MANDATORY ASSIGNMENT #2

IN3020/IN4020

Due Sunday 9th May 2021

Exercises in this mandatory assignment must be solved and submitted individually.

Completion and submission of the mandatory assignment must take place in accordance with current guidelines of the Department of Informatics (Institutt for Informatikk), see:

www.uio.no/studier/admin/obligatoriske-aktiviteter/mn-ifi-oblig.html

Submission of the mandatory assignment will be taken as the confirmation that guidelines have been read and understood.

Deadline: Sunday, May 9th at 23:59 (by midnight).

The deadline is absolute, and deferral will only be granted in accordance with the mandatory assignment rules (obligreglement).

All questions must be answered to get an approval.

Exercise 1 – Concurrency Control, Serialization

1.1 Concurrency Example and Execution Plans

a. We are going to be looking at an example of possible conflicts between **two concurrent transactions** in a store reservation system.

   A customer can log in to the electrical store reservation system and choose one electrical product that is marked as available (data element A) and one retail warehouse product marked as available (data element B) from their respective databases. When the customer has decided & ordered, both entities will be marked unavailable.

   Detail two such concurrent transactions T1 and T2 by dividing each into their read/write operations. Note: both transactions must have all necessary read and write operations.

   Ignore the date, period and quantity. Assume that we are operating with two data elements only: the electrical product and the retail warehouse product, marked as either available or not available).

b. Show at least two possible interactions between the transactions with different interleaving of their operations where at least one of them demonstrates a correct and at least one demonstrates an incorrect result. Note that you should say something about what your correctness (integrity) criterion is.

c. Create three different example execution plans.

d. Explain which are serial or serializable and why.

1.2 Conflict Serializability, Locks

a. What kind of conflicts can one have in a transaction execution plan? List them up and explain in your own words (if necessary, with examples).

b. Draw the precedence graphs of all your execution plans in 1.2 (a).

c. Just by looking at the precedence graphs, how can you determine which of them are conflict serializable?

d. Explain the reasoning behind 2-phase (2PL) locking. Note that you are not being asked for the formulation of 2PL only but for the reasoning behind the mechanism. What does it help us with?

Exercise 2 – Isolation levels

2.1 Snapshot Isolation and related mechanisms

Snapshot Isolation is a very popular technique that we have studied some time on. Define and then describe in your own words what Snapshot Isolation is and what it is used for. Why is it so popular? Does it guarantee serializability?

2.2 SI Plan

What is an SI plan? What does it mean that an SI plan is strict?

2.3 Isolation levels and concurrency mechanisms in practice

This is a "look up, learn and share" exercise: What concurrency and isolation mechanisms do the most recent versions of Postgres, MySQL and Oracle support? Look them up, list them and compare.
Note: Different DBMS' may focus upon different capabilities / mechanisms.

Exercise 3 – NoSQL DBMS

3.1 NoSQL DBMS

a. Briefly explain the key motivations behind the NoSQL databases.

b. A NoSQL database is often referred to as a schemaless database. What is a schema and in what considerations can a NoSQL database be said to be schemaless?

c. List down the common types of NoSQL databases that we have discussed and briefly explain each type.

d. What type of NoSQL database is Google Bigtable and Apache Hbase? Explain in your own words how an Hbase database is designed.

e. Mention three advantages and three disadvantages of a NoSQL database compared to a traditional relational database.

3.2 CAP theorem, NoSQL graphs, DQM, ETL/ELT

a. Explain the CAP theorem. What are the implications of the theorem on a NoSQL database?

b. Is it possible for a NoSQL database to be ACID-compliant? Justify your answer.

c. What are RDF and Property Graph database management systems? Compare and contrast these with examples.

d. Explain the concept of Data Quality Management (DQM). What are the key attributes that a DQM system aims at ensuring?

e. What do you understand by the terms ETL and ELT? Suppose, you are working on a machine learning (ML) workflow to predict the housing price in Oslo. How would you apply the ELT and/or the ETL process in your workflow?

Exercise 4 – Database Security

4.1 Security and Privacy

a. What are the incentives behind ensuring database security and privacy? Explain the key concepts and components of database security.

b. List down the common threats to a database security. Elaborate each type with appropriate examples.

c. What are the typical methods of data protection?

d. Suppose you have a database that contains personally identifiable information. You need to publish the contents of the database on the internet so that university researchers can utilize the data on their research. How can you ensure personal privacy, in addition to information security, in a shareable copy of the database, so that the researchers cannot identify any individual in the database but still find the data useful for their analysis? **Note:** the purpose of this exercise is to help you distinguish between information security and information privacy, and you can draw answers from your own thought processes and from other resources in books or the Internet.

4.2 IdM, Security architecture

a. Explain the terms Role and Identity management in the context of database security.

b. Briefly explain the oracle-based security architecture.