W4111 – Introduction to Databases Sections 002, V002; spring 2022

Homework 1 – Written Assignment

Instructions

- The homework submission date/time is 06-Feb-2022 at 11:59 PM.
- Submission format is a PDF version of this document with your answers. Place your answers in the document after the questions.
- The name of your PDF must be <UNI>_S22_W4111_HW1_Written.pdf. For example, mine would be dff9_S22_W4111_HW1_Written.pdf
- You must use the Gradescope functions to mark the location of your questions/answers in the submitted PDF. Failure to mark pages will cause point deductions.
- You can use online sources but you must cite your sources. You may not cut and paste text..
- Questions typically require less than five sentences for an answer. You will lose points if your answer runs on and wanders.

  “Verbosity wastes a portion of the reader’s or listener’s life.”

Questions

Question 1: Briefly explain the terms *structured data, semi-structured data* and *unstructured data.* Give an example of each type.

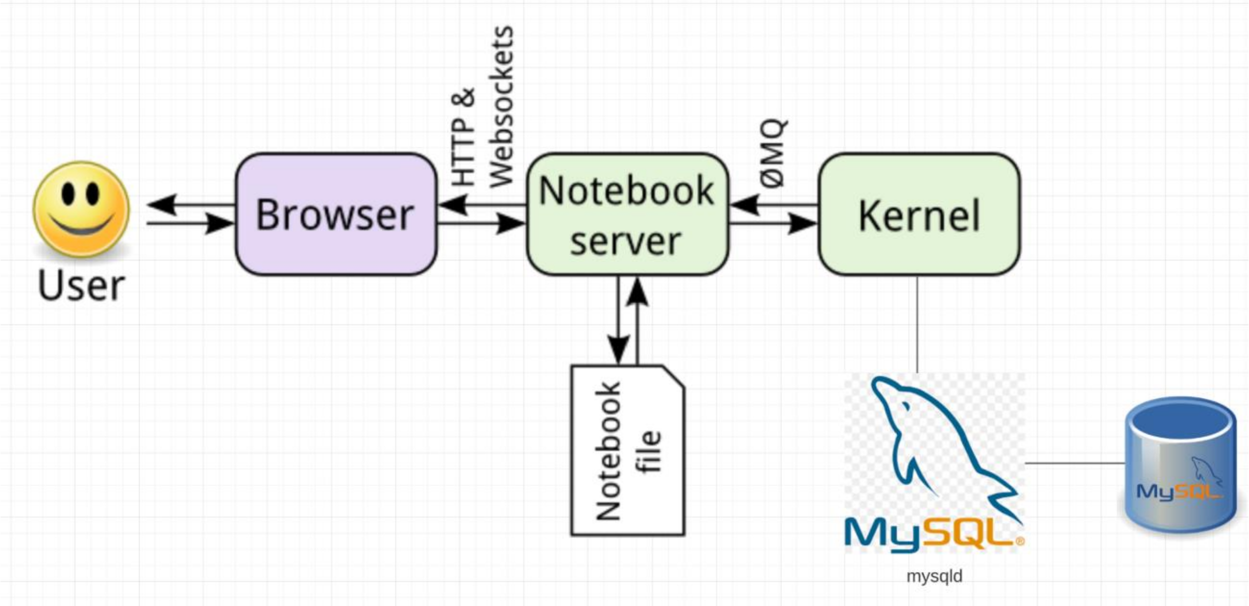
Structured data is highly organized data. It has entries with the same format and predefined length following the same order. Example would be Excel spreadsheets and relational database tables.
Semi-structured data is data organized in some degree. There will be some tags or markers separating elements and enforcing hierarchies, but the element size would vary and their order is not important. Example would be HTML files and JSON files.

Unstructured data is data that has no form predefined, thus able to be stored in any kind of file. Most of the data is unstructured data, example would be .png files, .mp4 files, .mp3 file, Word files, PDF files.

Question 2: Briefly explain the concept of *metadata*. For a presentation (PowerPoint, Google Slides), what would be some examples of metadata?

Metadata is data used to provide information about other data. It could summarize basic information about data so that tracking and working with that data would be much easier.
For a presentation, some example of metadata would be the size of the file, where is it stored and in what form is it stored, data quality, purpose of the data.

Question 3: The following diagram is an overview of Jupyter Notebook's runtime model when the notebook is using MySQL. Is this a 2-tier application or 3-tier application? Briefly explain why.



Above is a 3-tier application, as I can see that there are three main components, browser, notebook server and kernel. These correspond to presentation tier, application tier and data tier. Browser is the presentation tier which is built with HTML5. Notebook Server is the application tier which is written in a programming language and supports the application's core functions. Kernel is the data tier consisting of a database and a program for managing read and write access to a database.

Question 4: Briefly define and explain procedural and declarative languages. Is SQL procedural or declarative?

Procedural language is a computer programming language that follows a set of commands in order. It uses functions, conditional statements, and variables to create programs that calculate and display desired result.
Declarative language is a non-imperative style of programming language that describe its results without explicitly listing commands.
SQL is declarative language.

Question 5: List 4 advantages/differences of database management systems (DBMS) compared to programs and files for data processing. List two disadvantages of DBMS?

Advantage 1: DBMS solves the problem of redundancy
Advantage 2: DBMS has a high security level
Advantage 3: DBMS allows data integrity
Advantage 4: DBMS could support multiple users

Disadvantage 1: DBMS use huge memory size
Disadvantage 1: DBMS create huge size of database

Question 6: In a relational DBMS, columns/attributes should be *atomic.* Briefly explain what this means. If a table has a column *name* of the form "last name, first name", is this atomic?

Being atomic means that the column/attributes cannot be divided or split in smaller parts. If a table has a column name of the form "last name, first name", then it is not atomic.

Question 7: Attributes/columns have *types,* e.g. int, varchar(128), timestamp. An attribute/column values must be from a *domain*? What is the difference between a type and a domain (hint: domain constraints)?

An attribute column values mist be from a domain. A domain is a computer programming concept, which is the range of values permitted for a specific attribute, while data type of a column defines what value the column can hold: integer, character, money, date and time, binary, etc.

Question 8: There are four common types of people that interact with a database management system. List and briefly explain each of the four types.

1: Application programmers, who writes application programs such as PL and java to use the database
2: End users, who access the database from the terminal end. End users simply use the developed applications and don’t have any knowledge about the design and working of database.
3. Database Administrators, people who are responsible for everything relating to database such as making policies, strategies and providing technical supports
4. System Analyst, people who is responsible for the design, structure and properties of database

Question 9: Briefly explain the concepts of database *instance* and *schema*?

Database Instances are collection of information stored at a particular moment, or say, at a given point of time. Thus, instance is a dynamic value which keeps on changing.
Database schema is the overall design of the database, the logical structure of a database that will not be changed frequently

Question 10: Explain the concept of *physical data independence* and the importance of the concept.

Physical data Independence separates conceptual levels from the internal or physical levels. It allows providing a logical description of the database without the specifying physical structures. This concept is important because it could change the physical storage structures or devices while affecting the conceptual schema. Any change would be absorbed by the mapping between the conceptual and internal levels.