Database Modeling (Part 2)

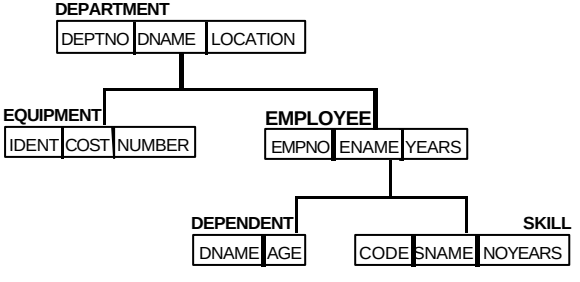
Hierarchical Data Model

- The first important logical database model.
- Primarily implemented on mainframe today.
- Records are arranged in a top-down structure that resembles an upside-down tree.
- The parent and child are often used in describing hierarchical model.
- An important characteristic is that a child is related to only one parent.
- The leading hierarchical DBMS in use today is IBM's Information Management System (IMS).

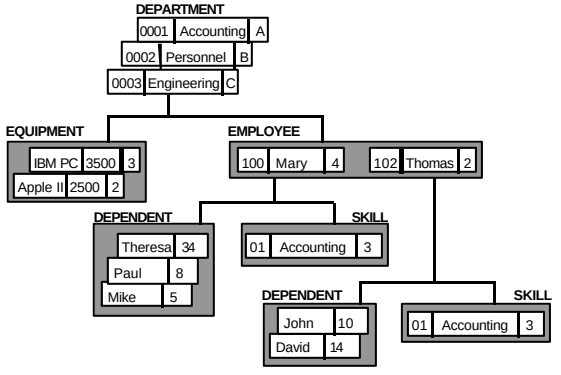
IMS Physical Databases

- The physical database record is a basic building block in IMS.
- A Physical Database Record (PDBR) consists of a set of related fields.
- A PDBR consists of a root segment and its subordinate segments called child segments.

IMS Physical Databases



PDBR Occurrences

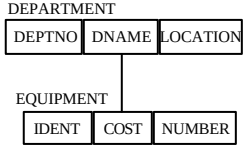


IMS Logical Database

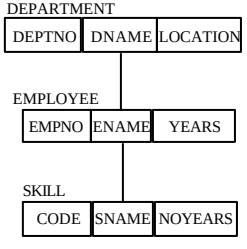
- External views of individual users in IMS are reflected in Logical Database Records (LDBR).
- Each LDBR type is a subset of a corresponding PDBR type.
- Any segment type (except the root segment) of a PDBR may be omitted from an LDBR.
- Any field types that occur in a PDBR may be omitted in the corresponding LDBR.

Examples of LDBR

(a) Equipment LDBR



(b) Personnel LDBR

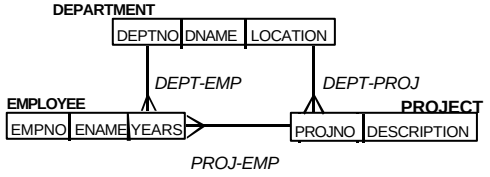


Network Data Model

- There is no distinction between parent and child record types.
- Any record type may be associated with an number of different record types.
- A set is the definition of a directed relationship from an owner record type to one or more member record types..

Network Data Model

- Generally, we can assume that set is implemented as a ring data structures with the owner at the head of the chain and with the last member pointing to the owner.



Network Data Model

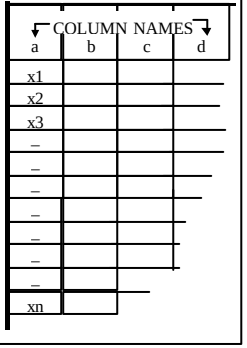
- The Conference on Data System Languages (CODASYL) through its Data Base Task Group (DBTG) is a standard organization that has developed and issued descriptions of language for defining and processing data in Network DBMS.
- IDMS (Integrated Database Management System) is the leading DBTG DBMS on IBM computers.

Relational Data Model

- A data model that represents data in the form of tables or relation.
- The relational database model consists of the following three components:
  - **Data Structure** : Data are organized in the form of tables or relations.
  - **Data Manipulation:** Powerful operations such as SQL languages or Query-by-example, are used to manipulate data stored in the database.
  - **Data Integrity:** Business rules are specified to maintain the integrity of data when they are manipulated.

Physical Properties

- A relation consists of 1 or more columns and 0 or more rows.
- A row is called a tuple.
- Each relation is given a unique name.
- Each column has a name unique within the relation.
- Each row contains an instance of the data associated with the relation.
- A relation with no rows is empty (contains no data), but still exists.



Logical Properties

- Ordering of columns
- Ordering of rows
- Uniqueness
- The sequence of columns (Left to right) is significant
- The sequence of rows (Top to bottom) is significant

Ordering of Columns

- Columns are unordered, left to right.
- This property is designed to preserve the independence of each column.

Ordering of Rows

- Rows are unordered, top to bottom.
- This is designed to preserve the independence of each row.

Uniqueness

- No row may be duplicated in a given relation. Uniqueness in a relation is guaranteed by the designation of a primary key for each relation.
- A candidate key in a relation is an attribute that uniquely identifies a row in that relation.
- A primary key is a candidate key that has been selected to be the unique identifier for each row.
- Primary key values cannot be null, since they would then not identify a row.

The Sequence of Columns (Left to right) is Significant

- The columns of a relation can be interchanged without changing the meaning or use of the relation.
- There is no hidden meaning implied by the ordering.

The Sequence of Rows (Top to bottom) is Significant

- The rows of a relation may be interchanged or stored in any sequences.
- It makes no differences as whether to insert a new row in front or at the end of the table.

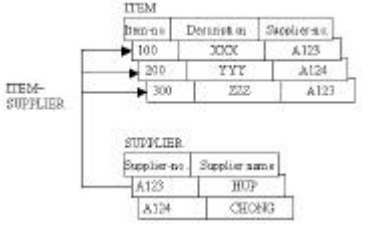
Relational vs. Network Models

- In relational model, connections between two relations are represented by including two attributes with the same domain – one in each of relations (Foreign Key).
- In network model, 1:N connections between two record types are explicitly represented by the set type construct.

Relational vs. Network Models

Item No	Description	Supplier-no.
100	XXX	A123
200	YYY	A124

Supplier-no.	Supplier name
A123	HUP
A124	CHONG



Relational vs. Network Models

- In relational model, individual tuples that have the same value for that attributes are logically related, even though they are not physically connected together.
- In network model, the DBMS connects related records together in a set instance by some physical method. Records are physical connected together when they participate in the same set instance.
- A set type physically represents a logical 1:N relationship type.

Hierarchical vs. Network Models

- Both represent relationship explicitly.
- A record type in the network model can be a member in any number of set types.
- In hierarchical, a record type can have one real parent. This creates problems when modeling M:N and n-ary relationship types.
- If many relationship types exists, we will have to duplicate records and pointers to design a hierarchical representation.

Hierarchical vs. Network vs. Relational Models

- The hierarchical model is considered inferior to both the relational and network models as for as modeling capability is concerned.

Strengths of OO Data Model

- A closer representation of real-world problem domains and has a greater productivity in applications productivity.
- It has ability to model complex data types such as images and documents.

Weakness of OO Data Model

- Lack of accepted standards
  - There are no initial standard at the national and international level yet.
- Lack of development tools
  - Tools such as CASE and 4GL are still under development, hence but not widely available.
- Performance
  - The performance of ODBMS technology with large numbers of concurrent users and frequent transactions has yet been tested or demonstrated.

Weakness of OO Data Model

- Data Management facilities
  - Some of the products do not have adequate facilities for concurrency control, backup and recovery.
- Query languages
  - Users cannot retrieve data about one or more objects based on his own defined criteria.

Review Questions

- Compare and contrast the following:
  - Inheritance vs generalization hierarchy
  - Generalization vs specialization
  - Candidate key vs primary key
  - Subtype vs super type
  - Physical database record vs logical database record
  - Encapsulation vs inheritance
  - Relational model vs network model
  - Hierarchical model vs network model
- What are the limitations of ODBMS technology?

Review Questions

- Draw an E-R and OO diagram for each of the following situations:
  - A company has a number of employees. The attributes of EMPLOYEE include NAME, ADDRESS, BIRTHDATE and DATEHIRED. One method that is required of all employees is claculateYesrsOfSevice. The company also has several projects. Attributes of PROJECT include CODE, DESCRIPTION and START DATE. Each employee may be assigned to one or more projects, or maybe assigned to a project. A project must have at least one employee assigned, and may not be assigned to a project. A project must have at least one employee assigned, and may have several employee assigned. One method required of all projects is CalculateTotalCostToDate.
  - In a vehicle-licensing application, there are three types of vehicle: passenger, truck, and trailer. Vehicle ID is an attribute of all vehicle types. Truck and trailer vehicles have an attribute named GROSS CAPACITY. The passenger and truck vehicle types required of all courses is ChangeCourseDescription.