classification of Database Management Systems

The Relational Data Model and Relational Database Constraints

CLASSIFICATION OF DATABASE MANAGEMENT SYSTEMS

- Database management systems can be classified based on several criteria, such as the **data model, user numbers and database distribution.**

Classification Based on Data Model

- Relational Data Model
- Hierarchical Data Model
- Network Data Model
- Object-Oriented Data Model

- The **Relational Data Model** in Database Management Systems (DBMS) organizes data into tables (relations), each consisting of rows (tuples) and columns (attributes).
- It uses primary keys to uniquely identify records and foreign keys to establish relationships between tables.
- This model ensures data integrity through constraints like primary keys, foreign keys, and unique constraints.
- Well-known DBMSs like Oracle, MS SQL Server, DB2 and MySQL support this model.

- The **Hierarchical Data Model** is a way of organizing data in a tree-like structure.
- Each node represents a record (or entity), and the edges represent relationships between the records.
- The topmost node is called the root, and the other nodes are subordinates of the root.
- The **Network Data Model** is an extension of the hierarchical model and allows more complex relationships between data entities.
- It organizes data as a graph where multiple parent-child relationships can exist, and records can have more than one parent.
- This is done through sets of relationships, which are used to form networks of data.

- The **Object-Oriented Data Model** (OODM) in Database Management Systems (DBMS) is a data model that integrates the principles of object-oriented programming (OOP) with the traditional database concepts.
- Object-oriented database management systems (OODBMS) combine database capabilities with object-oriented programming language capabilities.
- This model allows databases to store objects, which are instances of classes, in a way that mirrors the behavior of objects in object-oriented programming languages like Java, C++, and Python.

- The most popular data model in use today is the **relational data model**.
- Other traditional models, such as **hierarchical data models** and **network data models**, are still used in industry mainly on mainframe platforms.
- However, they are not commonly used due to their complexity.
- These are all referred to as traditional models because they preceded the relational model.

Classification Based on User Numbers

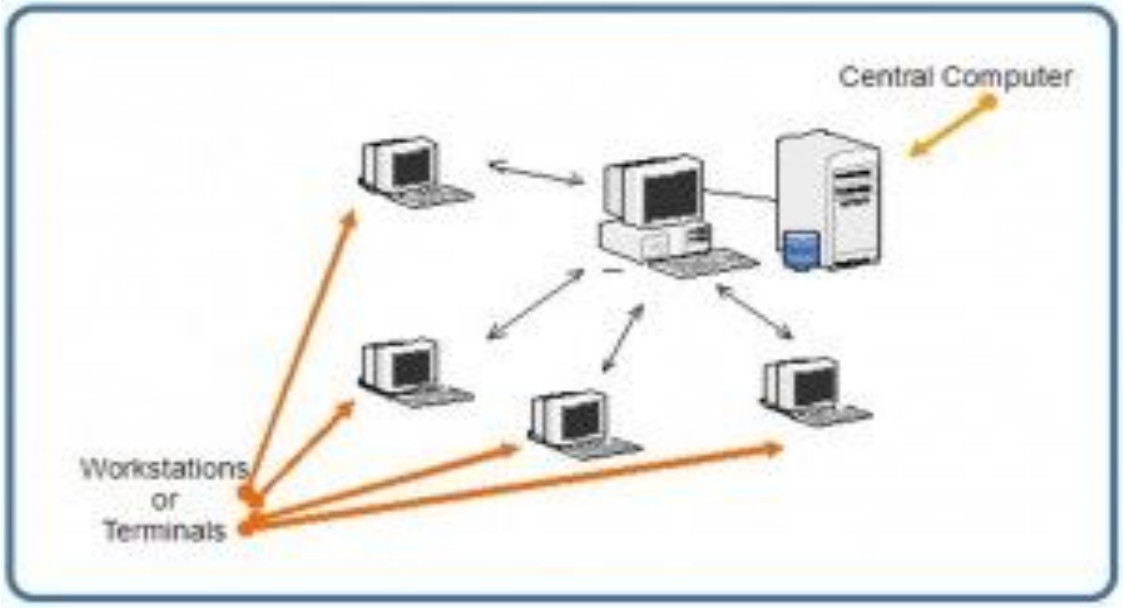
- Single-user database system
- Multi-user database system

- A DBMS can be classified based on the number of users it supports.
- It can be :
  - Single-user database system, which supports one user at a time, or
    - It typically serves smaller-scale applications where only one individual interacts with the database, such as a personal application or standalone system.
  - Multiuser database system, which supports multiple users concurrently.
    - Unlike single-user systems, multi-user systems enable many users to interact with the database concurrently without interfering with each other's operations.

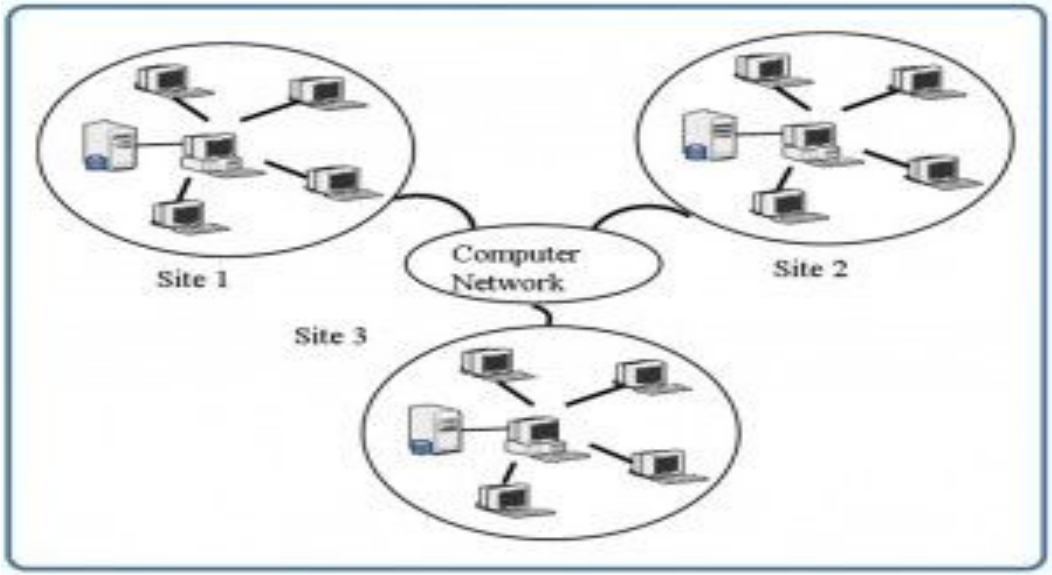
Classification Based on Database Distribution

- Centralized systems
- Distributed Systems

With a centralized database system, the DBMS and database are stored at a single site that is used by several other systems too.



In a distributed database system, the actual database and the DBMS software are distributed from various sites that are connected by a computer network.



- Distributed database system:
  - Homogeneous systems
  - Heterogeneous systems

Homogeneous distributed database systems

- Homogeneous distributed database systems use the same DBMS software from multiple sites.
- ***Data exchange between these various sites can be handled easily.***
- For example, library information systems by the same vendor may use the same DBMS software which allows easy data exchange between the various library sites.

Heterogeneous distributed database systems

- In a heterogeneous distributed database system, different sites might use different DBMS software, but there is ***additional common software to support data exchange between these sites***.
- For example, the various library database systems use the same machine-readable cataloguing (MARC) format to support library record data exchange.

Classification Based on Cost:

1. Low cost DBMS – The cost of these systems vary from $100 to $3000.
2. Medium cost DBMS – Cost varies from $10000 to $100000.
3. High cost DBMS –Cost of these systems are usually more than $100000

Classification Based on Access:

1. Sequential access – One after the other.
2. Direct access
3. Inverted file structures

Classification Based on the usage:

- Online transaction processing(OLTP) DBMS
- Online analytical processing(OLAP) DBMS
- Big data and analytics DBMS
- XML DBMS, GIS DBMS, Sensor DBMS, Mobile DBMS, Open source DBMS

Relational Model Concepts

- A Relation is a mathematical concept based on the ideas of sets
- The model was first proposed by Dr. E.F. Codd of IBM Research in 1970 in the following paper: "A Relational Model for Large Shared Data Banks," Communications of the ACM, June 1970
- The above paper caused a major revolution in the field of database management and earned Dr. Codd the coveted ACM Turing Award

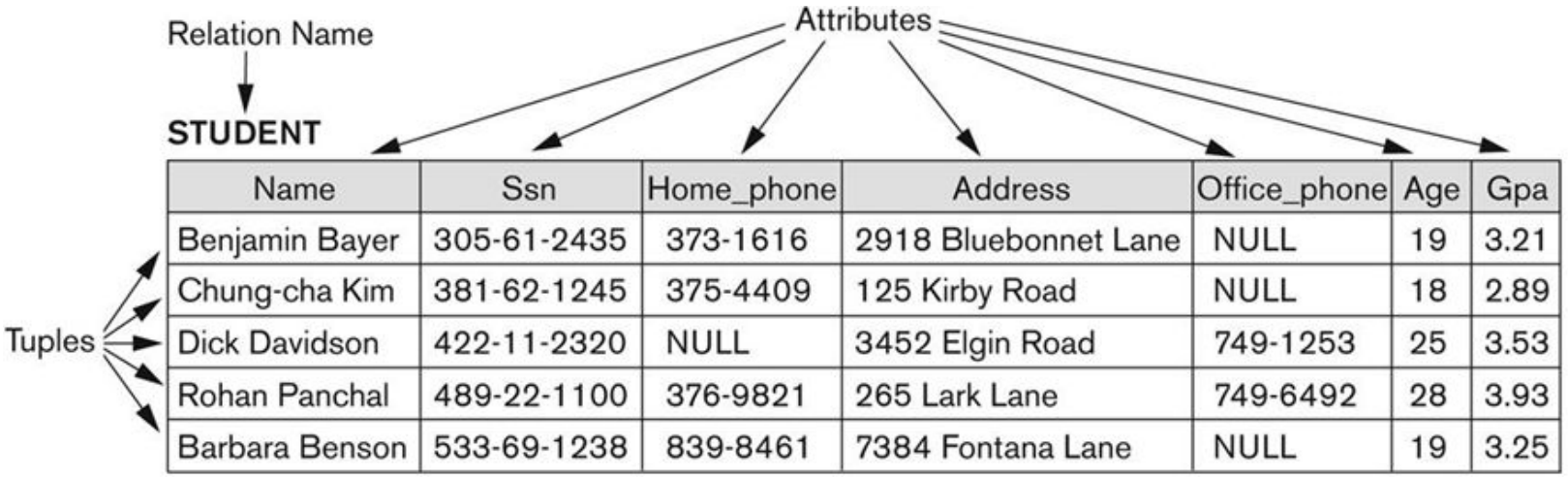
INFORMAL DEFINITIONS

RELATION: A table of values

- A relation may be thought of as a set of rows.
- Each row represents a fact that corresponds to a real-world entity or relationship.
- Each row has a value of an item or set of items that uniquely identifies that row in the table.
- Each column typically is called by its column name or column header or attribute name.

Key of a Relation: Called the key

- In the STUDENT table, SSN is the key
- Sometimes row-ids or sequential numbers are assigned as keys to identify the rows in a table



Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	NULL	19	3.21
Chung-cha Kim	381-62-1245	375-4409	125 Kirby Road	NULL	18	2.89
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	749-1253	25	3.53
Rohan Panchal	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	NULL	19	3.25

The attributes and tuples of a relation STUDENT.

FORMAL DEFINITIONS

- The **Schema** (or description) of a Relation:
  - Denoted by R(A1, A2, ...An)
  - R is the name of the relation
  - The attributes of the relation are A1, A2, ..., An

Example:

CUSTOMER (Cust-id, Cust-name, Address, Phone#)

- CUSTOMER is the relation name
- Defined over the four attributes: Cust-id, Cust-name, Address, Phone#
- Each attribute has a domain or a set of valid values.
- For example, the domain of Cust-id is 6 digit numbers.

- A row is called a tuple, which is an ordered set of values
- A column header is called an attribute

– Each attribute value is derived from an appropriate domain.

- The table is called a relation.

– A relation can be regarded as a set of tuples (rows).

- The data type describing the types of values an attribute can have is represented by a domain of possible values.
- Each row in the CUSTOMER table is a 4-tuple and consists of four values, for example.

<632895, "John Smith", "101 Main St. Atlanta, GA 30332",

"(404) 894-2000">

- A relation is a set of such tuples (rows).
- A domain has a logical definition.
- Example: “USA_phone_numbers” are the set of 10 digit phone numbers valid in the U.S.

- The relation is formed over the cartesian product of the sets; each set has values from a domain
- • The Cartesian product of two sets A and B is defined to be the set of all pairs (a, b) where $a \in A$ and $b \in B$. It is denoted $A \times B$, and is called the Cartesian product

SID	SName
111	Williams
222	Johnes

C_NO	SID
1000	111
2000	222

STUDENT × ENROLLMENT

STUDENT

SID	SName
111	Williams
111	Williams
222	Johns
222	Johns

ENROLLMENT

C_NO	SID
1000	111
2000	222
1000	111
2000	222

RESULT

SID	SName	C_NO	SID
111	**Williams**	**1000**	**111**
111	Williams	2000	222
222	Johns	1000	111
222	**Johns**	**2000**	**222**

- The degree of a relation is the number of attributes n of its relation schema.
- A relation schema R of degree n is denoted by R(A1 , A2 , …, An)
- The domain of Ai is denoted by dom(Ai).

DEFINITION SUMMARY

Informal Terms	Formal Terms
Table	Relation
Column	Attribute/Domain
Row	Tuple
Values in a column	Domain
Table Definition	Schema of a Relation
Populated Table	Extension

Characteristics of Relations

- ***Ordering of tuples in a relation r(R)***:
- – The tuples are not considered to be ordered, even though they appear to be in the tabular form.
- ***Ordering of attributes in a relation schema***

R (and of values within each tuple):

- – We will consider the attributes in R(A1, A2, ..., An) and the values in t=<v1, v2, ..., vn> to be ordered .
- ***Values in a tuple:*** All values are considered atomic (indivisible). A special null value is used to represent values that are unknown or inapplicable to certain tuples.
- Multivalued attributes-Represented by separate relations . Ex: A person may have many phone numbers and email addresses.
- Composite attributes -Represented by single component. Ex. Address. It can be divided into house number, street, city, and state.