DAMG6210 - Data Management and Database Design

Homework 01

1-28. Reread the definitions for *data* and *database* in this chapter. Database management systems only recently began to include the capability to store and retrieve more than numeric and textual data. What special data storage, retrieval, and maintenance capabilities do images, sound, video, and other advanced data types require that are not required or are simpler with numeric and textual data?

Definitions:

Data: A stored representation of objects and events that have meaning and importance in the user's environment.

Database: An organized collection of logically related data.

The special data storage, retrieval, and maintenance capabilities required for managing images, sound, video, and other advanced data types compared to numeric and textual data are:

1. ***Storage:***

- Numeric data typically requires minimal storage, and text data can vary in size but is generally measured in bytes or kilobytes. These data types can be stored in fixed-length fields or variable-length records in a database.

- Images, sound, and video files are usually much larger than text or numeric data, occupying more storage space. To manage this, we use special compression techniques, like JPEG for images or MP3 for audio, to reduce file sizes without losing too much quality. These media types also need more advanced systems for storage and retrieval because of their size and complexity.

2. ***Retrieval***

- Structured data, such as numbers, dates, and text, is stored in well-organized tables, like those in a relational database. To access this type of data, we use SQL (Structured Query Language), which makes retrieving information quick and efficient. Since the data is arranged in a predefined schema (like rows and columns), it is easy to filter, sort, and search through. Using indexes helps speed up the process, making data retrieval faster.

- Unstructured or complex data, such as images, videos, emails, or social media posts, doesn't follow a fixed format like structured data. Instead of relational databases, this data is often stored in NoSQL databases, data lakes, or object storage systems. Retrieving this data can be slower and more complicated because it lacks a strict organization. You might need specialized tools like full-text search engines for documents or machine learning for analyzing images and videos. Unstructured data often requires metadata (information about the data) to help locate and retrieve the needed information more efficiently.

3. ***Maintenance Capabilities***

- Maintenance of numeric and textual data is relatively straightforward due to the structured nature of relational databases. These databases adhere to a fixed schema, making updates, additions, and deletions predictable and manageable. Data integrity is preserved through constraints such as primary and foreign keys, and regular backups ensure data safety and quick recovery in case of issues. The consistent data format allows for efficient management and troubleshooting.

- On the other hand, maintaining images, sound, video, and other advanced data types involves greater complexity. These data types do not conform to a fixed schema and can exhibit high variability in format and size, thus requiring more sophisticated systems and processes for maintenance. The large volume and diverse formats of media files necessitate scalable storage solutions and advanced backup strategies. Regular updates and optimizations are necessary to handle growth and ensure efficient access and performance. Effective

metadata management becomes crucial for organizing and retrieving these files, and specialized tools are often required for tasks such as indexing and analyzing content. Maintaining the quality and integrity of media files over time can also be challenging, necessitating additional resources and techniques.

1-29. Table 1-1 shows example metadata for a set of data items. Identify three other columns for these data (i.e., three other metadata characteristics for the listed attributes) and complete the entries of the table in Table 1-1 for these three additional columns.

Table 1-1 Class Roster

Data Item	***Metadata***								
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**	**Default Value**	**Creation Date**	**Access Control**
Course	Alphanumeric	30			Course ID and name	Academic Unit	N/A	2023-01-18	Admin
Section	Integer	1	1	9	Section number	Registrar	1	2023-01-18	Admin
Semester	Alphanumeric	10			Semester and year	Registrar	Current Term	2023-01-18	Admin
Name	Alphanumeric	30			Student name	Student IS	N/A	2023-01-18	Student
ID	Integer	9			Student ID(SSN)	Student IS	N/A	2023-01-18	Admin
Major	Alphanumeric	4			Student major	Student IS	N/A	2023-01-18	Student
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit	0.0	2023-01-18	Admin

1-30. In the section "Disadvantages of File Processing Systems," the statement is made that the disadvantages of file processing systems can also be limitations of databases, depending on how an organization manages its databases. First, why do organizations create multiple databases, not just one all-inclusive database supporting all data processing needs? Second, what organizational and personal factors are at work that might lead an organization to have multiple, independently managed databases (and, hence, not completely follow the database approach)?

Reasons why organizations create multiple databases instead of one all-inclusive database:

- Different departments often have unique needs, leading them to develop separate databases tailored to their specific functions. This results in multiple, specialized databases within the organization.
- Many organizations have older databases that were established before modern systems were introduced. Integrating these legacy systems can be costly and risky, so they are often maintained alongside newer databases.
- Certain business functions, like customer management or financial analysis, may require databases optimized for their specific purposes. This specialization can lead to the creation of multiple, function-specific databases.
- As organizations grow and evolve, their database systems often develop incrementally. This gradual change can result in a collection of separate databases rather than a single, unified system.
- Implementing and managing a large, all-encompassing database can be very expensive and complex. To avoid these issues, organizations might opt for smaller, more manageable databases tailored to different needs.
- Concerns about data ownership and the need for strict access control can lead departments to maintain their own databases, resulting in multiple independent systems across the organization.

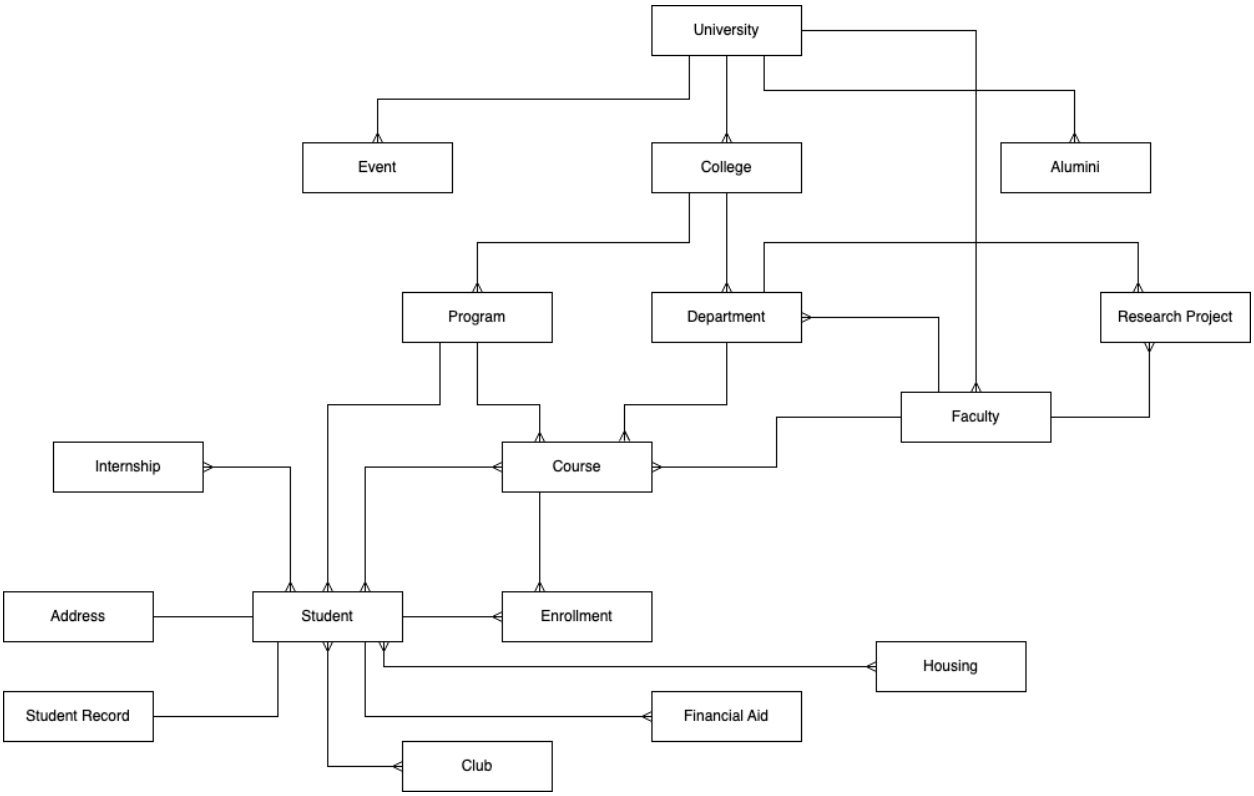
Organizational and personal factors leading to multiple, independently managed databases:

- Without clear direction from top management, departments may create their own databases independently, leading to a proliferation of systems.

- Disagreements over data standards and access rights can cause departments to develop their own databases to maintain control over their data.
- Converting legacy systems to a unified database can be costly and challenging, causing resistance and leading to the continued use of multiple databases.
- A lack of centralized database management can result in uncoordinated database development across different parts of the organization.
- Traditional organizational silos often lead departments to create and maintain their own databases rather than collaborating on a unified approach.
- Not all departments have access to the same level of expertise, leading some to create their own solutions based on available skills.
- The need for quick development and deployment of database applications may lead some departments to create standalone databases as a faster solution.
- Departments handling sensitive data may prefer to manage their own databases to ensure better control over access and security.
- Personal preferences among IT staff or limited departmental budgets may also influence the decision to maintain separate databases rather than investing in a centralized system.

1-38. Consider your business school or other academic unit as a business enterprise.

a. Define several major data entity types and draw a preliminary enterprise data model (similar in notation to Figure 1-3a).



References

Hoffer, J. A., Ramesh, V., & Topi, H. (2016). *Modern database management* (13th ed.). Pearson.