

Data Management and Database Design

INFO 6210

Fall 2016

Assignment 1

Submission Date – Sep/19/2016 Monday

Student Name: Malick Fairoz Sayeed Abuthahir

NUID: 001235450

Program: MS in Information Systems

Professor Name: Yusuf Ozbek

College: College Of Engineering

University: Northeastern University

((Part 1-A))

Review Question

1. Define each of the following terms:

a. Data

Data is the information stored in the tables in terms of rows and columns under the database. This includes stored representations of objects and events that have meaning and importance in the user's environment.

b. Information

The information is defined as data that have been processed in such a way that the knowledge of a person increases when using the data.

c. Metadata

Metadata are the data that describes the properties or characteristic of the end user data and the context of that data like source of the data, where the data are stored, ownership and usage.

d. Database Application

Database application is a set of related programs that is used to perform a series of database activities like create, update, delete, and select on behalf of database users.

e. Data Warehouse

Data warehouse is an integral decision support database whose content is derived from various operational databases, including personal, workgroup, department, and ERP databases. It also provide user with the opportunity to work with historical data to identify patterns and trends and answers to strategic business questions.

f. Constraint

A Constraint is a rule that cannot be violated by database user. For example if a customer places an order, the constraints that ensures that the customer and the order remains associated is called a relational integrity constrain and it prevents an order from being entered without specifying who placed the order.

g. Database

A database is an organized collection of logically related data which can be easily accessed, managed and updated.

h. Entity

An entity is like a noun in that it describes a person, a place, an object, an event, or a concept in a business environment for which information must be recorded and retained.

i. Database Management System

A Database Management System is a software system that is used for creating, updating, storing and retrieving the data stored in the database and provide controlled access to user database.

j. Client/Server Architecture

Client/server architecture is a computing model in which the server hosts, delivers and manages most of the resources and services to be consumed by the client. This type of architecture has one or more client computers connected to a central server over a network or Internet connection

k. Systems Development Life Cycle (SDLC)

Systems development life cycle is the traditional methodology used to develop, maintain, and replace information systems.

l. Agile Software Development

An approach to database and software development that emphasizes individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and response to change over following a plan.

m. Enterprise Data Model

Enterprise data modeling encompass many databases, describing the scope of data maintain by the organization

n. Conceptual Data Model

Conceptual data model identify scope of database requirements for proposed information systems and analyze the overall data requirements for business functions supported by database.

o. Logical Data Model

Logical data model describes the data in as much details as possible, without regard to how they will be physically implemented in the database. This also identifies data integrity and security requirements, and populate repository.

p. Physical Data Model

A physical data model shows all tables structures, including column name, column data type, column constraints, primary key, foreign key and relationship between tables.

2. Match the following terms and definitions:

Answers		Terms	Definitions	
C		data	A	a. data placed in context or summarized
K		Database application	B	b. application program(s)
L		constraint	C	c. facts, text, graphics, images, etc.
G		repository	D	d. a graphical model that shows the high-level entities for the organization and the relationships among those entities
F		metadata	E	e. organized collection of related data
M		data warehouse	F	f. includes data definitions and constraints
A		information	G	g. centralized storehouse for all data definitions
J		user view	H	h. separation of data description from programs
B		Database management system	I	i. a business management system that integrates all functions of the enterprise
H		Data independence	J	j. logical description of portion of database
E		database	K	k. a software application that is used to create, maintain, and provide controlled access to user databases
I		Enterprise resource planning (ERP)	L	l. a rule that cannot be violated by database users
R		Systems development life cycle (SDLC)	M	m. integrated decision support database
O		prototyping	N	n. consist of the enterprise data model and multiple user views
D		enterprise data model	O	o. a rapid approach to systems development
Q		Conceptual schema	P	p. consists of two data models: a logical model and a physical model
P		Internal schema	Q	q. a comprehensive description of business data
N		external schema	R	r. a structured, step-by-step approach to systems development

3. Contrast the following terms:

a)	Data dependence	Data independence
	This is property allows data to change and evolve without changing the application programs that process the data	This is property will not allows data to change and evolve without changing the application programs that process the data
b)	Structured data	Unstructured data
	The data which are under data types of numeric, characters and dates are Structured data.	The data which are under data types of picture, video and audio are Unstructured data.

c)	Data	Information
	Data is raw, unorganized facts that needs to be processed. Data can be something simple and seemingly random and useless until it is organized.	When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information.
d)	Repository	Database
	Repository is a database which stores metadata which is a centralized knowledge base of all data definition, data relationship, reports and other system components.	Database is an organized collection of logically related data. It contains occurrence of data.
e)	Entity	Enterprise data model
	The data or information in the user environment about which the organization wishes to maintain is known as Entity.	The scope and general content of the organizational database which is used for the database development is known as enterprise data model.
f)	Data warehouse	ERP system
	Data warehouse provides user to work with historical data to identify the patterns and trends of strategic business.	ERP system provides the data necessary for the enterprise to examines and manages activities such as manufacturing, sales, finance, marketing, human resources.
g)	Two-tier databases	Multitier databases
	There is a significant need for the functionality that needs to be programmed into the application on the user's computer. For each change in the business logic or user interface, each client computer that has the application needs to be updated.	In multitier there is no much need for the functionality that needs to be programmed into the application on the user's computer. The client computer need to be updated for every business logic or user interface changes.
h)	Systems development life cycle	Prototyping
	SDLC is a traditional methodology used to develop, maintain and replace information systems.	Prototyping is an iterative process of systems development in which requirements are converted to a working systems
i)	Enterprise data model	Conceptual data model
	It is similar to conceptual data modeling, but addresses the unique requirements of a specific business.	It identifies the highest level relationships between different entities.
j)	Prototyping	Agile software development
	Prototyping is an iterative process of systems development in which requirements are converted to a working systems	It is an approach to a database and software development that emphasizes individuals and interactions over process and tools.

4. List five disadvantages of file processing systems.

- 1) Program Data Dependence
- 2) Duplication of Data
- 3) Limited Data Sharing
- 4) Lengthy Development Times
- 5) Excessive Program Maintenance

5. List the nine major components in a database system environment.

- 1) Computer – aided software engineering (CASE) tool.
- 2) Repository
- 3) Database Management System (DBMS)
- 4) Database
- 5) Application Programs
- 6) User Interface
- 7) Data and Database Administrators
- 8) Systems Developers
- 9) End Users

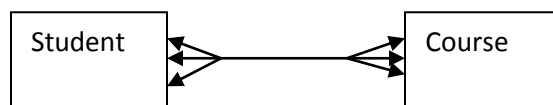
((Part 2-A))

Problems and Exercise

- 1) For each of the following pairs of related entities, indicate whether (under typical circumstances) there is a one-to-many or a many-to-many relationship. Then, using the shorthand notation introduced in the text, draw a diagram for each of the relationships.

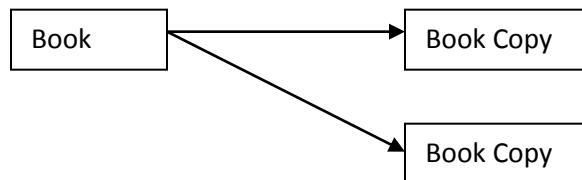
a. STUDENT and COURSE (students register for courses)

Since one student can take any number of courses and one course can be taken by any number of students, It is known as Many To Many Relationship.



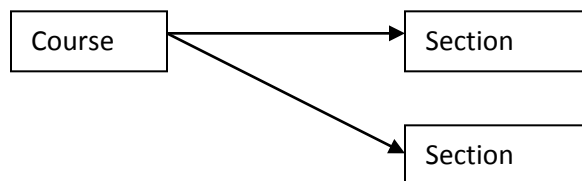
b. BOOK and BOOK COPY (books have copies)

Only a book can be made to many book copies so it is One To Many Relationships.



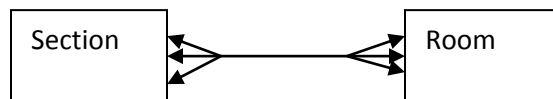
c. COURSE and SECTION (courses have sections)

Only courses can contains sections and the sections cannot have courses so it One To Many Relationships.



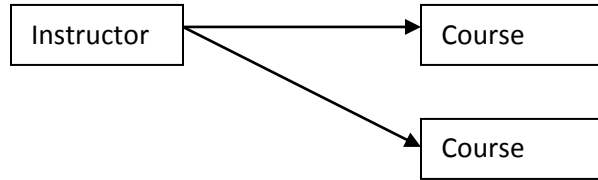
d. SECTION and ROOM (sections are scheduled in rooms)

Sections can accommodate many rooms and even a room can consist many sections so it is Many To Many Relationship.



e. INSTRUCTOR and COURSE

Instructor can take any number of courses but each course can be taught by a single instructor only so it is One To Many Relationship.



- 2) 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?

ANS) The data type which holds the data like video, audio, image has the capacity to store and retrieve. These are recently included in the DBMS and known as unstructured data types. These are not required for numeric and textual data.

- 3) 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.

ANS)

Name	Type	Length	Min	Max	Description	Source	Null	Schema	Field Order
Course	Alphanumeric	30			Course ID and name	Academic Unit	N	COE	1
Section	Integer	1	1	9	Section number	Registrar	N	COE	2
Semester	Alphanumeric	10			Semester and year	Registrar	N	NEU	3
Name	Alphanumeric	30			Student name	Student IS	Y	COE	4
ID	Integer	9			Student ID (SSN)	Student IS	Y	NEU	5
Major	Alphanumeric	4			Student major	Student IS	Y	COE	6
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit	N	COE	7

- 4) 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)?

ANS)

First: - It depends on the organization like if there is no much growth on data and with less functions then it may not require all-inclusive databases. But this might not fit for entire organization.

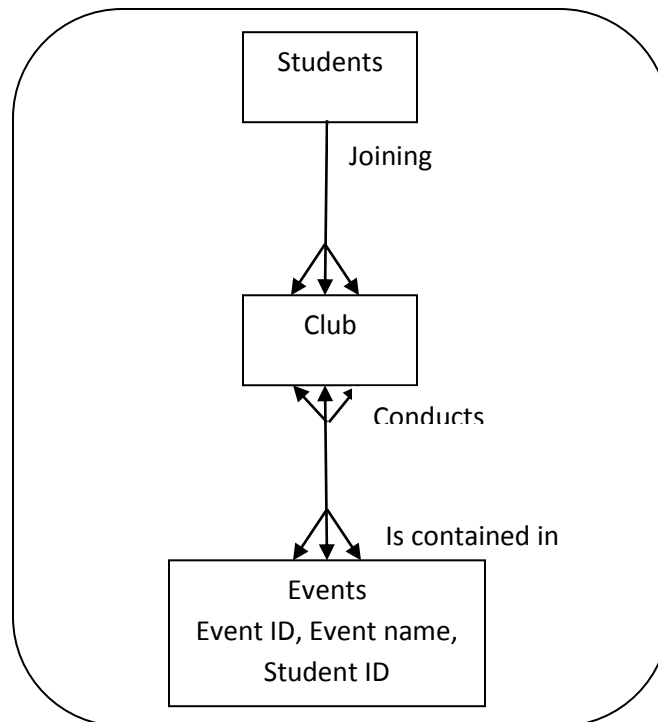
Second: - The person who holds the data ownership in an organization will decide whether to have multiple databases for maintaining the data high secure and performance.

- 5) Consider a student club or organization in which you are a member. What are the data entities of this enterprise? List and define each entity. Then, develop an enterprise data model (such as Figure 1-3a) showing these entities and important relationships between them.

ANS)

There are three entities

- 1) “Students” are joining the club
- 2) “Club” is an organization conducts the events for the students
- 3) “Events” are conducted for each activity for the students who perform it. Like dancing, signing, gaming etc., it contains entities like event id, event name and student’s id.



((PART 2-A))

Exercises

- 1) Compare this book to a database (assuming you don't just use it as a notebook, scribbling in the margins). What features does it provide? What features are missing?

ANS)

This book contains 2 chapters with overall of 47 pages. When comparing this to a database this book does not require more electricity, easy to carry, and does not need to boot before you can use it. We can see that the data is somewhat fine to access as it has a scrolling option which takes the user to the different pages by checking the page number through naked eyes. Even though there is a search option to find a particular text but without an index it make the job tough to find what is the content of this book. As far as the data is small it is better to access through paper or books but if the data is more, than this feature is not the best method to access the data, since database with good design has better access method to access the data. Moreover in this book we cannot add, insert, update or delete the data since it is not editable.

- 2) Describe two features that this book provides to help you look for particular pieces of data in different ways.

ANS)

The features are

- a. The Scroll Bar which takes the user top to bottom of the book in less than a second.
- b. The Find text (Ctrl+F) which searches the text which user requires.
- c. The Data is not editable so it is consistency.

- 3) What does CRUD stand for?What do the terms mean?

ANS)

CRUD stands for the four fundamental database operations they are Create, Read, Update and Delete. The each operation of this term are, is to create a new record by validating the new entry, updating the exiting record by validating the modified entry, deleting the exiting record by verifying that no user uses it and finally reading the record which finds the data you want in a quick speed.

- 4) How does a chalkboard implement the CRUD methods? How does a chalkboard's database features compare to those of this book?

ANS)

The chalkboard database performs all four fundamental operations of CRUD. Like we can create the data, update and delete the existing data and also we can read it. When comparing this chalkboard's database with this book we can see the following points.

- a. In chalkboard the data can be created easily but it is not possible in this book because it is not editable.
- b. In the book the data will not be lost as it is locked for CRUD method but in chalkboard the data can be lost easily just with hands.
- c. The book can be carried easily but it is not possible in case of chalkboard database.
- d. In both book and chalkboard they have their own storage limits so they cannot be expanded anymore.

- 5) Consider a recipe file that uses a single index card for each recipe with the cards stored alphabetically. How does that database's features compare to those of a book?

ANS)

The recipe file with index card will provide a good access method to fetch a data from it, so it is efficient when compared to this book. But if the data in the recipe file is more, then it will be hard to carry it anywhere since it is in physical form. The size and weight increases as the data increases in the recipe file, but the book is in PDF format so physically it will not grow.

- 6) What does ACID stand for? What do the terms mean?

ANS)

ACID stands for Atomicity, Consistency, Isolation and Durability.

- a) Atomicity is a kind of rule where all the operations in a transaction should be done or none should be done.
- b) Consistency makes the database in a consistent state before and after the transaction is made. If the operation in a transaction violates the database's rule then it will be completely rolled back.
- c) Isolation is like locking systems where only the transaction person will be accessing the records during the transaction.
- d) Durability means that once the transaction is made and the system got crashed down due to any reason, then when restarting the system, it will start where it ends.

- 7) Suppose Alice, Bob, and Cindy all have account balances of \$100 and the database does not allow an account's balance to ever drop below zero. Now consider three transactions: 1) Alice transfers \$125 to Bob, 2) Bob transfers \$150 to Cindy, and 3) Cindy transfers \$25 to Alice and \$50 to Bob. In what order(s) can the transactions be executed successfully?

ANS)

To execute the transaction successfully it can be done in either of the following order.

- (3) Cindy transfer 25\$ to Alice and 50\$ to Bob
- (1) Alice transfer 125\$ to Bob
- (2) Bob transfer 150\$ to Cindy

Or

- (3) Cindy transfer 25\$ to Alice and 50\$ to Bob
- (2) Bob transfer 150\$ to Cindy
- (1) Alice transfer 125\$ to Bob

- 8) Explain how a central database can protect your confidential data.

ANS)

The centralized database stores the user data from different locations and provides access to the user for their respective data. This will prevent the illegal user from performing something stupid or malicious. In addition to that, the user will be accessing the data remotely so there is no need for physical data; so even if the user loses or damages their computer, the data will be safe in the database's computer.

((PART 2-B))

Exercises

For the following scenarios, list the type(s) of database that might make good choices for storing the data.

- 1) A dog breeding database that records the ancestors of a single dog for five generations.

ANS)

Since the data is required to take the ancestors details of a single dog for five generation, there will not be more data development. Also it's the details of ancestors which is unchangeable so the data does not required to update. Under these criteria "Flat File Database" is the good choice for storing the data.

- 2) A similar dog breeding database that records the ancestors and descendants of a single dog for five generations each way.

ANS)

In this question, the data is need for both ancestors and descendants of a single dog, so it will form like a top down tree structure and they may be change in the records too. For this situation the "Hierarchical Database" would be a good choice for storing the data.

- 3) Application settings that record which windows a user had open and where they were positioned the last time the application was used.

ANS)

This requires tracking the details of the user's activities like window selection, when and where it is used. For this "Temporal database" would be the good choice because it can track the details over the time.

- 4) Total sales figures by month, arranged to make it easy to see trends graphically.

ANS)

To make graphical trends from the sales figure it is good to use "Spreadsheets Database" since it can represent the data in charts, graphs and other formats.

- 5) The same as Exercise 4 but the users want to be able to draw similar data for several product lines on the same graph.

ANS)

For this we can use "Dimensional Database" because it can represent the data in multi-dimensional view. Like not only for sales, it can represent the data for sales, product and year.