IS Course 4 Database Modeling

1- Database Modeling and Design Steps

- *Data modeling* is the *first step* of the database design process. High-level (entity-relationship)
  - This step is sometimes considered a high-level *abstract design* phase, also called *conceptual design*.
  - The objective of this phase is to describe:
  - The data contained in the database (e.g., entities: students, professors, courses, subjects)
  - The relationships between data elements (e.g., students are supervised by professors; professors teach courses)
  - The constraints on the data (e.g., the student ID has exactly eight digits; a subject has only four or six credits)
- *The second step* is to express the data elements, relationships, and constraints using the concepts provided by the high-level data model.
  - Because these concepts do not include implementation details, the result of the data modeling process is a (semi-)formal representation of the database structure.
  - This result is fairly easy to understand, so it is used as a reference to ensure that all user requirements are met.
- *The third step* is database design.
  - During this step, we can have two sub-steps:
  - One called logical database design, which defines a database within a specific DBMS data model (internal, all three types, including relational)
  - and another called physical database design, which defines the database's internal storage structure, file organization, or indexing techniques.
  - These two sub-steps are the steps for implementing the database and creating user operations/interfaces.

2- Defining a Data Model

- In the database design phase, data is represented using a specific data model.
- A data model is a collection of concepts or notations for describing data, relationships between data, data semantics, and data constraints.
- Most data models also include a set of basic operations for manipulating data in the database.

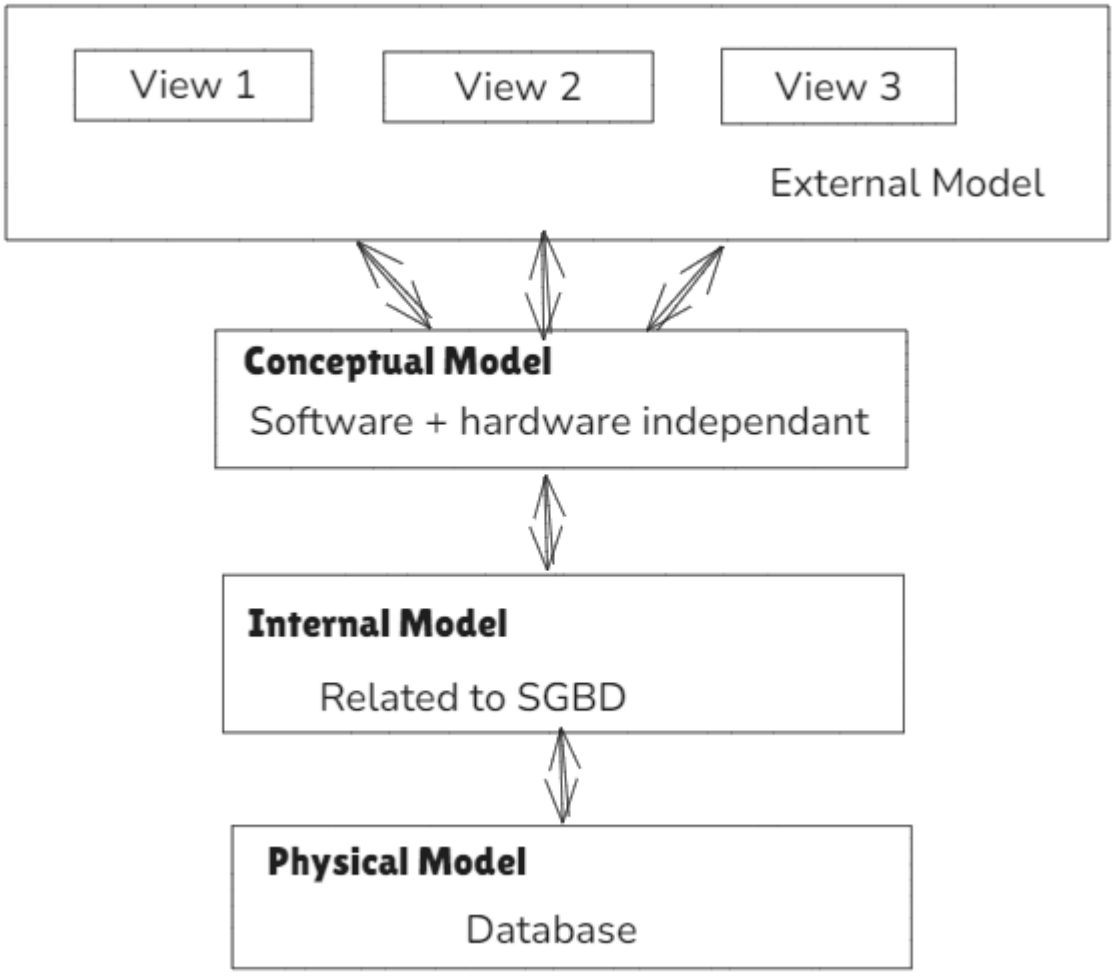
3- Degrees of Data Abstraction

There are four types of data models: external, conceptual, internal, and physical.

- External models represent the user's view of the database.
- Conceptual models provide flexible data structuring capabilities.
- Internal models view a database as a collection of fixed-size records (relational data model, network data model, hierarchical data model).
- Physical models: Represent the physical representation of the database. They have the lowest level of abstraction and describe how data is stored.

4- Data Abstraction Layer

- The data abstraction layer hides the complexity of data from database users.
- It provides a consistent interface to the data, regardless of how the data is physically stored and organized.
- The data abstraction layer allows application developers to focus on the business logic of their applications, rather than the details of data storage.



5- Types of Record-Based Logical Data Models

- These models provide concepts that users can understand, but they are not too far removed from how data is stored on a computer.

- Three well-known data models of this type are relational data models, network data models, and hierarchical data models.

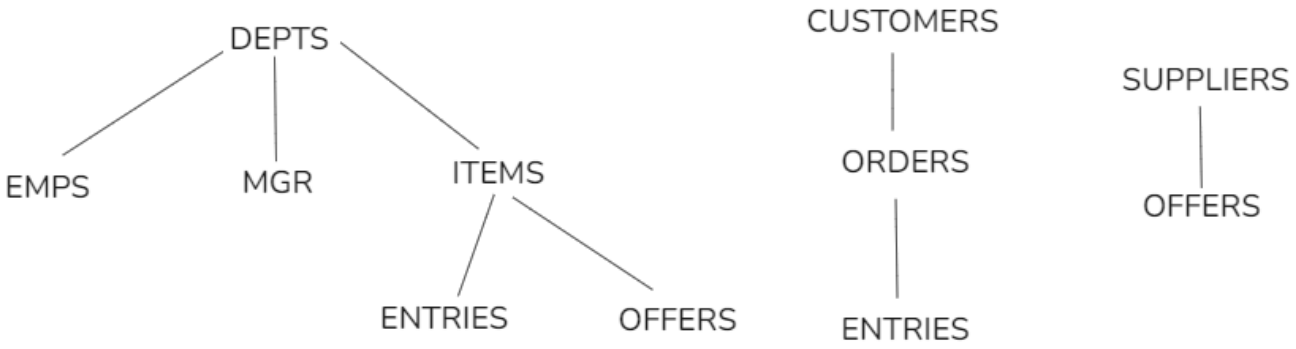
5.1 Relational Model

- This model represents data as relationships or tables.
- For example, in the Science World membership system, each membership has many members.
- The membership ID, expiration date, and address information are fields in the membership.
- Members are individuals such as Mickey, Minnie, Mighty, Door, Tom, King, Man, and Moose.
- Each record is said to be an instance of the membership table. - terms used: relation or table, fields-attributes, record: record (row, tuple) instance

customer_id	store_id	first_name	last_name	email
1	1	MARY	SMITH	MARY.SMITH@sakilacustomer.org
2	1	PATRICIA	JOHNSON	PATRICIA.JOHNSON@sakilacustomer.org
3	1	LINDA	WILLIAMS	LINDA.WILLIAMS@sakilacustomer.org
4	2	BARBARA	JONES	BARBARA.JONES@sakilacustomer.org
5	1	ELIZABETH	BROWN	ELIZABETH.BROWN@sakilacustomer.org
6	2	JENNIFER	DAVIS	JENNIFER.DAVIS@sakilacustomer.org
7	1	MARIA	MILLER	MARIA.MILLER@sakilacustomer.org

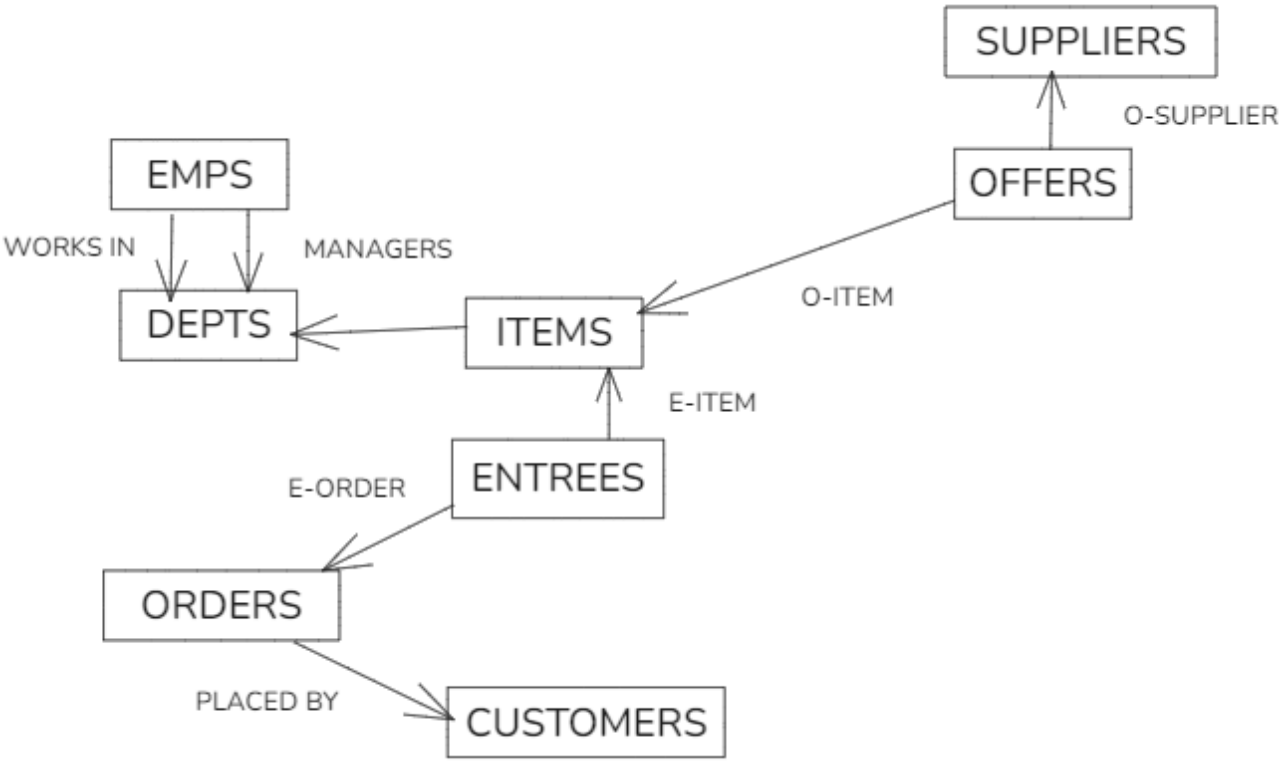
5.2 Hierarchical Model

- This model represents data in a hierarchical tree structure.
- Each branch of the hierarchy represents a number of related records.



5.3 Network Model

- This model represents data as record types.
- This model also represents a limited type of one-to-many relationship called a set type
- Terms used: record, set



6- The Relational Data Model

The relational data model was introduced by E. F. Codd in 1970. It is currently the most widely used data model.

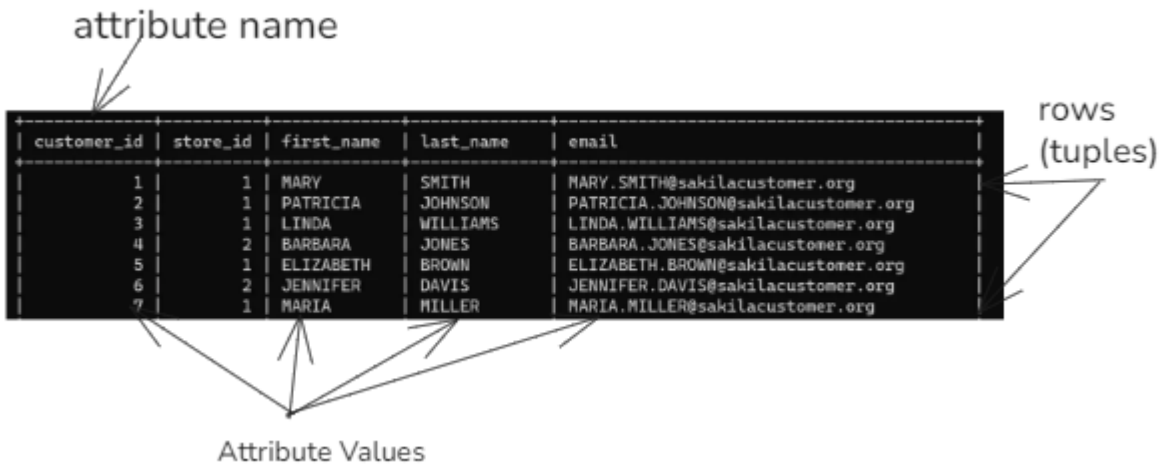
- The relational model has provided the basis for:
- Research on data/relationship/constraint theory
- Many database design methodologies
- The standard database access language called Structured Query Language (SQL)
- Nearly all modern commercial database management systems
- The relational data model describes the world as "a set of interdependent relationships (or tables)."

6-1 Fundamental Concepts of the Relational Data Model

- **Relation = Table**
  - A relation, also called a table or file, is a subset of the Cartesian product of a named list of domains.
  - In a table, each row represents a group of related data values. - A row, or record, is also called a tuple.
  - A column in a table is a field and is also called an attribute.
  - *Formal definition:*

$$\begin{aligned}&\text{- Given } n \text{ domains}\\&\text{- And } r \text{ a relation defined on these domains}\\&\text{- Then: } r \subseteq D_1 \times D_2 \times \ldots D_n\end{aligned}$$

- **Table**:
  - A database is compsed of multiple tables, and each table holds the data
- **Column**:
  - The main storage unit in a database is called a column, field, or attribute.
  - They house the basic data components into which your content can be decomposed.
  - When deciding which fields to create, you can, for example:
  - Select the common components of the information you will store in the database
  - Avoid details that distinguish one element from another.
- **Domain**
  - A domain is the original set of atomic values used to model data.
  - An atomic value means that each value in the domain is indivisible with respect to the relational model.
  - Examples: The First Name domain is the set of strings that represent people's names.
  - The Marital Status domain offers a set of possibilities: Married, Single, Divorced.
- **Records**
  - Records contain fields that are related, such as a customer or an employee.
  - As mentioned earlier, a tuple is another term used for record. (or a line)



- **Degree**
  - The degree is the number of attributes in a table. In the example above, the degree is 5.

6-2 Table Properties

- A table has a name that is distinct from all other tables in the database.
- There are no duplicate rows; each row is distinct.
- Column entries are atomic. The table does not contain repeating groups or multi-valued attributes.
- Column entries belong to the same domain based on their data type, including:

- number (numeric, integer, float, smallint, etc.)
- character (string)
- date
- logical (true or false)
- Operations combining different data types are not allowed.
- Each attribute has a distinct name.
- The sequence of columns is not significant.
- The sequence of rows is not significant.

Bibliographic References

Watt, Adrienne, and Nelson Eng. *Database design*. 2nd Edition. BCcampus, 2014