

DATA MANAGEMENT AND DATABASE DESIGN

HOMEWORK: WEEK- 1

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?

- **STORAGE:** Unstructured data such as images, audio, video, location/GPS etc., require considerably more storage volume when compared to structured data such as tables, arrays, or spreadsheets.
- **RETRIEVAL:** Since multimedia objects use relatively more storage, it affects the performance and consumes more bandwidth. This will in turn affect the speed of content retrieval and content management. Locating and retrieving multimedia objects are also difficult since they do not have a particular notation (like keys or index) for quick access or recovery.
- **MAINTENANCE:** In case of multimedia objects, updating contents are relatively difficult when compared to structured data. In case of any discrepancy, the whole object must be replaced, while with structured data it can be edited more easily. Multimedia objects would also result in loss of data when original data is removed from source location and would also have quality issues.

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.

DATA ITEM	METADATA									
NAME	TYPE	LENGTH	MIN	MAX	DESCRIPTION	SOURCE	TARGET	DATE OF CREATION	DATE OF UPDATION	PRIMARY KEY
Course	Alphanumeric	30			Course ID and Name	Academic Unit	Academic Unit	2/15/2022	NULL	No
Section	Integer	1	1	9	Section number	Registrar	Student IS	7/13/2022	NULL	No
Semester	Alphanumeric	10			Semester and year	Registrar	Student IS	5/19/2022	8/1/2023	No
Name	Alphanumeric	30			Student Name	Student IS	Student IS	5/19/2022	NULL	No
ID	Integer	9			Student ID(SSN)	Student IS	Academic Unit	5/19/2022	NULL	Yes
Major	Alphanumeric	4			Student Major	Student IS	Academic Unit	5/19/2022	NULL	No
GPA	Decimal	3	0	4	Student grade point average	Academic unit	Academic Unit	NULL	12/23/2022	No

Metadata Keys:

- **TARGET:** Where the data will eventually be stored for further reference.
- **DATE OF CREATION:** Date when was the data for that particular data item was created.

(Eg: Every data item field would be created after every point of the admission process.

(i.e.) Course would be created after admit is accepted,

Semester, ID, name, major would be created after the student is enrolled

Section would be allocated after student registers for a course,

GPA would be null in the beginning.)

- **DATE OF UPDATION:** Date when the data for that particular data item was updated.

(Eg: Few data item fields would be created after every point of the admission process. GPA would be updated at end of every semester,

Semester would be updated after end of every semester)

- **PRIMARY KEY:** If that particular data item is a primary key or not.

(Eg: ID is the primary key for this table)

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)?

Part A: Why do organizations create multiple databases, not just one all-inclusive database supporting all data processing needs?

While all-inclusive databases store all the information under one roof, they have their own disadvantages:

- To begin with, most organizations generate enormous amounts of data and storing all this data under one single database would lead to performance issues, sooner or later, depending on how much data is being processed at one given point. When the data is split across multiple databases, the performance increases, since only one part of the data is served at once.
- When there are multiple databases in the place of one, it is more secure as all the components are split across various databases and it also gives the database administrators more power to modify and retrieve data without affecting the whole database.
- When all the data is incorporated in a single database, it makes the readability relatively difficult. When it comes to several databases, there a more structured format, which makes locating and retrieving data easier.

Part B: Why do organizations create multiple databases, not just one all-inclusive database supporting all data processing needs?

- Multiple databases simplify the process of administration and editing data when data is split across several databases is easier when compared to a single database. When databases are split according to subsystems, locating issues is much simpler and each database can be separately scaled without affecting all the other databases.
- There would be data/databases from several third-party organizations or clients with which the organization must have to work and hence it is easier to incorporate all the data whilst having multiple databases instead of a single database.
- Creating a database with huge volume of data, at one go, will relatively cost more than building one database at a time. Depending on funding, it would be easier on the organization to build databases one at a time instead of a single huge database at once.

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 Fig 1-3a)

Assumption: Admission process in a university.

Major entities - Flow process:

- Student fills up application form.
- Admissions office verifies the application
- Admission office sends an admit
- Student accepts the admit and proceeds to pay fees
- Student center reviews the application and acknowledges the fee payment
- Student center gives out and unconditional offer
- Student accepts the unconditional offer

Enterprise data model notation:

