**Topic 3: Data Models and Conceptual Modeling**

>A data model: is an integrated collection of concepts for describing and manipulating data, relationships between data, and constraints on the data in an organization.

>A model is a representation of ‘real world’ objects and events, and their associations.

- Basic concepts and notations for db designers and end users.

**A data model 3 components:**

* **Structural part:** sets of rules according to which databases can be constructed.
* **Manipulative part:** types of operation that will be allowed on data.
* **Set of integrity constraints:** ensures if data is accurate**.**

**Data model board categories:**

* **Object based** (describes data at the conceptual and external level)
* **Record based** (describes data at the conceptual and external level)
* **Physical based** (describes internal)

1. **Object Based Data Model**

* Uses concepts such as entities, attributes and relationships.
* **Entity :** is a distinct object (a person, place, thing, concept, event) in the organization that is to be represented in the database.
* An attribute is a property that describes some aspect of the object that we wish to record.
* **Relationship** is an association between entities.

**Types** of object-based data model are :-

o Entity–Relationship

o Semantic

o Functional

o Object-Oriented.

* The object-oriented data model extends the definition of an entity to include not only the attributes that describe the state of the object but also the actions that are associated with the object, that is, its behavior.
* The object is said to encapsulate both state and behavior.

1. **Record Based Data Models**

- defines a fixed number of fields, each typically of a fixed length.

There are 3 principal types of record-based logical data model:

* the relational data model,
* the network data model, and
* hierarchical data model

1. **Physical Data Model**

* Physical data models describe how data is stored in the computer.

1. **Conceptual Modeling**

* ‘Heart’ of the database.
* It supports all the external views and is, in turn, supported by the internal schema.
* **Definition:** is the process of constructing a model of the information use in an enterprise that is independent of implementation details, such as the target DBMS, application programs, programming languages, or any other physical considerations.

1. **Hierarchical data model**

* is a restricted type of network model. Again, data is represented as collections of records and relationships are represented by sets.
* allows a node to have only one parent.
* can be represented as a tree graph, with records appearing as nodes.
* Its basic logical structure is represented by an upside-down tree. The hierarchical structure contains levels, or segments.
* A segment is the equivalent of a file system’s record type.
* The hierarchical model depicts a set of one-to-many (1:M) relationships between a parent and its children segments. (Each parent can have many children, but each child has only one parent.

**Disadvantages:**

They do not provide adequate facilities for explicitly specifying constraints on the data, whereas the object-based data models lack the means of logical structure specification but provide more semantic substance by allowing the user to specify constraints on the data.

1. **Network data model**

* represented as collections of **records**, and relationships are represented by **sets** (which become pointers in the implementation.)
* The records are organized as generalized graph structures with records appearing as **nodes** (also called **segments**) and sets as **edges** in the graph
* allows a record to have more than one parent.
* Not used today

Some important concepts that were defined at this time are:

* The **schema**, which is the conceptual organization of the entire database.
* The **subschema**, which defines the portion of the database “seen” by the application programs that actually produce the desired information from the data contained within the database.
* **A data management language (DML)**, which defines the environment in which data can be managed and to work with the data in the database.
* A schema **data definition language (DDL),** which enables the database administrator to define the schema components
* relational systems adopt a declarative approach to database processing (that is, they specify what data is to be retrieved), network and hierarchical systems adopt a navigational approach (that is, they specify how the

data is to be retrieved)

1. **Relational data model**

* data and relationships are represented as tables, each of which has a number of columns with a unique name.
* database be perceived by the user as tables.
* However, this perception applies only to the logical structure of the database, that is, the external and conceptual levels of the ANSI-SPARC architecture.
* It does not apply to the physical structure of the database, which can be implemented using a variety of storage structures.

**The Entity Relationship Model**

* designers prefer to use a graphical tool in which entities and their relationships are pictured. Thus, the entity relationship (ER) model, or ERM, has become a widely accepted standard for data modeling.
* The relational data model **+** ERM = tightly structured database design.
* ER models are normally represented in an entity relationship diagram (ERD), which uses graphical representations to muodel database components.

**The Object-Oriented (OO) Model**

* a data model that more closely represented the real world.
* both data and their relationships are contained in a single structure known as an object.
* a semantic data model because semantic indicates meaning.
* But quite unlike an entity, an object includes information about relationships between the facts within the object, as well as information about its relationships with other objects. Therefore, the facts within the object are given greater meaning.
* Subsequent OODM development has allowed an object to also contain all operations that can be performed on it, such as changing its data values.
* Object-oriented data models are typically depicted using Unified Modeling Language (UML) class diagrams.
* Unified Modeling Language (UML) is a language based on OO concepts that describes a set of diagrams and symbols that can be used to graphically model a system.
* UML class diagrams are used to represent data and their relationships within the larger UML object-oriented system’s modeling language.

**Object/Relational and XML**

* relational model’s main vendors evolved the model further and created the extended relational data model (ERDM).
* The ERDM gave birth to a new generation of relational databases supporting OO features such as objects (encapsulated data and methods)
* That’s why a DBMS based on the ERDM is often described as an object/relational database management system (O/R DBMS).
* Extensible Markup Language (XML) emerged as the de facto standard for the efficient and effective exchange of structured, semistructured, and unstructured data.
* To address this need, XML databases emerged to manage unstructured data within a native XML format.



