

# 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 organized data that always has rows and columns as tables. They have relational keys in database. E.g., table “actors” in IMDB database.

Semi-structured data still has some format but is not as organized with rows and columns as structured data. E.g., IMDB actors documents in JSON format.

Unstructured data is raw data without organizing. E.g., a list of students' papers word documents.

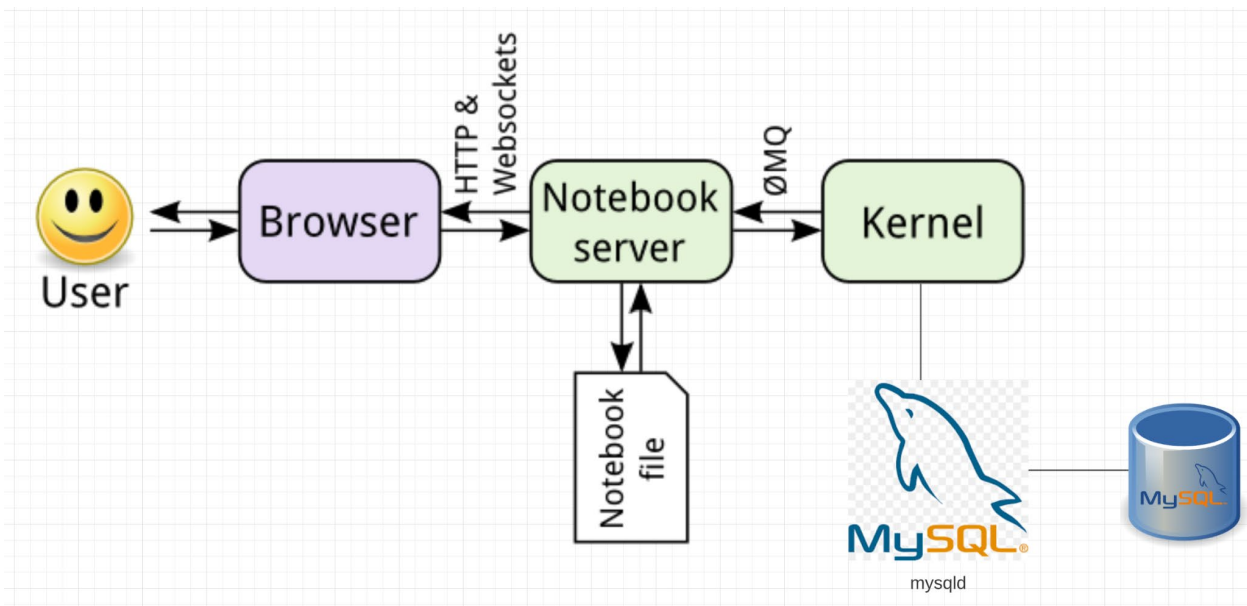
Reference: <https://k21academy.com/microsoft-azure/dp-900/structured-data-vs-unstructured-data-vs-semi-structured-data/>

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

Metadata is data to describe other data.

Example: author, size, file modification date.

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.



It's a 3-tier application. Because the user doesn't use jupyter notebook with direct database calls, it communicates with mysql to access data.

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

Procedural languages require the user to specify how to get the needed data, including all the steps.

Declarative languages do not require the user to specify how to get the needed data.

SQL is declarative since the queries don't specify exact steps to get the result.

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

Advantages: Improve data security, can set the password to protect the dataset.

Display data more organized, easy to use.

Integrate different datasets.

Better access data, declarative languages do not require the user to specify how to get the needed data.

Disadvantages:

Need more hardware and software.

Frequently update the system and the environment.

Reference: <https://tutorialink.com/dbms/advantage-and-disadvantages-of-dbms.dbms>

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?

It means indivisible.

No, we can partition it as "first name", "last name". For example, if we want to email the person, we can use Mr./Mrs + "last name".

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 should be from a domain.

Type is a data type like varchar(8), domain is a combination of data type and domain constraints, like varchar(8) NOT NULL.

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.

Naive users. Interact DBMS with a previously written application program.

Application programmers. The people who write application programs.

Sophisticated users. Interact DBMS without writing programs, but using database queries.

Specialized users. The people who write untraditional database applications like video.

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

Instance: The actual content in a database.

Schema: The structure of the database, without actual content.

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

It's the ability to adjust the physical schema without changing the logical schema. So the changes in some parts would not influence other parts of the data seriously. Applications rely on the logical schema.