Basics of DBMS

Lec-02

A database is a computer-based record-keeping system whose overall purpose is to record and maintain information

What is DBMS ?

(DataBase Management System)

- A collection of programs that enables you to store, modify, and extract information from a database

- The related information when placed in an organized form makes a database

- Database is a collection of information organized in such a way that a computer program can quickly select desired pieces of data
  - The organization of data/information is necessary because unorganized information has no meaning

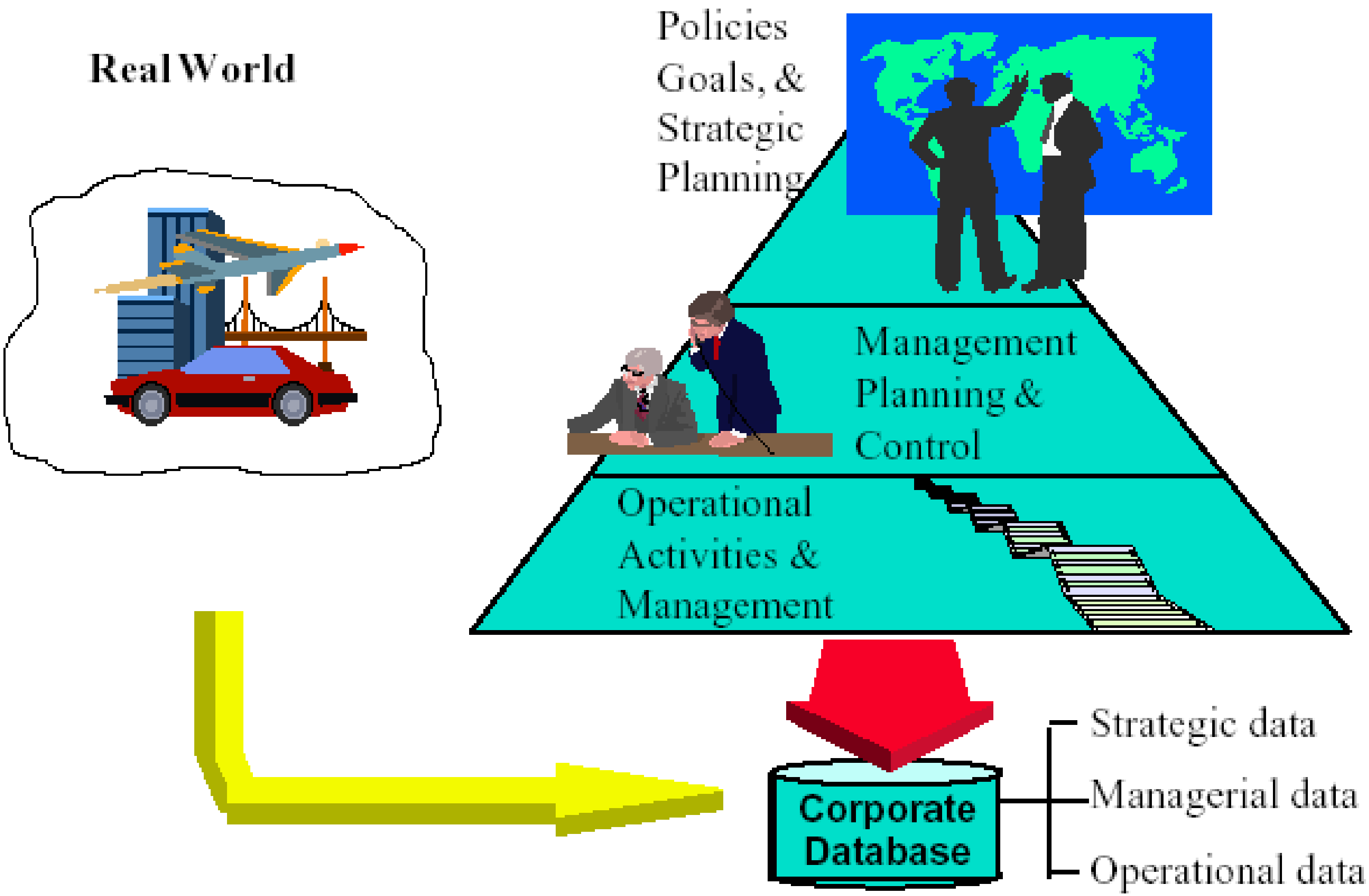
- The general purpose of a DBMS is to provide for the definition, storage, and management of data that can be shared by many users

Why Database Management System?

Database management systems (DBMS) contain information about a particular enterprise

- Collection of interrelated data
- Set of programs to access/process the data
- An environment that is both *convenient and efficient* to use

The Role of DBMS



Real World

Policies
Goals, &
Strategic
Planning

Management
Planning &
Control

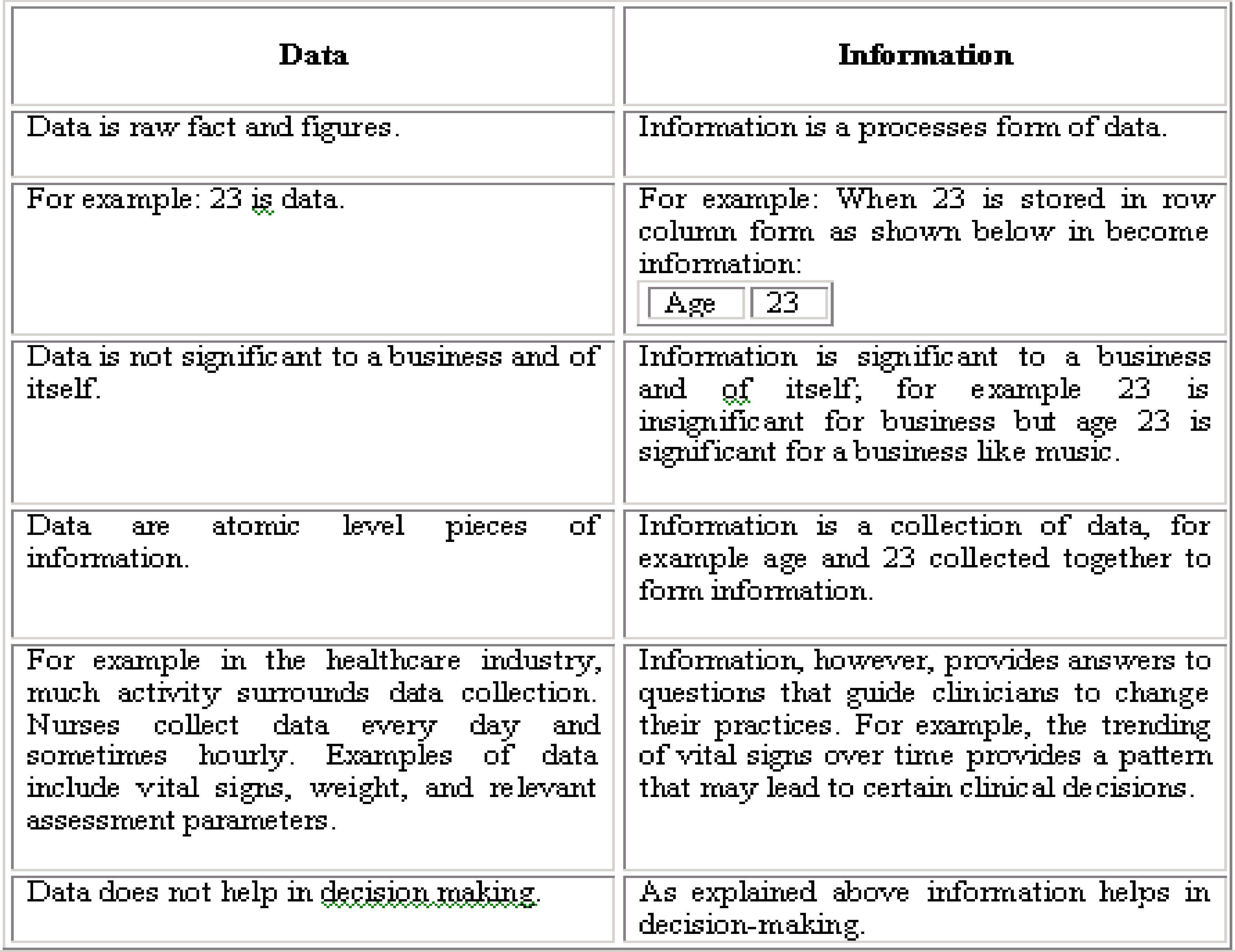
Operational
Activities &
Management

Corporate Database

Strategic data

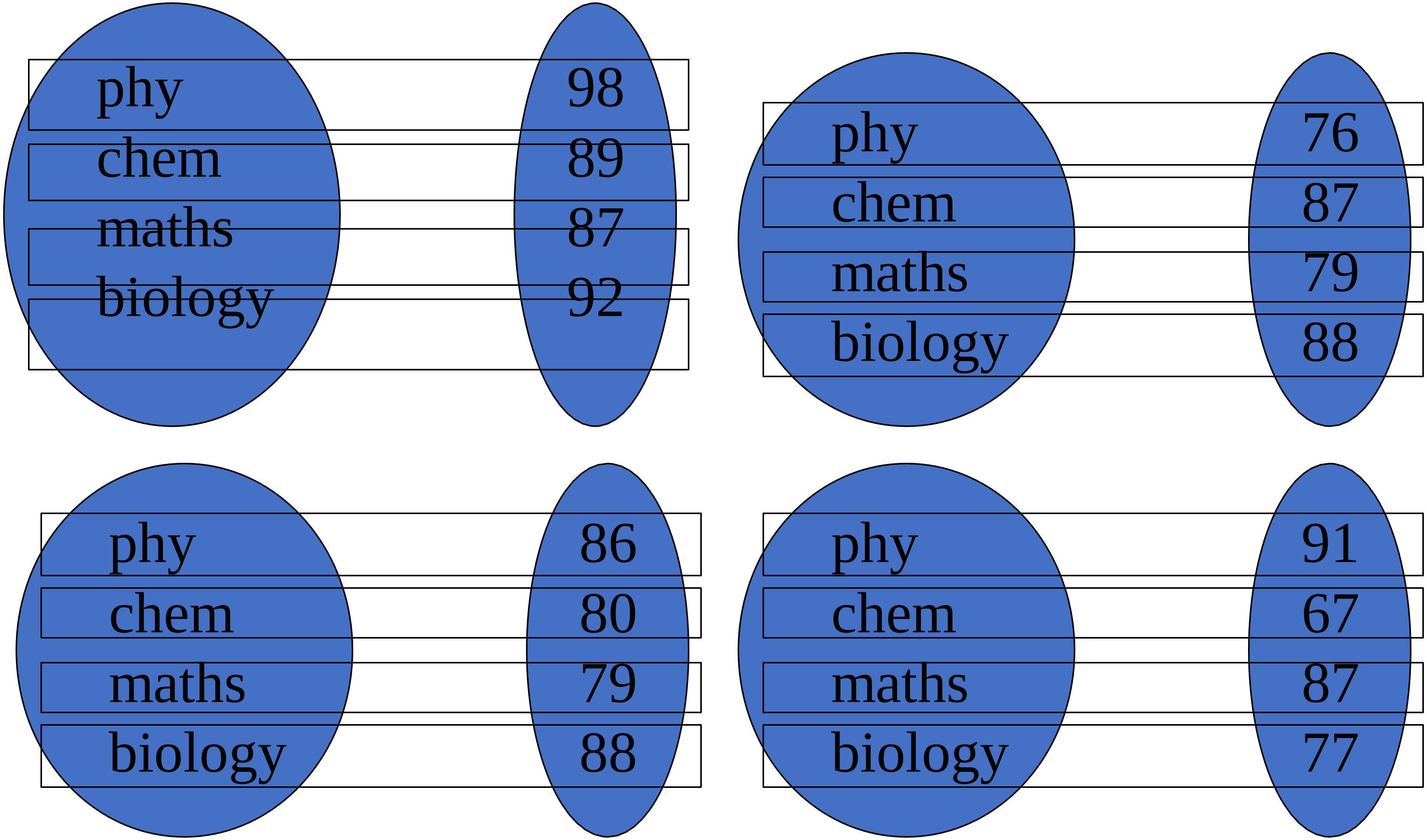
Managerial data

Operational data



Data	Information
Data is raw fact and figures.	Information is a processes form of data.
For example: 23 is data.	For example: When 23 is stored in row column form as shown below in become information: Age 23
Data is not significant to a business and of itself.	Information is significant to a business and of itself; for example 23 is insignificant for business but age 23 is significant for a business like music.
Data are atomic level pieces of information.	Information is a collection of data, for example age and 23 collected together to form information.
For example in the healthcare industry, much activity surrounds data collection. Nurses collect data every day and sometimes hourly. Examples of data include vital signs, weight, and relevant assessment parameters.	Information, however, provides answers to questions that guide clinicians to change their practices. For example, the trending of vital signs over time provides a pattern that may lead to certain clinical decisions.
Data does not help in decision making.	As explained above information helps in decision-making.

DATABASE



phy	98
chem	89
maths	87
biology	92

phy	76
chem	87
maths	79
biology	88

phy	86
chem	80
maths	79
biology	88

phy	91
chem	67
maths	87
biology	77

Operations on Databases

- To **add new information**
- To **view or retrieve** the stored information
- To **modify or edit** the existing information
- To **remove or delete** the unwanted information
- **Arranging the information** in a desired order
- etc.

Manual database and its problems

- Wastage of skills and intelligence of human beings on repetitive calculations.

- Error prone.

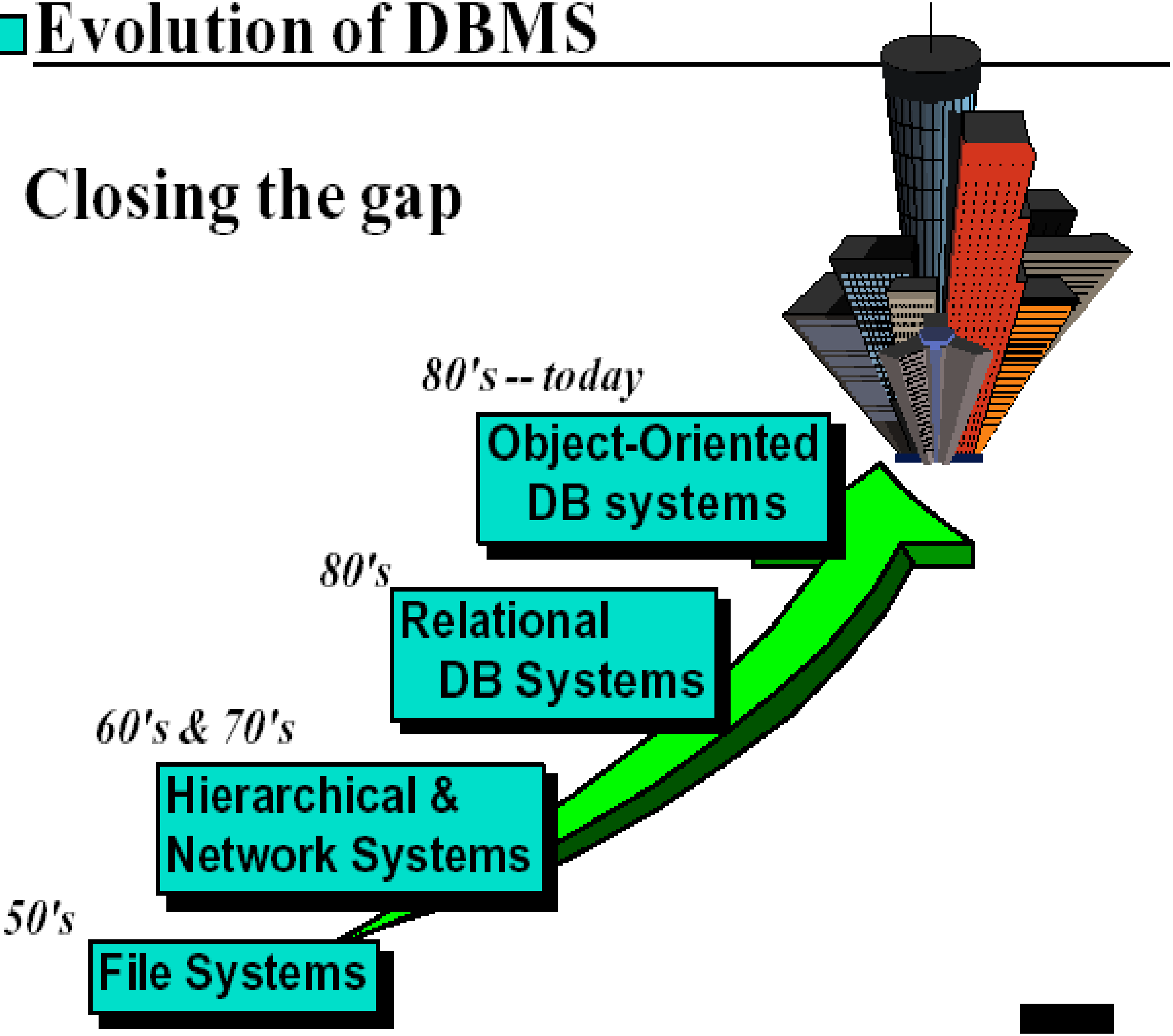
Database and Computers

- Large storage capacity
- It has high speed
- Computer is more accurate.

There are two approaches for storing data in computers such as File based approach and Database approach.

Evolution of DBMS

Closing the gap



80's -- today

Object-Oriented DB systems

80's

Relational DB Systems

60's & 70's

Hierarchical & Network Systems

50's

File Systems

[redacted]

File Based Approach

- File-based systems were an early attempt to computerize the manual filing system that we are all familiar with

- File systems may use a storage device such as a hard disk or CD-ROM and involve maintaining the physical location of the files

- Programmers used programming languages such as C, C++, etc. to write applications that directly accessed flat files to perform data management services and provide information for users.

Drawbacks of using File System

- Data Redundancy and inconsistency
  - Multiple file formats, duplications of information in different files

- Difficulty in accessing data in a convenient and efficient manner
  - Need to write a new program to carry out each new task

- Data Isolation
  - Multiple files and formats
  - Writing new application programs to retrieve the appropriate data is difficult

Drawbacks of using File System

- Integrity Problem
  - Integrity constraints; e.g. minimum balance
  - Hard to add new constraints or change existing one
- Security Problem
  - Hard to provide user access to some but not all, data
- Concurrency
  - Access by multiple users, which is needed for the performance
- Atomicity of updates
  - Failures may leave DB in an inconsistent state with partial updates carried out
- Low productivity and high maintenance cost

Database Approach

Database systems offer solutions to all these problems

Database Approach

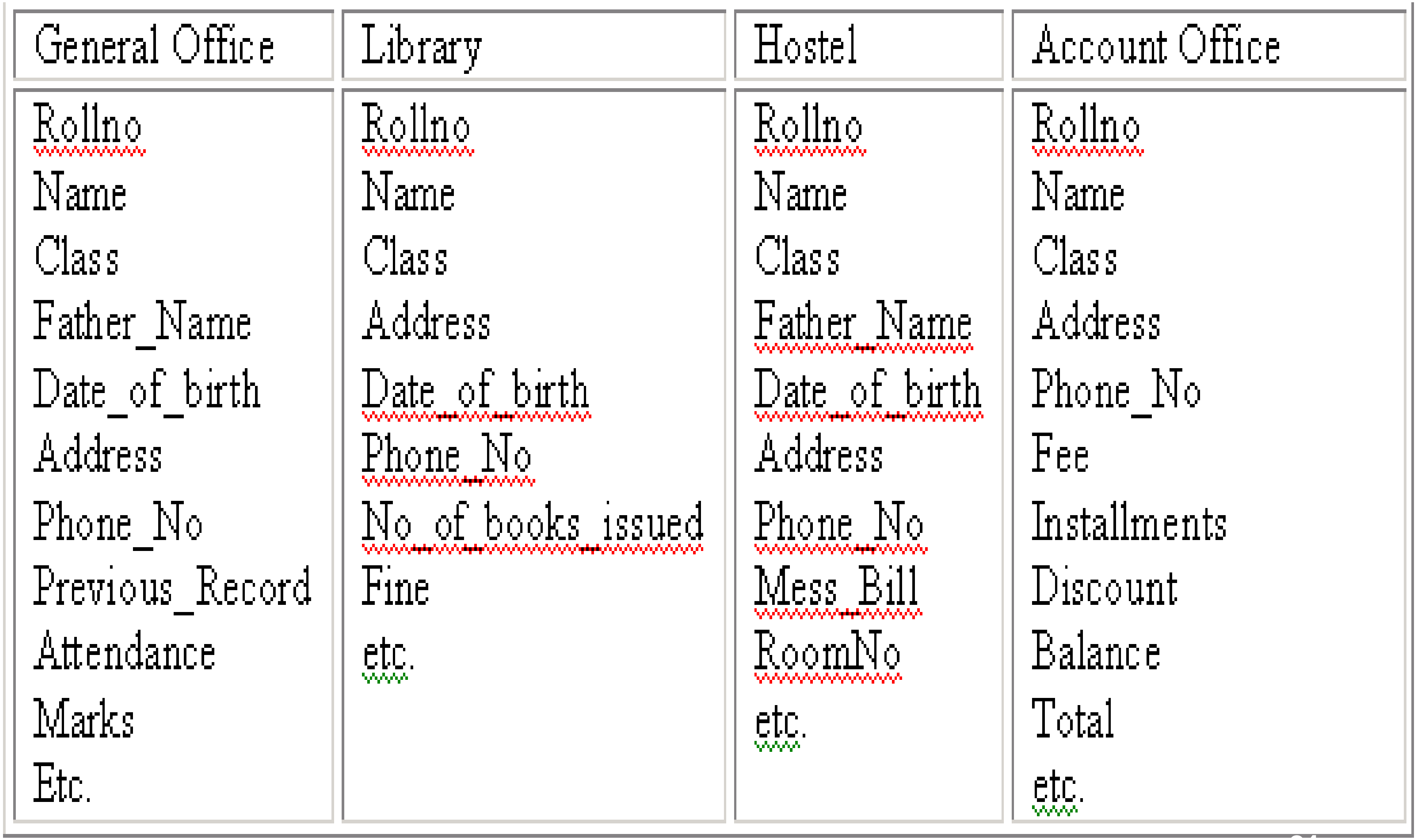
- Database is a collection of information organized in such a way that a computer program can quickly select desired pieces of data
- A DBMS is a software system that allows users to define, create and maintain a database and provides controlled access to the data.
- A database management system (DBMS) is a collection of programs that enables users to store, modify, and extract information from a database per the requirements.

Examples of Database Applications

The following are main examples of database applications:

- **Computerized library systems**
- **Automated teller machines**
- **Flight reservation systems**
- **Computerized inventory systems**
- **Commercially available Database management systems in the market are dbase, Foxpro and Oracle etc.**

Example
University Database in File Based System



General Office	Library	Hostel	Account Office
Rollno	Rollno	Rollno	Rollno
Name	Name	Name	Name
Class	Class	Class	Class
Father_Name	Address	Father_Name	Address
Date_of_birth	Date_of_birth	Date_of_birth	Phone_No
Address	Phone_No	Address	Fee
Phone_No	No_of_books_issued	Phone_No	Installments
Previous_Record	Fine	Mess_Bill	Discount
Attendance	etc.	RoomNo	Balance
Marks		etc.	Total
Etc.			etc.

Advantages of DBMS

- **Controlling Redundancy**
- **Integrity can be enforced**

  The integrity of data means that data in a database is always accurate, such that incorrect information cannot be stored in it
- **Inconsistency can be avoided**

  When the same data is duplicated and changes are made at one site, which is not propagated to the other site, it gives rise to inconsistency and the two entries regarding the same data will not agree
- **Data can be shared**
- **Providing backup and recovery**
- **Restricting unauthorized access**
- **Solving enterprise requirements than an individual requirement**

View of Data

- A database system is a collection of interrelated data and a set of programs that allow users to access and modify these data

- A major purpose of a database system is to provide users with an abstract view of the data
  - Data abstraction
    - Hide the complexity of data structures to represent data in the database from users through several levels of data abstraction
  - Data models
    - A collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints

Levels of Abstraction

- **Physical level:**
  - describes how a record (e.g., instructor) is stored

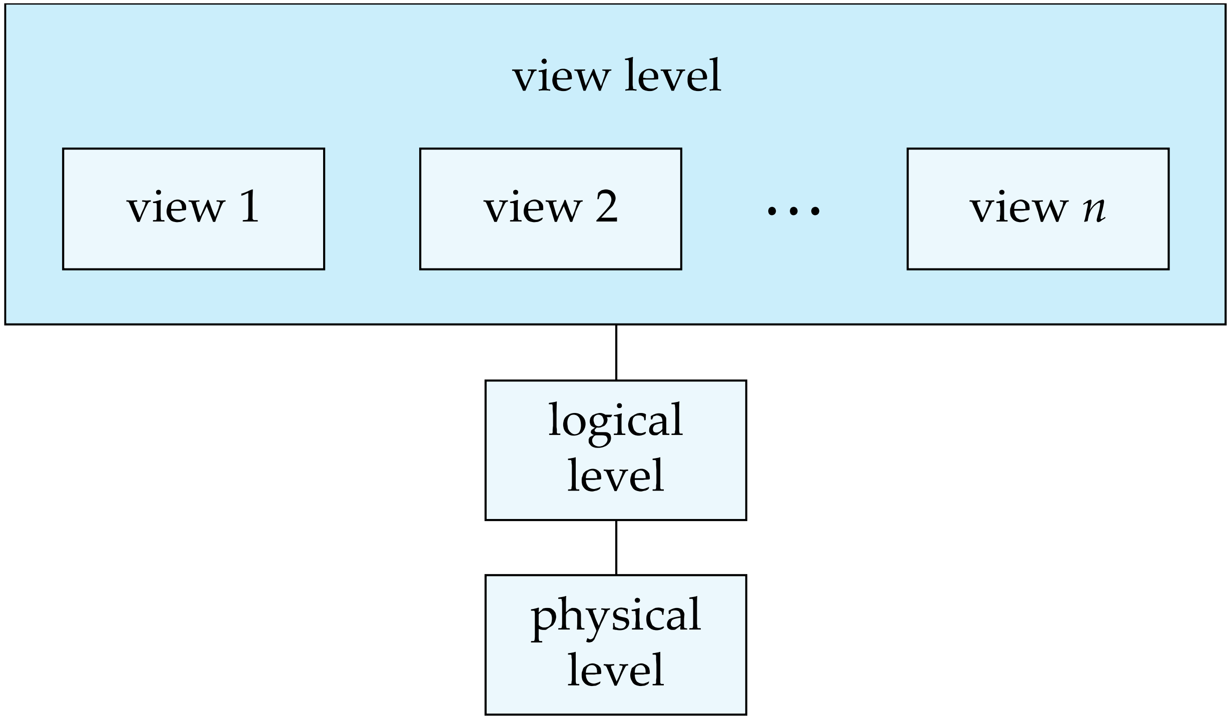
- **Logical level:**
  - describes data stored in database, and the relationships among the data.

```
type instructor = record
            ID : string;
            name : string;
            dept_name : string;
            salary : integer;
            end;
```

- **View level:**
  - application programs hide details of data types
    - Views can also hide information (such as an employee’ s salary) for security purposes.

View of Data

An architecture for a database system



Instances and Schemas

- Similar to types and variables in programming languages

- **Logical Schema** – the overall logical structure of the database
  - Analogous to type information of a variable in a program
  - Example: The database consists of information about a set of customers and accounts in a bank and the relationship between them
    - Customer Schema:

      Name, Customer ID, Account #, Aadhar ID, Mobile #
    - Account Schema:

      Account #, Account Type, Interest Rates, Balance

- **Physical schema** – the overall physical structure of the database

Instances and Schemas

- **Instance** – the actual content of the database at a particular point in time
  - Analogous to the value of a variable

Customer Instances:

Name	Customer ID	Account #	Aadhar ID	Mobile #
Lavish	6728	256389	2943115306352	9689562302
Surbhi	8912	125635	9186325686123	8956236523
Piyush	6617	372912	1233543135438	7895423059

Physical Data Independence

- **Physical Data Independence** –
  The ability to modify the physical schema without changing the logical schema

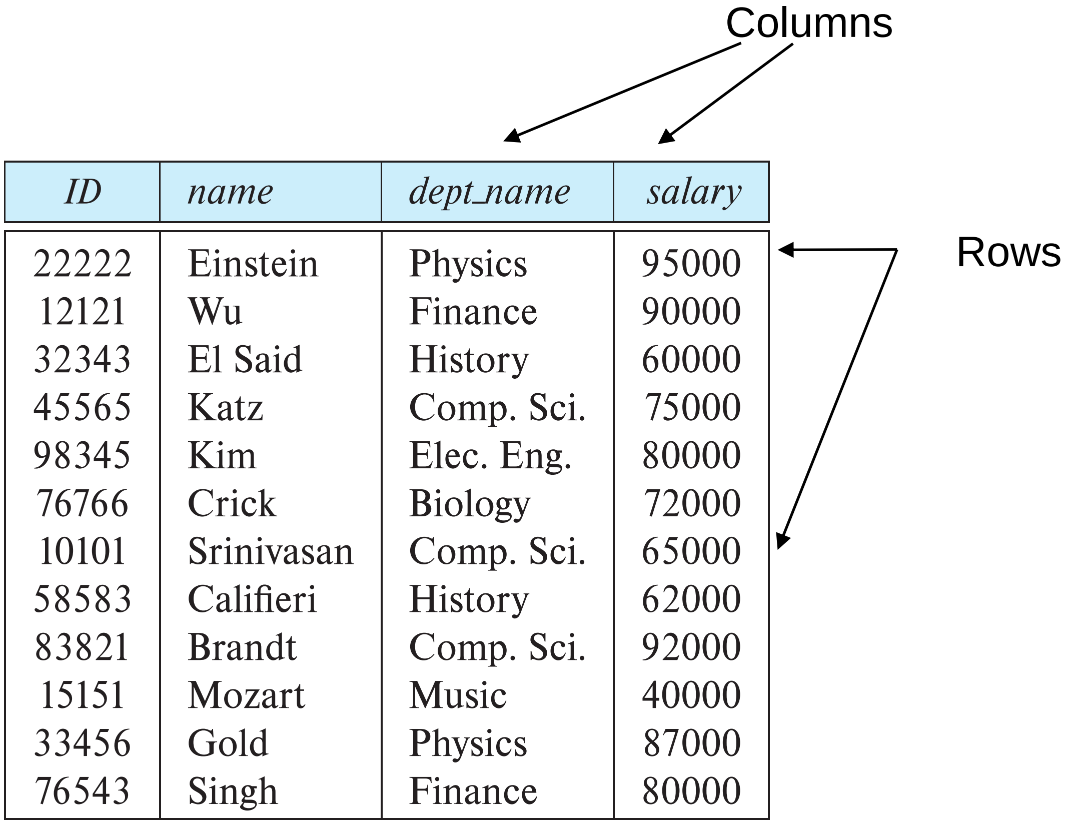
  - Applications depend on the logical schema
  - In general, the interfaces between the various levels and components should be well-defined so that changes in some parts do not seriously influence others.

Data Models

- A collection of tools for describing
  - Data
  - Data relationships
  - Data semantics
  - Data constraints
- Relational model
- Entity-Relationship data model (mainly for database design)
- Object-based data models (Object-oriented and Object-relational)
- Semi-structured or unstructured data models (XML)
- Other older models:
  - Network model
  - Hierarchical model

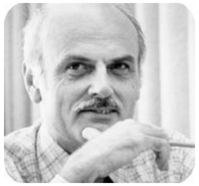
Relational Models

- All the data is stored in various tables.
- Example of tabular data in the relational model


Columns
Rows

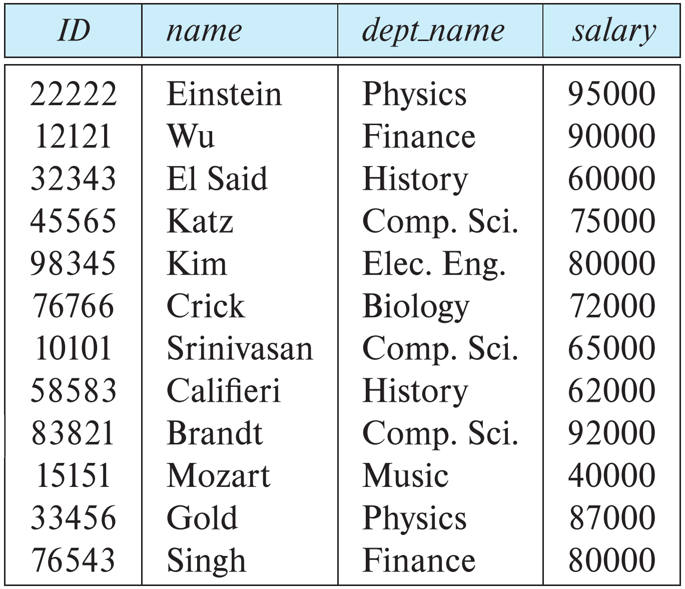
ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table



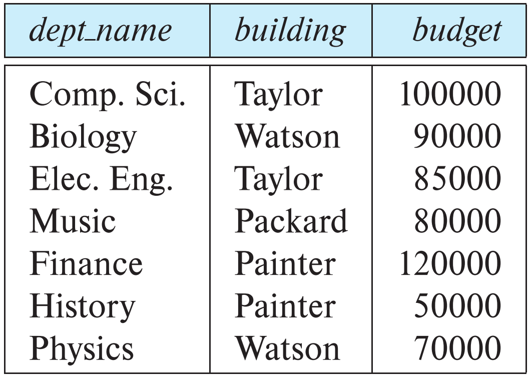
Ted Codd
Turing Award 1981

A Sample Relational Database



ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table



dept_name	*building*	*budget*
Comp. Sci.	Taylor	100000
Biology	Watson	90000
Elec. Eng.	Taylor	85000
Music	Packard	80000
Finance	Painter	120000
History	Painter	50000
Physics	Watson	70000

(b) The *department* table

Database Languages

A database system provides a

- **Data-Definition Language (DDL)** to specify the database schema

- **Data-Manipulation Language (DML)** to express database queries and updates

Data Definition Language (DDL)

- Specification notation for defining the database schema

  Example:

  ```
  create table instructor (
                     ID           char(5),
                     name         varchar(20),
                     dept_name    varchar(20),
                     salary       numeric(8,2))
  ```

- DDL compiler generates a set of table templates stored in a ***data dictionary***

- Data dictionary contains metadata (i.e., data about data)
  - Database schema
  - Integrity constraints
    - Primary key (ID uniquely identifies instructors)
  - Authorization
    - Who can access what

Data Manipulation Language (DML)

- Language for accessing and updating the data organized by the appropriate data model

The types of access are:

- Retrieval of information stored in the database
- Insertion of new information into the database
- Deletion of information from the database
- Modification of information stored in the database

A query is a statement requesting the retrieval of information

- The portion of a DML that involves information retrieval is called a query language
- DML also known as a query language

Data Manipulation Language (DML)

- Two classes of languages

- Pure- used for proving properties about computational power and for optimizations
  - *Relational Algebra*
  - Tuple relational calculus
  - Domain relational calculus

- Commercial- used in commercial systems
  - SQL is the most widely used commercial language

SQL

- SQL query language is nonprocedural. A query takes as input several tables (possibly only one) and always returns a single table
- Example to find all instructors in Comp. Sci. dept
- **select** *name*
  from *instructor*
  where *dept_name* = 'Comp. Sci.'

- SQL is **NOT** a Turing machine equivalent language
- To be able to compute complex functions SQL is usually embedded in some higher-level language

- Application programs generally access databases through one of
  - Language extensions to allow embedded SQL
  - Application program interface allows SQL queries to be sent to a database