

Chapter 11

Database Concepts

Database Concepts

INTRODUCTION

Database is collection of interrelated data and database system is basically a computer based record keeping system.

It contains the information about one particular enterprise. It maintains any information that may be necessary to the decision making process involved in the management of that organisation.

Database Concepts

WHY DATABASE?

A typical file processing system suffers from some major limitations like,

- ✓Data Redundancy
- ✓Data Inconsistency
- ✓Un-sharable Data
- ✓Un-standardized Data
- ✓Insecure Data
- ✓Incorrect Data

On the other hand, a database system overcomes all these limitations and ensures continues efficiency.

Database Concepts

ADVANTAGES OF DATABASE

The advantages provided by a database system are :

1. Reduced redundancy
2. Controlled data inconsistency
3. Shared Data
4. Standardized Data
5. Secured Data
6. Integrated Data

Database Concepts

ADVANTAGES OF DATABASE

1. Reduced redundancy:

Database redundancy means duplication of data. Non-database systems maintain separate copy of data for each application.

For example:

In college, student records are maintained and hostel also maintains the student records for those students who live in hostel. Though the records of hosteller students are already being maintained by the college, the hostel keeps separate copy of it. But this duplication of data leads to inconstancy or incorrect data.

Database Concepts

ADVANTAGES OF DATABASE

2. Controlled Data Inconsistency:

When the redundancy is not controlled, there may be occasions on which two entries about the same data do not agree (that is one of them stores updated information and the other does not.) at such times, database is said to be inconsistency.

By controlling redundancy, the inconsistency is also controlled. The database ensures any change is made to either of the two entries is automatically made to the other. This process is known as propagating update.

Database Concepts

ADVANTAGES OF DATABASE

3. Shared Data

Sharing of data mean that individual piece of data in the database may be shared among several different users, in the sense that each of those users may have access to the same piece of data and each of them may use it for different purposes.

Database Concepts

ADVANTAGES OF DATABASE

4. Standardized Data

The database management systems can ensure that all the data (that is stored centrally) follow the applicable standards. There may be certain standards laid by the company or organization using the database.

Similarly, there may be national and or international standards. Standardizing stored data formats is particularly desirable as an aid to data interchange or migration between systems.

Database Concepts

ADVANTAGES OF DATABASE

5. Secured Data

A database management system ensures data security and privacy by ensuring that the only means of access is through proper channel and also by carrying out authorization checks whenever access to sensitive data is attempted.

Data Security: Data security refers to protection of data against accidental or intentional disclosure to unauthorized persons, or unauthorized modification or destructions.

Privacy of Data: it refers to the rights of individuals and organizations to determine for themselves when, how and what extent information about them is to be transmitted to others

Database Concepts

ADVANTAGES OF DATABASE

6. Integrated Data

When database contains data employed by many different users it is important that association between data items not to be destroyed. Hardware failures and various types of accidents will occur occasionally. The storage of data and its updation, and insertion procedures defined by the database, are such that the system can easily recover from these circumstances without harm to the data.

The database management system designs certain integrity checks to ensure that data values confirm to some specified rules. *For example date cant be like 25/25/12; it is invalid date.*

Database Concepts

DATABASE ABSTRACTION

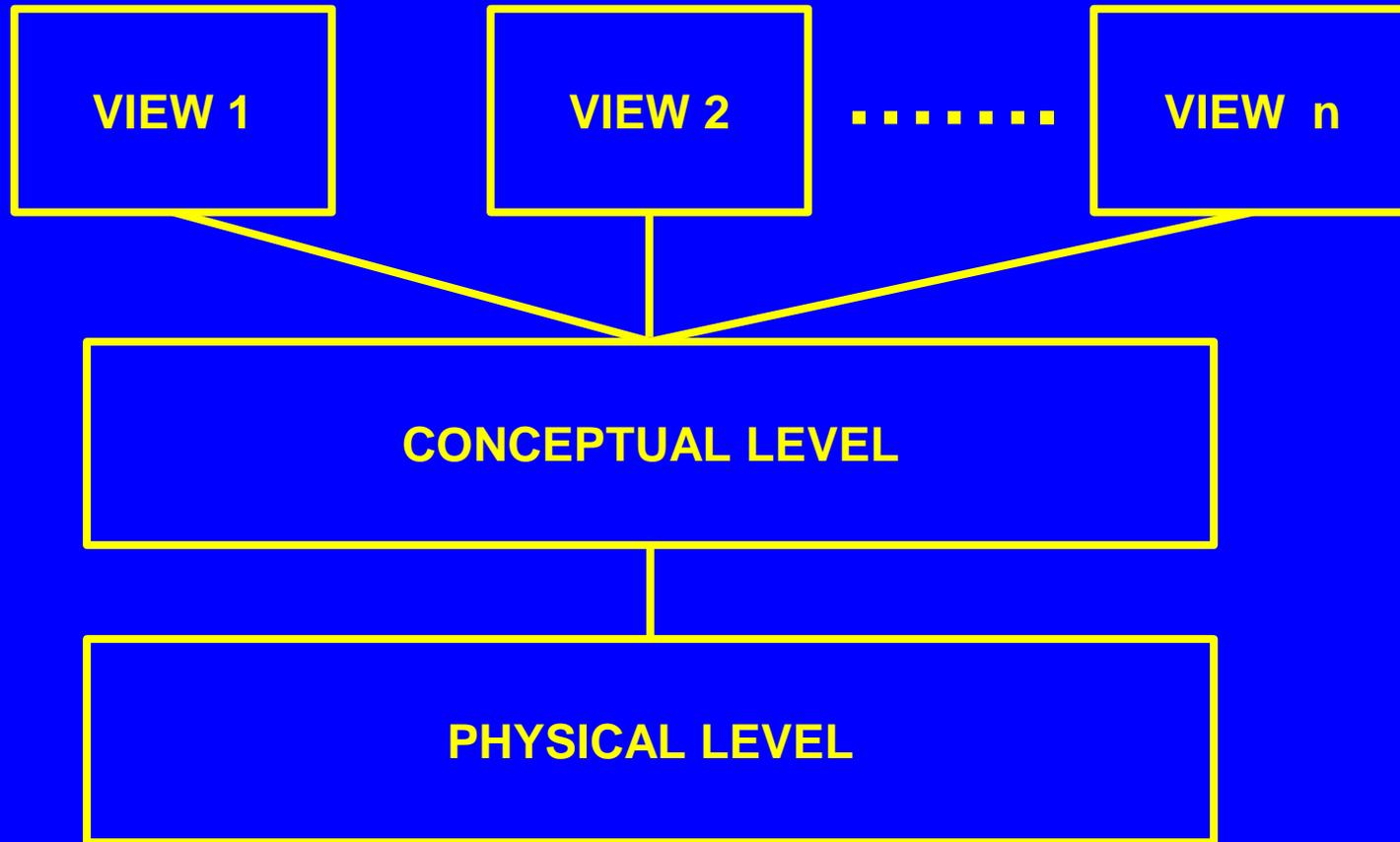
Data abstraction simplifies database design. The major purpose of a database system is to provide users with an **abstract view** of the system. The system hides certain details of how data is stored and created and maintained. Complexity should be hidden from database users.

Three Levels of Abstractions are,

1. Physical Level (Internal Level)
2. Conceptual Level
3. External Level (View Level)

Database Concepts

THREE LEVELS OF ABSTRACTION



Database Concepts

1. PHYSICAL LEVEL (INTERNAL OR LOW LEVEL)

It defines how data is stored. It is very complex and used by developers. & it deals with,

- ✓ How the data are stored.
- ✓ E.g. index, B-tree, hashing.
- ✓ Interface between OS and record structure.
- ✓ Lowest level of abstraction.
- ✓ Complex low-level structures described in detail.

Database Concepts

1. CONCEPTUAL LEVEL

It defines data in terms of a data model. It tells what data is stored and described in small numbers. This level is usually used by **DBA** (Database administrator) & it deals with,

- ✓ Next highest level of abstraction.
- ✓ Describes *what* data are stored.
- ✓ Describes the relationships among data.
- ✓ Database administrator level.

Database Concepts

1. EXTERNAL LEVEL OR VIEW LEVEL

It defines a number of simplified domain-specific views. It describes only part of databases. This level is used by users.

✓ Highest level.

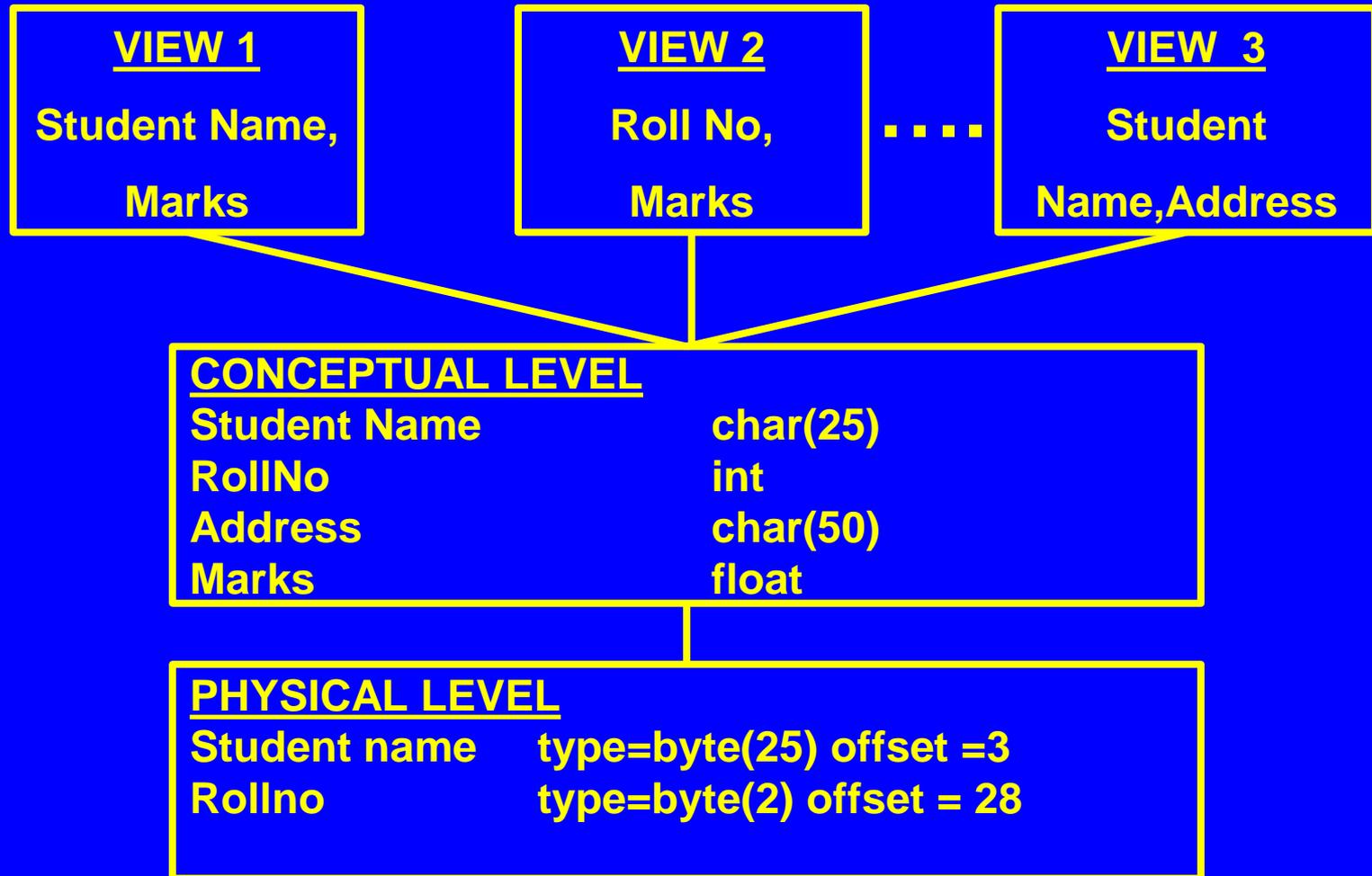
✓ Describes *part* of the database for a particular group of users.

✓ Can be many different views of a database.

✓ E.g. In a School get a view of Student details, but not of payroll data of employees or teachers of the school.

Database Concepts

EXAMPLE OF THREE LEVELS OF ABSTRACTION



Database Concepts

DATA INDEPENDENCE

The ability to modify a scheme definition in one level without affecting a scheme definition in the next higher level is called **DATA INDEPENDENCE**

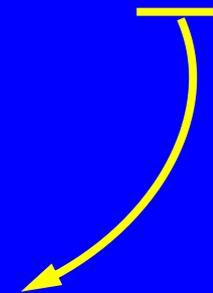
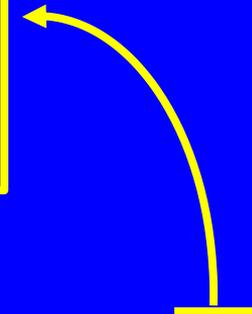
A major objective for three-level architecture is to provide data independence, which means that upper levels are unaffected by changes in lower levels.

There are two kinds of data independence:

- ✓ Logical data independence
- ✓ Physical data independence

Database Concepts

DATA INDEPENDENCE



Database Concepts

DATA INDEPENDENCE

SCHEMA

schema is nothing but definition of objects/tables in a database.

A schema is the set of metadata (data dictionary) used by the database, typically generated using DDL. It defines attributes of the database, such as tables, columns, and properties. A database schema is a description of the data in a database.

One or more schemas can reside on the same database.

Database Concepts

DATA INDEPENDENCE

SCHEMA

A **logical schema** won't exist in your database. A logical schema is a design-centric database structure built to meet your business requirements. It is a model that exists on a white board or in a diagramming tool. It is like the architect's drawings of your database.

Logical schema is made out of actually what you plan(your initial data model) and it's objects are:

- ✓ Tablespaces
- ✓ Tables
- ✓ Views
- ✓ procedures

Database Concepts

DATA INDEPENDENCE

SCHEMA

A **Physical schema** is a term used in data management to describe how data is to be represented and stored (files, indices, *et al.*) in secondary storage using a particular database management system (DBMS) (e.g., Oracle RDBMS, Sybase SQL Server, etc.).

A Physical schema is one where you have your data files, redo logs, control files etc, physical objects that reside on a O.S. and are in row state. These objects are used by logical structures to make our data make sense and readable.

Database Concepts

DATA INDEPENDENCE

LOGICAL DATA INDEPENDENCE

Logical data independence indicates that the conceptual schema can be changed without affecting the existing external schemas.

Logical data is data about database, that is, it stores information about how data is managed inside. For example, a table (relation) stored in the database and all constraints, which are applied on that relation.

Logical data independence is a kind of mechanism, which liberalizes itself from actual data stored on the disk. If we do some changes on table format it should not change the data residing on disk.

Database Concepts

DATA INDEPENDENCE

PHYSICAL DATA INDEPENDENCE

Physical data independence indicates that the physical storage structures or devices could be changed without affecting conceptual schema.

All schemas are logical and actual data is stored in bit format on the disk. Physical data independence is the power to change the physical data without impacting the schema or logical data.

For example, in case we want to change or upgrade the storage system itself, that is, using SSD instead of Hard-disks should not have any impact on logical data or schemas.

Database Concepts

DATA MODELS

A database model is a specification describing how a database is structured and used.

A **data model** organizes data elements and standardizes how the data elements relate to one another.

Data model tells how the logical structure of a database is modelled. Data Models are fundamental entities to introduce abstraction in DBMS. Data models define how data is connected to each other and how it will be processed and stored inside the system.

Database Concepts

DATA MODELS

There are three types of Data Models,

- ✓ Relational Model
- ✓ Network Model
- ✓ Hierarchical Model

Database Concepts

DATA MODELS

RELATIONAL MODEL

Relational data model is the primary data model, which is used widely around the world for data storage and processing. This model is simple and have all the properties and capabilities required to process data with storage efficiency.

In the relational model, all data must be stored in relations (tables), and each relation consists of rows and columns. Each relation must have a header and body. The header is simply the list of columns in the relation. The body is the set of data that actually populates the relation, organized into rows. You can extrapolate (extend) that the junction of one column and one row will result in a unique value - this value is called a tuple.

Database Concepts

DATA MODELS

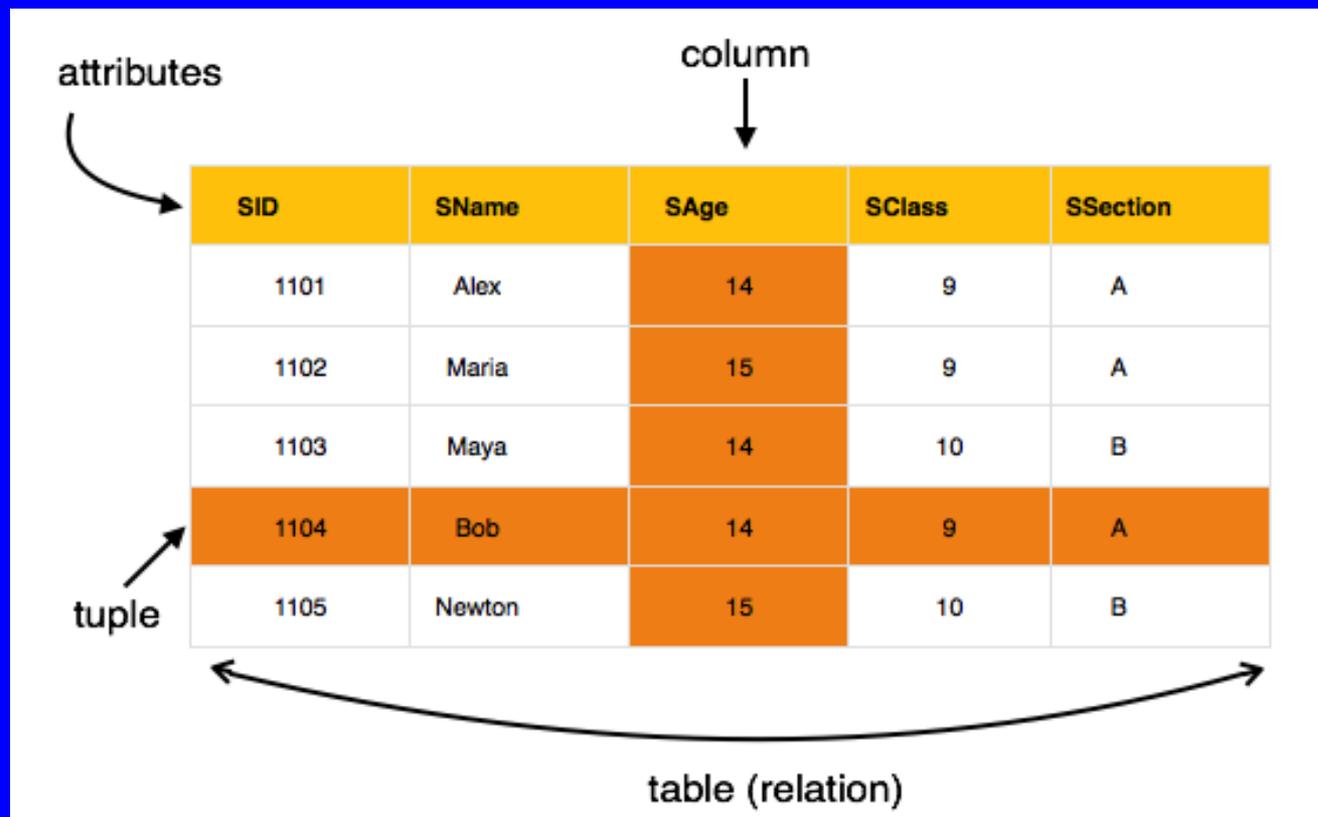
RELATIONAL MODEL

The second major characteristic of the relational model is the usage of keys. These are specially designated columns within a relation, used to order data or relate data to other relations. One of the most important keys is the primary key, which is used to uniquely identify each row of data. To make querying for data easier, most relational databases go further and physically order the data by the primary key. Foreign keys relate data in one relation to the primary key of another relation.

Database Concepts

DATA MODELS

RELATIONAL MODEL – AN EXAMPLE



Database Concepts

DATA MODELS

RELATIONAL MODEL - CONCEPT

Tables: In relation data model, relations are saved in the format of Tables. This format stores the relation among entities. A table has rows and columns, where rows represent records and columns represents the attributes.

Tuple: A single row of a table, which contains a single record for that relation is called a tuple.

Attribute : columns are referred as attributes.

Domain: A pool of values from which the actual values appearing in a given column are drawn.

Database Concepts

DATA MODELS

RELATIONAL MODEL - CONCEPT

What does *Primary Key* mean?

A primary key is a special relational database table column (or combination of columns) designated to uniquely identify all table records.

A primary key's main features are:

- ✓ It must contain a unique value for each row of data.
- ✓ It cannot contain null values.

A primary key is either an existing table column or a column that is specifically generated by the database according to a defined sequence.

Database Concepts

DATA MODELS

RELATIONAL MODEL - CONCEPT

PRIMARY KEY - TABLE - CUSTOMERS

Customer ID	Name	Company	Phone Number	E-mail Address
6273	Vedat Diker	CLIS/UMD	(301) 405 9814	vedat@umd.edu
3245	Bugs Bunny	Acme, Inc.	(123) 555 9876	bugs@acme.com
1324	Will E. Coyote	Acme, Inc.	(123) 555 9821	will@acme.com

primary key field

Primary key is a unique identifier of records in a table.

Primary key values may be generated manually or automatically.

Database Concepts

DATA MODELS

RELATIONAL MODEL - CONCEPT

What does *Foreign Key* mean?

A foreign key (**REFERENTIAL INTEGRITY**) is a column or group of columns in a relational database table that provides a link between data in two tables. It acts as a cross-reference between tables because it references the primary key of another table, thereby establishing a link between them.

The majority of tables in a relational database system adhere to the foreign key concept. In complex databases and data warehouses, data in a domain must be added across multiple tables, thus maintaining a relationship between them. The concept of referential integrity is derived from foreign key theory. Foreign keys and their implementation are more complex than primary keys.

FOREIGN KEY

Primary key field

Parent table

Directors

Director ID	Name	Date of Birth	Place of Birth	Biography
785	John Frankenheimer	19-Feb-30	New York, NY	Born in New York and raised in Queens, ...
235	Ridley Scott	30-Nov-37	South Shields, UK	Education: Royal College of Art, London...
976	James Foley	28-Dec-53	Brooklyn, NY	Attended the USC Film School...

Relationship

Movies

Child table

Movie ID	Title	Director ID	Genre	...
4532	Gladiator	235	Action	
8357	Swwet and Lowdown	497	Comedy	
7465	Confidence	976	Drama	

Foreign key field

Database Concepts

DATA MODELS

RELATIONAL MODEL - CONCEPT

What does *Candidate Key* mean?

All the attribute combinations inside a relation that can serve primary key are candidate keys as they are candidates for the primary key position.

What does *Alternate Key* mean?

A candidate key that is not the primary key is called alternate key.

Database Concepts

DATA MODELS

NETWORK MODEL – CONCEPT

Developed in mid 1960s as part of work of CODASYL (Conference on Data Systems Languages) which proposed programming language COBOL (1966) and then network model (1971). Objective of network model is to separate data structure from physical storage, eliminate unnecessary duplication of data with associated errors and costs.

- ✓ uses concept of a data definition language, data manipulation language.
- ✓ uses concept of m:n linkages or relationships.
 - an owner record can have many member records.
 - a member record can have several owners.

Database Concepts

DATA MODELS

NETWORK MODEL – CONCEPT

Example of a Network Database

- ✓ a hospital database has three record types:
 - patient: name, date of admission, etc.
 - doctor: name, etc.
 - ward: number of beds, name of staff nurse, etc.
- ✓ Need to link patients to doctor, also to ward
- ✓ Doctor record can own many patient records
- ✓ Patient record can be owned by both doctor and ward records.
- ✓ Network DBMSs include methods for building and redefining linkages, e.g. when patient is assigned to ward.

Database Concepts

DATA MODELS

NETWORK MODEL – CONCEPT

Restrictions

Links between records of the same type are not allowed while a record can be owned by several records of different types, it cannot be owned by more than one record of the same type (patient can have only one doctor, only one ward).

Summary

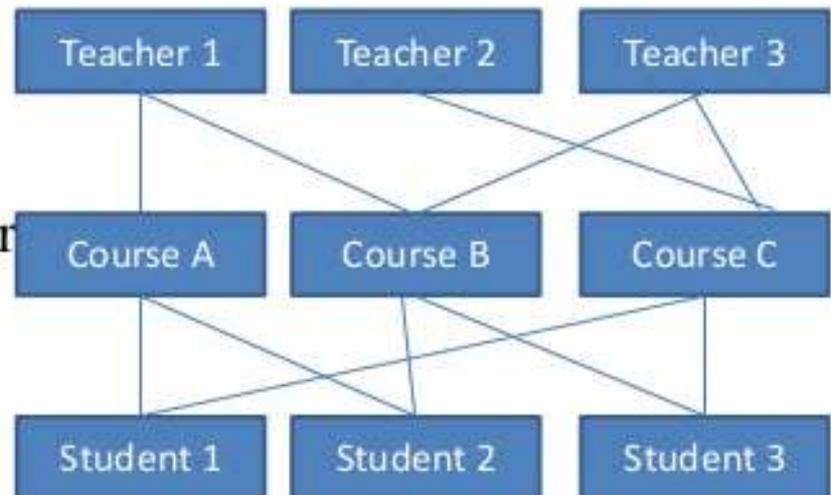
- ✓The network model has greater flexibility than the hierarchical model for handling complex spatial relationships.
- ✓it has not had widespread use as a basis for GIS because of the greater flexibility of the relational model.

Database Concepts

DATA MODELS

NETWORK MODEL – CONCEPT

- Represented using a Data-Structure Diagram.
- Boxes represents the records & lines the links.
- Based on ‘owner-member relationship.’
- Members of an owner may be many but for many member owner is one.
- Can represent one-to-one and many-to-many as well.



Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

Early 1960s, IBM saw business world organizing data in the form of a hierarchy.

Rather than one record type (flat file), a business has to deal with several types which are hierarchically related to each other

e.g. Company has several departments, each with attributes: name of director, number of staff, address

- ✓ each department requires several parts to make its product, with attributes: part number, number in stock
- each part may have several suppliers, with attributes: address, price.

Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

certain types of geographical data may fit the hierarchical model well.

e.g. Census data organized by state, within state by city, within city by census.

The database keeps track of the different record types, their attributes, and the hierarchical relationships between them the attribute which assigns records to levels in the database structure is called the key.

Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

Summary of features

- ✓ a set of record "types"
 - e.g. supplier record type, department record type, part record type
- ✓ a set of links connecting all record types in one data structure diagram (tree).
- ✓ at most one link between two record types, hence links need not be named.
- ✓ for every record, there is only one parent record at the next level up in the tree.
 - e.g. every county has exactly one state, every part has exactly one department.

Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

Summary of features

✓ no connections between occurrences of the same record type
cannot go between records at the same level unless they share the same parent.

Advantages and disadvantages (Next Slide)

Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

Advantages and disadvantages

- ✓ data must possess a tree structure
 - tree structure is natural for geographical data

- ✓ data access is easy via the key attribute, but difficult for other attributes.
 - in the business case, easy to find record given its type (department, part or supplier).
 - in the geographical case, easy to find record given its geographical level (state, county, city, census tract), but difficult to find it given any other attribute
 - e.g. find the records with population 5,000 or less

Database Concepts

DATA MODELS

HIERARCHICAL MODEL – CONCEPT

Advantages and disadvantages

✓ tree structure is inflexible

➤ cannot define new linkages between records once the tree is established.

e.g. in the geographical case, new relationships between objects cannot define linkages laterally or diagonally in the tree, only vertically.

✓ the only geographical relationships which can be coded easily are "is contained in" or "belongs to".