Assignment 6

Transactions, Security, Data Ethics, NoSQL

Assigned: Nov 15, 2024
Due: Dec 6, 2024. 11:59:59PM
Total Points: 100

Objectives

The purpose of this assignment is to go through some exercise of basic transaction, security and NoSQL concepts. Specifically, this assignment will help you practice your knowledge on the following points:

- What is the basic concept for a database transaction.
- Basic knowledge of transaction anomalies and concurrency control.
- Identify data ethical issues and come up with some solutions in real-world events.
- How to conduct basic access control through SQL.
- When NoSQL database is more preferable than RDBMS.

Part I: Transactions [45pts]

1. [10pts] Describe in your own word what are the four properties that ACID provides for a transaction. Explain each with a concrete example.

2. [18pts] Assume that there is a table S with a non-unique integer column x. For updates and deletion below, the specified records do exist in the database initially. Assuming all actions in the following schedules are immediately visible to all transactions, what is the **highest possible ANSI'92 isolation level** the following schedule satisfies? Please briefly explain why.

(1) **[6pts]** $R_1(x = 100), \text{Update}_2(\text{SET x = 102 WHERE x = 100}), C(T_2), R_1(x < 200), C(T_1)$

(2) **[6pts]** $R_1(x = 100), \text{Ins}_2(\{x : 102\}), C(T_2), R_1(x < 200), C(T_1)$

(3) **[6pts]** $R_1(x = 100), \text{Ins}_2(\{x : 100\}), C(T_2), R_1(x < 200), C(T_1)$

3. [17pts] Suppose we have three transactions T_1, T_2 and T_3. The content of each transaction is shown as below. (Note that R and W stands for read and write, while A, B, C, D are values in the database.)

- T1: R(A), W(B), W(D), R(A)
- T2: R(B), W(B), W(A), W(C)
- T3: R(B), W(A), R(D), W(C)

Answer the following questions:

(1) **[5pts]** Consider the schedule below:

$$R_3(B), R_2(B), W_2(B), R_1(A), W_3(A), W_1(B), W_2(A), R_3(D), W_3(C), W_2(C), W_1(D), R_1(A)$$

Is it a ***serializable*** schedule? If so, why? If not, explain why not using transaction anomalies.

(2) **[6pts]** Consider the following schedule:

$$R_1(A), W_1(B), R_3(B), R_2(B), W_2(B), W_1(D), R_1(A), W_3(A), R_3(D), W_2(A), W_2(C), W_3(C)$$

Show a schedule that is ***conflict equivalent*** to it. Also, briefly explain why they are conflict equivalent.

(3) **[6pts]** Write two ***conflict serializable*** schedules of T1, T2, and T3.

Part II: Security & Data Ethics [45pts]

1. [15pts] Imagine you are the DBA working for a university and create a relation called Enrollment with the schema below:

$$Enrollment(\underline{sname}, \underline{dname}, \underline{cno}, term, grade)$$

For authorization reasons, you need to define two views: StudentNames (with `sname` as the only attribute) and StudentGradeReport with fields `sname` and `avg_grade`. The former shows all distinct student names enrolled in some course at the university, the latter lists the average grade for each student. Answer the following questions:

(1) **[5pts]** Show the view definition statements for StudentNames and StudentGradeReport.

(2) **[5pts]** What are the SQL commands (more than one) you should issue to granted user `director` the privilege to know only average department course grades for the `CSE` and `MATH` departments?

(3) **[5pts]** What is the SQL command you should issue to grant user `secretary` the privilege to read the StudentNames view and to allow other people to read the view as well.

2. [30pts] Read the following article about a data broker company called National Public Data (NPD) and a recent incident happened around it. Focus on reading and thinking about what services they are providing, what data they need on the way, and why a data breach on a company like this can be disruptive. You can find useful context in the following links:

- https://nationalpublicdata.com/
- https://en.wikipedia.org/wiki/2024_National_Public_Data_breach
- https://securityintelligence.com/news/national-public-data-breach-publishes-private-data-billions-us-citizens/
- https://therecord.media/social-security-numbers-leak-national-public-data
- https://www.usatoday.com/story/tech/2024/08/17/social-security-hack-national-public-data-confirms/74843810007/

After getting enough context, answer the following questions:

(1) **[10pts]** Identify at least two different kinds of potential data ethical issues that may happen in this company. Discuss why it can be sensitive by itself and what danger it can cause when their data is mismanaged and/or a massive data breach described above happened.

(2) **[10pts]** Describe a potential solution based on what you have learned in the class on one concrete scenario that may happen in NPD. Discuss why it can help solve part of the ethical issues.

(3) **[10pts]** Reflect on this event and link it to another event you know that causes similar ethical issues. Discuss in which area we should take extra cautious and what can we do in general to prevent dire consequences.

Part III: NoSQL [10pts]

1. [10pts] Provide a concrete scenario when a database application developer will prefer to use a NoSQL document DBMS (e.g. MongoDB, Couchbase, etc.) rather than a traditional DBMS. In addition, name at least two advantages and one disadvantage for this decision.