

4.14: Study Questions - Data and Databases

Study Questions

? Study Question 4.14.1

What is the difference between data, information, and knowledge?

Answer

Data are raw facts and figures. Information is data with context and organization, answering who, what, when, where. Knowledge is the application of information and experience to gain insight for action and decision-making.

? Study Question 4.14.2

Explain in your own words the difference between hardware and software components of information systems in your own words.

Answer

Hardware is the physical components of an information system like computers, networks, and devices. Software is the programs and instructions running on the hardware, including operating systems, applications, utilities, etc.

? Study Question 4.14.3

What is the difference between quantitative data and qualitative data? In what situations could the number 63 be considered qualitative data?

Answer

Quantitative data are numeric values from measurement or calculation. Qualitative data are descriptive qualities that cannot be measured numerically. The number 63 could be qualitative if it represents a code or identifier, not a measured quantity.

? Study Question 4.14.4

What are the characteristics of a relational database?

Answer

Characteristics of a relational database include organized into tables with rows and columns; each row has a unique key, columns have field definitions, tables are related via keys linking rows.

? Study Question 4.14.5

When would using a personal DBMS make sense?

Answer

A personal DBMS makes sense for small data management needs for one person or a small workgroup, not enterprise-wide sharing.

? Study Question 4.14.6

What is the difference between a spreadsheet and a database? List three differences between them.

Answer

A spreadsheet is a simple data organization tool optimized for basic data manipulation by an individual. A database is designed for complex querying, large data volumes, multi-user access, and ensuring data integrity.

? Study Question 4.14.7

Describe what the term normalization means.

Answer

Normalization is designing a database to reduce duplication and redundancy and improve flexibility to change. It involves separating data into multiple tables and establishing relationships.

? Study Question 4.14.8

What is Big Data?

Answer

Big data refers to large, complex datasets with high volume, velocity, and variety that require advanced storage, processing, and analyzing techniques.

? Study Question 4.14.9

Name a database you interact with frequently. What would some of the field names be?

Answer

A database could be something like contacts on your phone - fields would include name, phone number, email, birthdate, etc.

? Study Question 4.14.10

Describe the benefits and what open-source data is.

Answer

Open source data is freely shared data that users can access and modify as needed since the source code is publicly available. Benefits include flexibility, transparency, and cost.

? Study Question 4.14.11

Name three advantages of using a data warehouse.

Answer

Data warehouse advantages: Integrates data from disparate sources, provides enterprise-wide analytics, and creates historical records of data over time.

? Study Question 4.14.12

What is data mining?

Answer

Data mining is analyzing large datasets to discover patterns and automatically extract meaningful information and insights.

? Study Question 4.14.13

What is metadata? Provide one example.

Answer

Metadata is "data about data" - it provides information and context about the actual data in a database. An example is a database column's data type (e.g. text, number, date, etc.), which describes the type of data contained in that column.

? Study Question 4.14.14

What are some potential ethical concerns when utilizing personal data for business analytics and data mining?

Answer

Some ethical concerns include:

- Lack of consent from individuals about data collection and use
- Violating privacy expectations by combining various data sources
- Misuse of data profiling to discriminate against groups of people
- Drawing faulty conclusions from incomplete data or incorrect analytics
- Data breaches that expose personal information
- Manipulating consumers through personalized advertising
- Perpetuating existing biases by using biased or unrepresentative data
- Quantitative data are numeric values from measurement or calculation.

Exercises

1. Review the design of the Student Clubs database earlier in this chapter. Reviewing the lists of data types given, what data types would you assign to each of the fields in each of the tables? What lengths would you assign to the text fields?
2. Review structured and unstructured data and list five reasons to use each.
3. Using Microsoft Access, download the database file of comprehensive baseball statistics from the website
4. SeanLahman.com. (If you don't have Microsoft Access, you can download an abridged version of the file [here](#) that is compatible with Apache Open Office). Review the structure of the tables included in the database. Identify three different data-mining experiments you would like to try, and explain which fields in which tables would have to be analyzed.
5. Do some original research and find two examples of data mining. Summarize each example and then write about what the two examples have in common.
6. Conduct some independent research on the process of business intelligence. Using at least two scholarly or practitioner sources, write a two-page paper giving examples of how business intelligence is used.
7. Use the Internet to research two software tools or technologies focused on managing knowledge. One tool should be for personal use, and one for business/organizational use. Write a one-page comparison explaining what each tool does, key features, who would use it, and how it facilitates knowledge management. Reflect on how the personal and organizational tools differ in their knowledge management approaches.
8. Install a simple personal DBMS like OpenOffice Base and design a basic database to store information on books, music, or movies. Focus on key fields and data types.

9. Search public data repositories and find 3 datasets of interest. Explain how basic data analysis could provide insights.
10. Interview 2 peers about databases they use and how data is valuable to them. Summarize key learnings in a paragraph.
11. Research a widely used database platform. Explain the key capabilities and a simple use case in 2-3 sentences.
12. Draw an entity relationship diagram for a basic scenario like university enrollment with 5-6 entities.
13. Find an article discussing an ethics case about personal data. Summarize the key issue and perspective in a paragraph.
14. Take a spreadsheet containing duplicated student data and restructure it into a simple, normalized database format.
15. Import a public dataset into a spreadsheet. Make basic charts and graphs to visualize and summarize top-level trends.

This page titled [4.14: Study Questions - Data and Databases](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Ly-Huong T. Pham and Tejal Desai-Naik \(Evergreen Valley College\)](#) .

- [4.14: Study Questions](#) by Ly-Huong T. Pham, Tejal Desai-Naik, Laurie Hammond, & Wael Abdeljabbar is licensed [CC BY 3.0](#).