Lesson 8

- Database Design

Monish Singh, Adjunct Faculty of Information Systems

Databases

- A database is a group of logically related files that stores data, and the associations among them.
- It is simply a collection of organized information, usually as a set of related lists of similar entries. The data is often organized so that it is easily accessible.

Databases

- **BUT WHY CAN'T WE JUST USE A SPREADSHEET ?**

Advantages of using a Database

1. Long term data storage.
2. Multiple access to the data.
3. Protection against erroneous entries.
4. Protection against inadvertent corruption.
5. Minimizes data redundancy.

Who uses Databases ?

Databases Basics

- A database consists of a number of interrelated tables.
- Each table has a number of records which are used to represent real world objects.
  - For example Marist has a record for each of its student.
- Each record has a number of fields which are data items used to specify a characteristics of the record.
- Examples of the fields are:
  - CWID
  - Name
  - CourseId
  - Address

Keys

- Primary Key
  - Identifies a unique row in each table.
  - Example: Student New (CWID) is the primary key of Student table.
- Foreign Key
  - Key to a different table than the one in which they reside.
  - Example: CourseId in the Student table.

ACID Model

1. Atomicity: Each transaction should be “all or nothing”. If one part of the transaction fails, the entire transaction should fail.
2. Consistency: Only valid data should be written to the database. If data does not meet validation rules, transaction should be rolled back.
3. Isolation: It requires that multiple transactions occurring at the same time not impact each others execution.
4. Durability: Ensures that any transaction committed to the database will not be lost.

Normalization

- Normalization is the process of efficiently organizing data in a database.
- There are two goals of the normalization process.
  - Eliminating redundant data.
  - Ensuring (only storing related data in a table).

The Normal Forms

- Normal forms are a series of guidelines for ensuring that databases are normalized.
- These are referred to as normal forms and are numbered from one through five.
- In practical applications, you'll often see 1NF, 2 NF and 3NF along with 4NF.
- 5NF is very rarely seen.

First Normal Form (1NF)

1. Eliminate duplicative columns from the same table.
2. Identify each row with a unique column or set of columns (Primary Key).

Second Normal Form (2NF)

1. Meet all the requirements of 1NF.
2. Remove subsets of data that apply to multiple rows of a table and place them in separate tables.
3. Create relationships between these new tables and their predecessors through the use of foreign keys.

Third Normal Form (3NF)

1. Meet all the requirements of 2NF.
2. Remove columns that are not dependent upon the primary key.

Structured Query Language (SQL)

- SQL is a set of instructions used to interact with a relational database.

Major SQL commands

- CREATE
- INSERT
- SELECT
- DELETE

- http://www.w3schools.com/sql/sql_create_table.asp

Entity Relationship Diagrams

- ERD is a database design model that provides graphical representation of database tables, the columns in tables and the relationships between tables.
- ERD can provide sufficient information for database administrator to follow when developing and maintaining databases.

Entity Relationship Diagrams

- There are generally 3 ways to model entity relationship diagrams.
    1. Conceptual Model.
    2. Logical Model.
    3. Physical Model

Entity Relationship Diagrams – Data Models

Feature	Conceptual	Logical	Physical
Entity names	X	X	
Entity relationships	X	X	
Attributes		X	
Primary keys		X	X
Foreign keys		X	X
Table names			X
Column names			X
Column data types			X

Components of an ERD

- Entity – A person, place or thing about which we want to collect and store multiple instances of data. It has a name which is a noun and attributes that describes the data we are interested in storing.

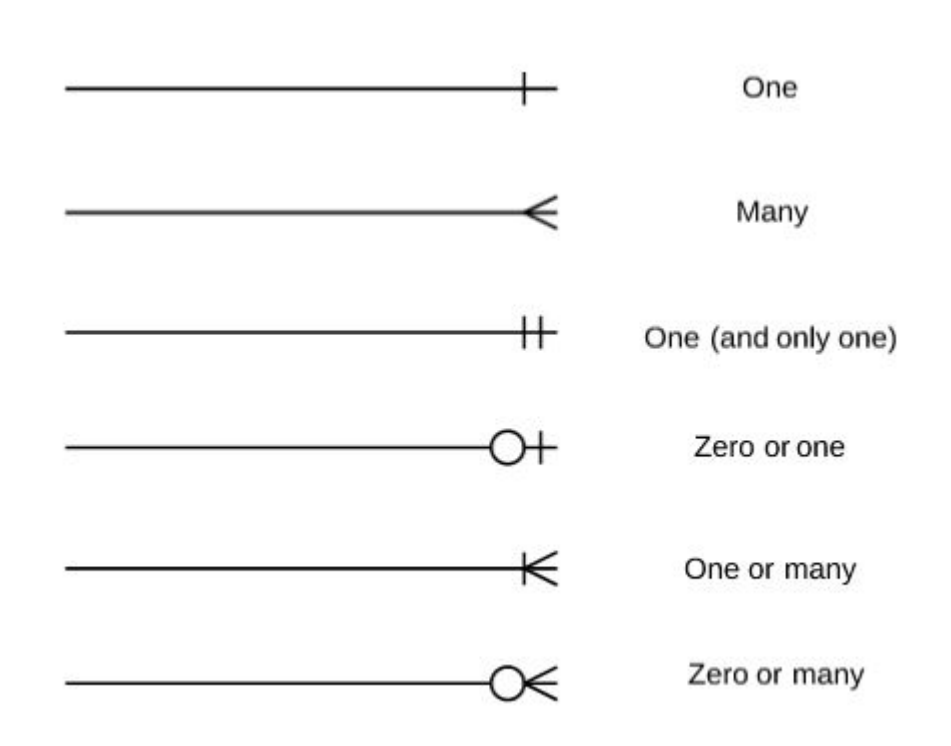
Components of an ERD

- Relationship – Relationships are meaningful associations between or among entities. They are usually verbs, e.g. assign, associates, track, reports, etc. In physical model, relationships are represented by lines.

Components of an ERD

- Cardinality – Cardinality refers to the maximum number of times an instance in one entity can be associated with the instances in the related entity. They are denoted using notations usually Crow's foot notation.

Crow's Foot Notation



BUT BEWARE OF MANY TO MANY RELATIONSHIPS

Issues with Many to Many

- Creates complex data set.
- Usually do not return correct results.
- Uses excessive computing resources.

Resolving Many to Many

- Break many to many relationships into two separate one-to-many relationships.

Book:

Dennis, A., Wixom, B., & Tegarden, D. (2012). Systems analysis design, UML version 2.0: An object oriented approach (4th ed.). Hoboken, NJ: John Wiley & Sons.