to precede T_1 in a serial 1C schedule, the final write on x in the serial 1C schedule would be performed by T_1, but this is inconsistent with the RD schedule where the final write on both x_A and x_B was performed by T_4.
- Hence, the RD schedule is not 1SR.

3. Consider the execution of the following two transactions

$$\begin{array}{ll} T_1: & R_1(x),\ R_1(y),\ W_1(x) \\ T_2: & R_2(x),\ R_2(y),\ W_2(y) \end{array}$$

on a replicated database distributed across 4 sites:

$$\text{Site A} = \{x\},\ \text{Site B} = \{x, y\},\ \text{Site C} = \{x, y\},\ \text{Site D} = \{y\}$$

The local schedules are as follows:

$$\begin{array}{ll} S_A: & R_1(x_A), W_1(x_A), C_1 \\ S_B: & R_1(y_B), W_1^r(x_B), C_1^r, W_2^r(y_B), C_2^r \\ S_C: & R_2(x_C), W_2^r(y_C), C_2^r, W_1^r(x_C), C_1^r \\ S_D: & R_2(y_D), W_2(y_D), C_2 \end{array}$$

(a) Which of the replication protocols (eager centralized, eager distributed, lazy centralized, lazy distributed) could have been used for the above execution?

(b) Is the schedule one-copy serializable?

Solution:

(a) Since the schedules include refresh transactions, the RD schedule is produced by a lazy protocol. Since reads are always performed on local copies whenever available, T_1 must be issued at S_A (due to $R_1(x_A)$ and $R_1(y_B)$) and T_2 must be issued at S_D (due to $R_2(y_D)$ and $R_2(x_C)$). The protocol could be the lazy centralized protocol with S_A as the master site for x and S_D as the master site for y. The protocol could also be the lazy distributed protocol.

(b) No. In a 1SR schedule, either T_1 read y from T_2 or T_2 read x from T_1. Since neither T_1 nor T_2 is reading from each other in the RD schedule, the schedule is not 1SR.

4. Consider the execution of the following two transactions

$$\begin{array}{ll} T_1: & R_1(x),\ R_1(y),\ W_1(x) \\ T_2: & R_2(x),\ R_2(y),\ W_2(y) \end{array}$$

on a replicated database distributed across 4 sites:

$$\text{Site A} = \{x\},\ \text{Site B} = \{x, y\},\ \text{Site C} = \{x, y\},\ \text{Site D} = \{y\}$$

The local schedules are as follows:

$$\begin{array}{ll} S_A: & R_1(x_A), W_1^r(x_A), C_1^r \\ S_B: & R_1(y_B), W_2^r(y_B), C_2^r, W_1^r(x_B), C_1^r \\ S_C: & R_2(x_C), W_2(y_C), C_2, W_1(x_C), C_1 \\ S_D: & R_2(y_D), W_2^r(y_D), C_2^r \end{array}$$

(a) Which of the replication protocols (eager centralized, eager distributed, lazy centralized, lazy distributed) could have been used for the above execution?

(b) Is the schedule one-copy serializable?

Solution:

(a) Since the schedules include refresh transactions, the RD schedule is produced by a lazy protocol. Since reads are always performed on local copies whenever available, T_1 must be issued at S_A and T_2 must be issued at S_D. Since $W_1(x)$ is performed at S_C and not at S_A, the protocol can't be a lazy distributed protocol. Hence, the protocol is the lazy centralized protocol with S_C as the master site for both x and y.

(b) No. Similar reasoning as the previous question.

5. For each of the following applications, discuss the suitability of using each of the four replication protocols (eager centralized, eager distributed, lazy centralized, lazy distributed).

   (a) Consider a bank application with its database fully replicated across 20 branches within the same country. The database consists of the following four relations:
   - `Customers` relation stores information about customers (e.g., name, address),
   - `Savings` relation stores information about customers' savings accounts (e.g., customer id, account id, balance), and
   - `Checking` relation stores information about customers' checking accounts (e.g., customer id, account id, balance).

   (b) Consider an online shopping application with its database fully replicated across 20 sites in different countries. The database consists of the following four relations:
   - `Product` relation stores information about products (e.g., product description and price),
   - `Customer` relation stores information about customers (e.g., name, shipping address, billing information),
   - `ShopBasket` relation stores information about products that customers have selected for purchase but have not yet ordered (e.g., product id, quantity), and
   - `Order` relation stores information about customers' orders (e.g., customer id, order date, amount paid).

   (c) Consider a sales database that is fully replicated across one server and 20 laptops. Each laptop belongs to a salesperson and the laptop may not be always connected to the server (e.g., while the salesperson is travelling). The database consists of the following three relations:
   - `Customer` relation stores information about customers (e.g., name, address, contact number),
   - `Order` relation stores information about customers' orders (e.g., customer id, order date, sales amount), and
   - `CustomerVisit` relation stores information about the customer visits made by each salesperson (e.g., salesperson id, customer id, date of visit, visit report).

Solution:

(a) eager centralized/distributed.

(b) lazy centralized/distributed.

(c) lazy centralized/distributed.