

## Main exercises

Saturday, February 10, 2024 3:50 PM

### QUESTION 1

We probably do not have 500 GB of main memory to hold all the data. We must therefore store data in a storage device such as a disk or tape and bring relevant parts into main memory for processing as needed. Even if we have 500 GB of main memory, on computer systems with 32-bit addressing, we cannot refer directly to more than about 4 GB of data. We have to program some method of identifying all data items. We have to write special programs to answer each question a user may want to ask about the data. These programs are likely to be complex because of the large volume of data to be searched. We must protect the data from inconsistent changes made by different users accessing the data concurrently. If applications must address the details of such concurrent access, this adds greatly to their complexity. We must ensure that data is restored to a consistent state if the system crashes while changes are being made. Operating systems provide only a password mechanism for security. This is not sufficiently flexible to enforce security policies in which different users have permission to access different subsets of the data.

### QUESTION 2 and 3

Logical Data Independence is a property of a database that can be used to change the logic behind the logical level without affecting the other layers of the database. Logical data independence is usually required for changing the conceptual schema without having to change the external schema or application programs. It allows us to make changes in a conceptual structure like adding, modifying, or deleting an attribute in the database. A very important advantage of using a DBMS is that it offers data independence. That is, application programs are insulated from changes in the way the data is structured and stored. Data independence is achieved through use of the three levels of data abstraction; in particular, the conceptual schema and the external schema provide distinct benefits in this area. Intuitively, some confidential information about faculty has been placed in a separate relation and information about offices has been added. The Courseinfo view relation can be redefined in terms of Faculty public and Faculty private, which together contain all the information in Faculty, so that a user who queries Courseinfo will get the same answers as before. Thus, users can be shielded from changes in the logical structure of the data, or changes in the choice of relations to be stored. This property is called logical data independence. In turn, the conceptual schema insulates users from changes in physical storage details. This property is referred to as physical data independence. The conceptual schema hides details such as how the data is actually laid out on disk, the file structure, and the choice of indexes.

### QUESTION 4

The conceptual schema (sometimes called the logical schema) describes the stored data in terms of the data model of the DBMS. In a relational DBMS, the conceptual schema describes all relations that are stored in the database. In our sample university database, these relations contain information about entities, such as students and faculty, and about relationships, such as students' enrollment in courses.

The physical schema specifies additional storage details. Essentially, the physical schema summarizes how the relations described in the conceptual schema are actually stored on secondary storage devices such as disks and tapes. We must decide what file organizations to use to store the relations and create auxiliary data structures, called indexes, to speed up data retrieval operations.

External schemas, which usually are also in terms of the data model of the DBMS, allow data access to be customized (and authorized) at the level of individual users or groups of users. Any given database has exactly one conceptual schema and one physical schema because it has just one set of stored relations, but it may have several external schemas, each tailored to a particular group of users. Each external schema consists of a collection of one or more views and relations from the conceptual schema. A view is conceptually a relation, but the records in a view are not stored in the DBMS.

### QUESTION 5

Database application programmers develop packages that facilitate data access for end users, who are usually not computer professionals, using the host or data languages and software tools that DBMS vendors provide. (Such tools include report writers, spreadsheets, statistical packages, and the like.) Application programs should ideally access data through the external schema. It is possible to write applications that access data at a lower level, but such applications would compromise data independence.

Design of the Conceptual and Physical Schemas: The DBA is responsible for interacting with the users of the system to understand what data is to be stored in the DBMS and how it is likely to be used. Based on this knowledge, the DBA must design the conceptual schema (decide what relations to store) and the physical schema (decide how to store them). The DBA may also design widely used portions of the external schema, although users probably augment this schema by creating additional views. Security and Authorization: The DBA is responsible for ensuring that unauthorized data access is not permitted. In general, not everyone should be able to access all the data. In a relational DBMS, users can be granted permission to access only certain views and relations. For example, although you might allow students to find out course enrollments and who teaches a given course, you would not want students to see faculty salaries or each other's grade information. The DBA can enforce this policy by giving students

permission to read only the Courseinfo view. Data Availability and Recovery from Failures: The DBA must take steps to ensure that if the system fails, users can continue to access as much of the uncorrupted data as possible. The DBA must also work to restore the data to a consistent state. The DBMS provides software support for these functions, but the DBA is responsible for implementing procedures to back up the data periodically and maintain logs of system activity (to facilitate recovery from a crash). Database Tuning: Users' needs are likely to evolve with time. The DBA is responsible for modifying the database, in particular the conceptual and physical schemas, to ensure adequate performance as requirements change.

A DBA needs to understand query optimization even if s/he is not interested in running his or her own queries because some of these responsibilities (database design and tuning) are related to query optimization. Unless the DBA understands the performance needs of widely used queries, and how the DBMS will optimize and execute these queries, good design and tuning decisions cannot be made

#### QUESTION 6

Security facility: there is a need as only one computer has access to the database-a person can access his computer and can get access to full database

No transactions are happening except by him on the database so no concurrency control needed

Crash recovery needed as crash can happen on the only computer having access to the database

View mechanism also needed when queries are made

For making queries, a query language is needed

- ❖ A security facility.
- ❖ A security facility is necessary because Scrooge does not plan to share his list with anyone else. Even though he is running it on his stand-alone PC, a rival duckster could break in and attempt to query his database. The database's security features would foil the intruder.
- ❖ Concurrency control.
- ❖ Concurrency control is not needed because only he uses the database.
- ❖ Crash recovery.
- ❖ Crash recovery is essential for any database; Scrooge would not want to lose his data if the power was interrupted while he was using the system.
- ❖ A view mechanism.
- ❖ A view mechanism is needed. Scrooge could use this to develop "custom screens" that he could conveniently bring up without writing long queries repeatedly.
- ❖ A query language.
- ❖ A query language is necessary since Scrooge must be able to analyze the dark secrets of his victims. In particular, the query language is also used to define views.

#### QUESTION 7

Ddl-schemas define the characteristics of some entity in the real world

Dml-actual values or rows define actual instances/examples of entities in the world

Buffer model-not needed except when we have to store a lot of things in buffers and need to manage buffers

Data model-needed to represent relationships between entities of the real world

#### QUESTION 8

Changes are made to the disk-so disk space manager

The disk space management layer has to be rewritten to take advantage of the new functions on OS files. It is likely that the buffer management layer will also be affected.

# QUESTION 9-READ FROM SOLUTIONS PDF