

CS 448 Spring 2024 Homework 1

Sean Rettig

January 20, 2024

Question 1. (0.50 pts)

- (a) Discuss the main capabilities that should be provided by a DBMS
- (b) Highlight the difference of using a database system with respect to the following approaches. Please, make sure to highlight the benefits offered by a DBMS.

1. Traditional file systems
 2. In-memory storage
-

- (a) One of the main capabilities that should be provided by a database include a self describing nature. The database shouldn't just store its own data, but it should also store a complete definition of the database structure and its constraints.

Another capability should be an insulation between programs and data, otherwise called program-data independence. This allows changing data structures and storage organization without having to change the DBMS access programs.

Thirdly, there is data abstraction. The DBMS should provide a conceptual representation of its data, and should hide the actual physical implementation.

The database should also support multiple view of the data. The user should not be limited to the actual positioning of the data, and should be able to view it in multiple ways.

Finally, the database should facilitate sharing of data and multiuser transaction processing. Multiple users should be able to access the data at the same time.

- (b)
 1. The traditional filesystem does not include the main characteristics of a DBMS well. There is no self describing nature in a filesystem, The program or software needs to know where the information is stored to be able to access it. There is no conceptual representation of the data either. Additionally, there are not multiple views of the data. Finally, sharing information is very difficult because traditional filesystems are very user focused, so if someone has information stored in their own filesystem place, it is not readily accessible by anyone else.
 2. It is not explicitly designated how in-memory storage would be designed. However, if the application software is storing everything in-memory and in the program that is running, all it takes is for the program to stop or the computer to shut down and the data will be gone.

Question 2. (0.50 pts)

Cite some examples of integrity constraints that you think can apply to the database shown in Figure 1.2 in the textbook. Please, provide at least two examples for each type of the following constraints: unique constraints and referential integrity constraints

Unique: Lets say that each section must have a unique section_identifier, and each student must have a unique student_number.

Referential: Each student_number in the grade report must refer to a student in STUDENT. Each section identifier in GRADE_REPORT must refer to a section_identifier in SECTION

Question 3. (0.25 pts)

What is the difference between a database schema and a database state? Please, use examples to support your answer.

The database schema is the description of the database and what information it holds, and how it all relates to each other. The database state is the current data actually in the database.

For example, in Figure 1.2, the database schema for the STUDENT table says that it includes the students name, Student_number, Class, and Major. The database state includes actual student names, and the rest of the information like Name: Smith, Student_number: 17, Class: 1, Major: CS and then other students as well.

Question 4. (0.50 pts)

- (a) What is logical independence and physical independence? Please, use examples to support your answer.
 - (b) Which one is harder to achieve? Explain why.
-

- (a) Data independence refers to the capacity to change different levels in the three scheme architecture without affecting each other. Logical independence is the ability to change the conceptual schema, while physical independence refers to the ability to change the internal schema.

If you had logical independence, you could add a row to GRADE_REPORT in Figure 1.2 and all the external views that call for data that is in the other columns will not be affected at all.

If you had physical independence, you could change the way that GRADE_REPORT is actually stored in the DBMS by creating a better way of accessing the data. The conceptual schema wouldn't change and neither would the external schema.

- (b) Logical independence is harder to achieve. This is because it allows structural and constraint changes, but you cannot affect application programs.

Question 5. (0.50 pts)

What are the different components of a DBMS? Discuss the importance and need of each component.

- **Interfaces for DBA staff, Casual Users, Application Programmers, and Parametric Users:** This is important because the data is needed to be accessed, but not everyone needs the same type of access. Some need more advanced while others need simpler access.
- **DDL Compiler:** This processes schema definitions. This is important to create the catalog.
- **Query Optimizer:** This component takes queries and looks for ways to optimize them. Either by consulting the catalog or through other means. It is important because it improves performance.
- **Precompiler:** This extracts DML commands and sends them to the DML compiler for compilation into object code. This is important to be able to communicate to the database.
- **Runtime Database processor:** This executes the commands, query plans, and transactions. This is important in accessing the database.
- **Stored Data Manager:** This uses basic operating system services for carrying out low-level IO to disk and main memory. It's important because the low-level IO needs to reach the higher level calls.
- **Concurrency control and Backup and Recovery systems:** Handle concurrency access as well as backup and recovery of the database. These are important to support the main capabilities of the database being concurrency access and then another important feature of backup.

Question 6. (0.50 pts)

- (a) What is a user transaction?
- (b) What properties must be maintained for a transaction by the database system? Please, discuss each property using an example to support your argument.
-

- (a) A transaction, as the textbook says, is an executing program or process. It can include one or more database accesses. It can read or update records as well.

- (b) Properties that the DBMS have must include isolation as well as atomicity.

Isolation ensures that each transaction takes place without affecting other transactions. This means that if I had one transaction asking for all the students in one class, but in the middle of that transaction I update the students in the class, the first transaction will have the same result as if I didn't do the second one.

Atomicity ensures that either all the database operations in a transaction are executed or none are. For example, if, in my transaction, I calculated GPA of all students and added that to their records, but my second operation doesn't complete successfully, then the first part of of my transaction will not execute.

Question 7. (0.25 pts)

Why is accessing a disk block expensive? Discuss the time components involved in accessing a disk block.

It is expensive for a couple of reasons. Given the address, the disk controller must physically position the read/write head on the on the correct track. This is called seek time and can take anywhere from 5 to 10 msec.

After seek time there is also rotational delay. This is waiting for the beginning of the desired block to rotate into position under the read/write head. For a 15,000 rpm (pretty fast) disk, the average rotational delay is 2 msec.

After that there is actually transferring the data. This is called the block transfer time, which can differ between disks and OSes, but is generally less than the seek time and rotational delay.

Thus, disk access time = seek time + rotational delay + block transfer time.