

Database Management System(DBMS)

...

What Is Data ???

- Data is a collection of raw facts.
- Which is isolated , un organized and un-interpreted but able to organize into useful information.
- Eg: 1, Ram, KTM

What Is Information ???

- Information is the processed or interpreted data.
- Eg :

Address Of Students		
SN	Name	Address
1	Ram	KTM
2	Shyam	JHAPA

Data vs Information

Data	Information
Data is a collection of raw facts, figures, videos, images etc	Information is the processed data.
Data is Un organized in nature	Information is organized.
Data is isolated in nature	Information is dependent to others
Data is un interpreted	Information is interpreted in nature

What Is Database???

- It is a collection of related information about a subject organized in a useful manner that provides a base or foundation for procedure, such as retrieving information, drawing conclusion and make decision.
- A database is a collection of interrelated data of a person or place in such a way that it can be accessed by DBMS.(Database Management System)
- Example : Banking record of the customers, attendance record of the students in school ,etc

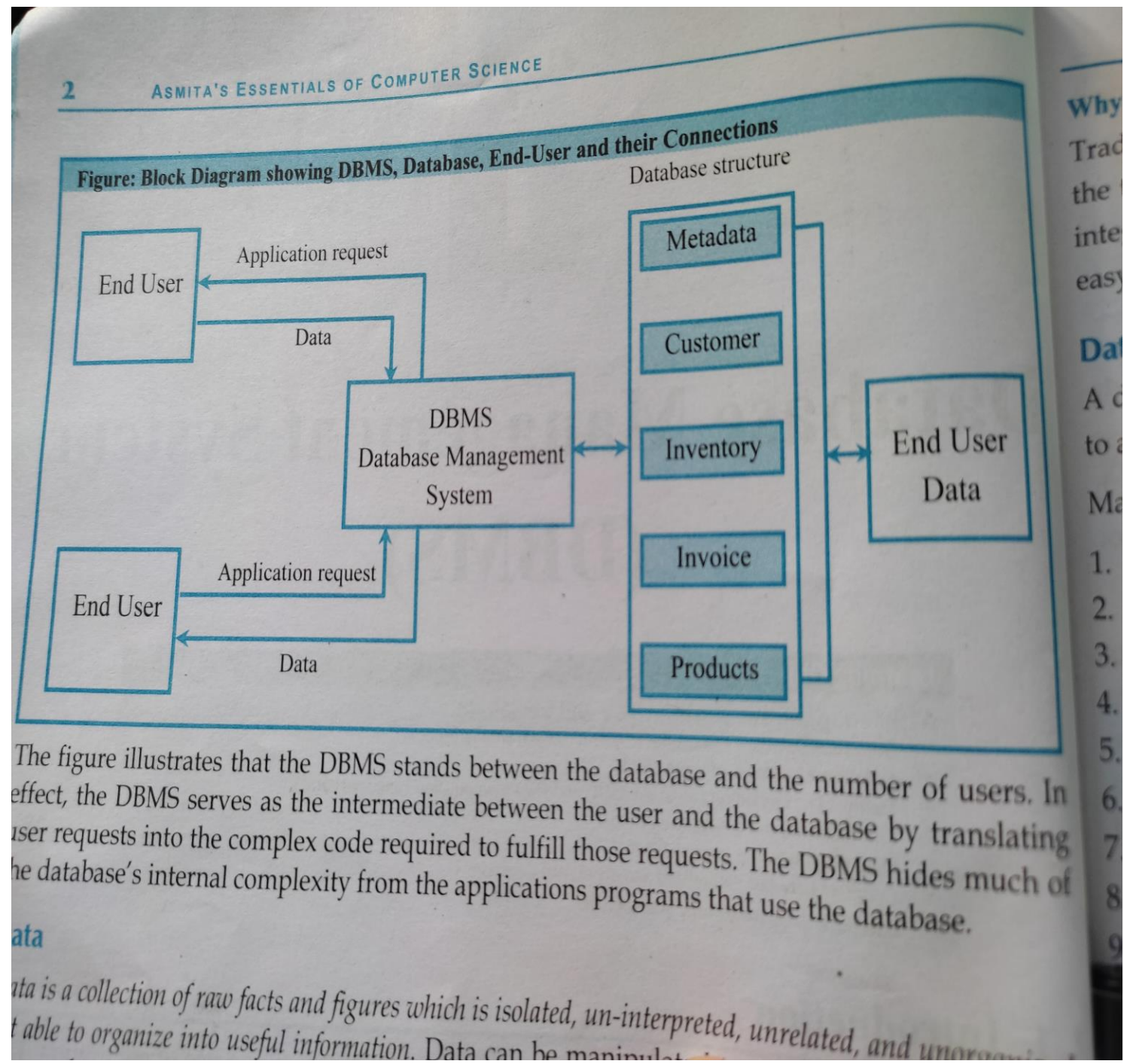
Database System ???

- Database systems consists of set of programs to access, update and modify the data.
- Database and database management system, combine it forms a **Database System**.
- In other words, both the database and database management system are the parts of Database System (DS).

Database Management System(DBMS)

- Database management system is a software which is used to manage the database File or database.
- DBMS provides an interface to perform various operations like database creation, storing data in it, updating data, creating a table in the database and a lot more.
- Example : oracle, MySQL, MongoDB

Block diagram showing DBMS, Database, End-user(Users) and their connections.



Advantages Of DBMS

- **Controls database redundancy:** It can control data redundancy because it stores all the data in one single database file and that recorded data is placed in the database.
- **Data sharing:** In DBMS, the authorized users of an organization can share the data among multiple users.
- **Integrity Can be Maintained:** It Ensures that data stored in database is correct.

Advantages Of DBMS (Continue)

- **Data Security:** Database can be secured by imposing or applying some of the restrictions provided by DBMS.
- **Easily Maintenance:** It can be easily maintainable due to the centralized nature of the database system.
- **multiple user interface:** It provides different types of user interfaces like graphical user interfaces, application program interfaces
- **Backup:** It provides backup and recovery subsystems which create automatic backup of data

Dis-advantages Of DBMS

- **Cost of Hardware and Software:** It requires a high speed of data processor and large memory size to run DBMS software.
- **Size:** It occupies a large space of disks and large memory to run them efficiently.
- **Complexity:** Database system creates additional complexity and requirements.
- **Higher impact of failure:** Failure is highly impacted the database because in most of the organization, all the data stored in a single database and if the database is damaged due to electric failure or database corruption then the data may be lost forever.

Problems With File Processing

- Data Duplication.
- Data Integrity Problem.
- Difficulty In Accessing Data(difficult to retrieve data when needed).
- Security Problem.

Objective Of DBMS

- To provide a storage area for the mass amount of data.
- To make access to the data easy to the end user(Users).
- To eliminate data redundancy(duplication of data)
- To allow multiple users to access the data.
- To protect the data using securities provided by DBMS.

Some DBMS tools or keyword.

- **Bit** : bit is the smallest unit of memory. (Either 0 or 1).
- **Character**: character is a collection of 8 bits. It includes alphabets, number and symbols(symbols like @*&-). It is usually known as CHAR in MYSQL.
- **Field** : A database field is a set of data values, of the same data type, in a table. Field is also called Column or Attribute.
- **Record**: A record is a group of data saved in a table.
- **File** : file is a collection of related records.
- **Database**: database is a collection of files.

Structured Query Language(SQL)

- SQL stands for Structured Query Language. It is used for storing and managing data in relational database management system (RDMS).
- It is a standard language for Relational Database System. It enables a user to create, read, update and delete relational databases and tables.

Language for DBMS (Commands)

- **Data Definition Language(DDL):** is used to define a database schema.
- Commands: CREATE, DROP, ALTER etc
- **Data Manipulation Language(DML):** is used to manipulate the data which already exists in the database.
- Commands: UPDATE, DELETE etc

Database Model

Types Of Database Model

- 1. Hierarchical Database Model.
- 2. Network Database Model.
- 3. Relational Database Model.
- 4. Object-oriented Model.

Hierarchical Database Model?

- These are the oldest and old-fashioned database model in which records are logically organized into a hierarchy of relationships. A hierarchical database is arranged logically in an inverted tree pattern. All records in hierarchy are called nodes i.e. either parent node or child node which are related to the each others in a parent-child relationship. Each parent record may have one or more child records, each can have only one parent. The top level record in the hierarchy is called the root record.

Hierarchical Database Model

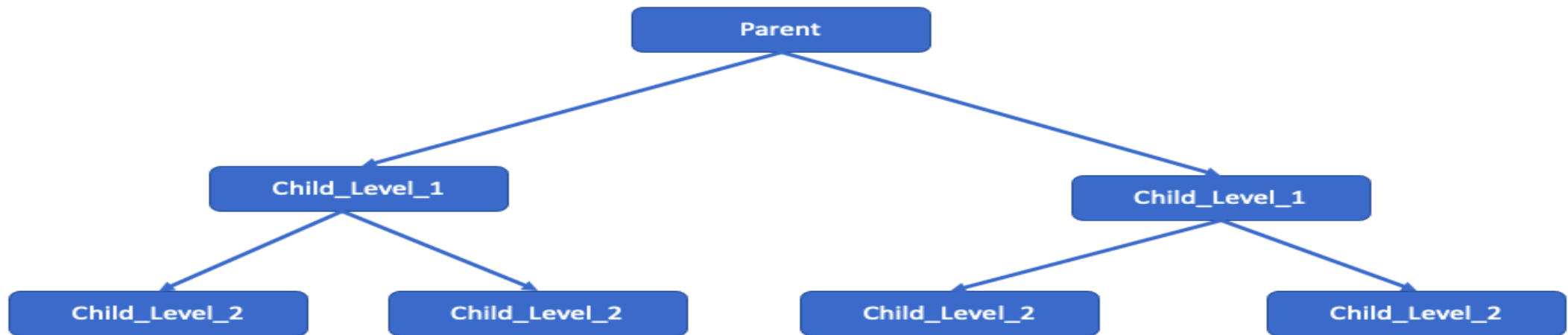


Fig: Hierarchical Database Model

Advantages of Hierarchical Database Model

- It is the easiest Model of the database.
- It is More secure.
- Searching is easy and fast if the parent is known.
- It helps to build complex system from simple components.
- It is suitable for 'One-to-many' relationship handling.
- It follows top down approach.

Dis-advantages of Hierarchical Database Model

- It is old fashioned, outdated database model.
- It is non-flexible database model. That means modification and addition of child without consulting the parent is impossible or very complex process.
- It can't handle Many-to-Many relationships.
- It increases data redundancy as same data can be stored in different places.
- If parent node is deleted, all the child nodes associated with it are deleted automatically.

Network Database Model?

- Network database model is a modified version of the hierarchical database that overcomes the limitation of hierarchical model. In Hierarchical single child not can't have multiple parents rather in network database model each node (child) may have several parents. The network database model has a higher level of flexibility than hierarchical model. It is easier to restructure the information stored in this structure. It is less user-friendly and secured when compared to the hierarchical structure.

Network Database Model

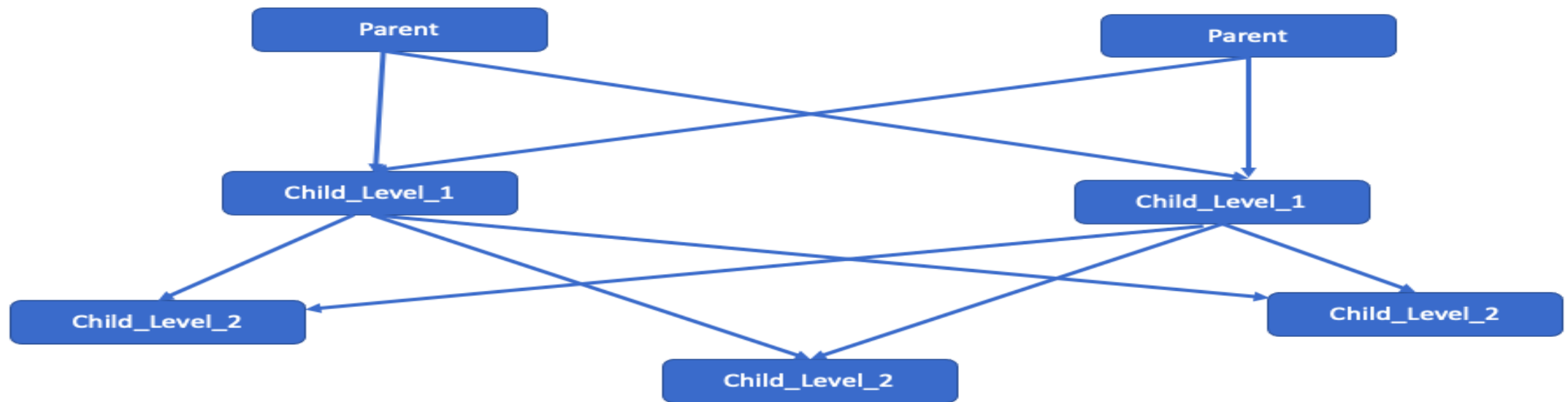


Fig: Network Database Model

Advantages of Network Database Model

- 'many-to-many' relationship is possible which makes it more flexible than hierarchical
- It reduces data redundancy i.e. repetition of data.
- Searching is fast because of multi-directional pointers.

Dis-advantages of Network Database Model

- It is one of the complex database models. (as it can also follow bottom up approach)
- As nodes increases, table and relationship increases which needs larger program to handle.
- It is less secured compared to hierarchical model, as its nodes are open to all.
- Pointers needed in this model increases over head of storage.

Relational Database Model?

- In a relational database model, data are arranged in two-dimensional table I.e. in rows and column, which are easy for a user to develop and understand. In general people often work with tabular form data, it is easy for most of them to understand the structure used in a relational database as data are in table. It is a highly flexible to program and retrieve data. It can be efficiently used even with a computer which has limited memory and processing capability. It is much easier to use since it enables the computer system to accommodate different types of enquiries in an efficient manner

Relational Database Model

The diagram illustrates a table in a Relational Database Model. The table has three columns: ID, Name, and Salary. It contains four rows of data. Annotations include a blue arrow labeled 'Column' pointing to the column headers and a blue arrow labeled 'Rows' pointing to the data rows.

ID	Name	Salary
1001	Jhon	10000
1002	Mike	20000
1003	Paula	30000
1004	Sunny	40000

Advantages of Relational Database Model

- It has very less data redundancy.
- Normalization (Breaking of single complex table into several simpler one) of the database is possible.
- If new data is to be added to an existing database, it is not necessary to redesign the database afresh. Rather, new table containing the new data can be added to the database and these tables can be related to the existing tables.
- Since one table is linked with other, some common fields and rules implemented on one table can easily be implemented to another.

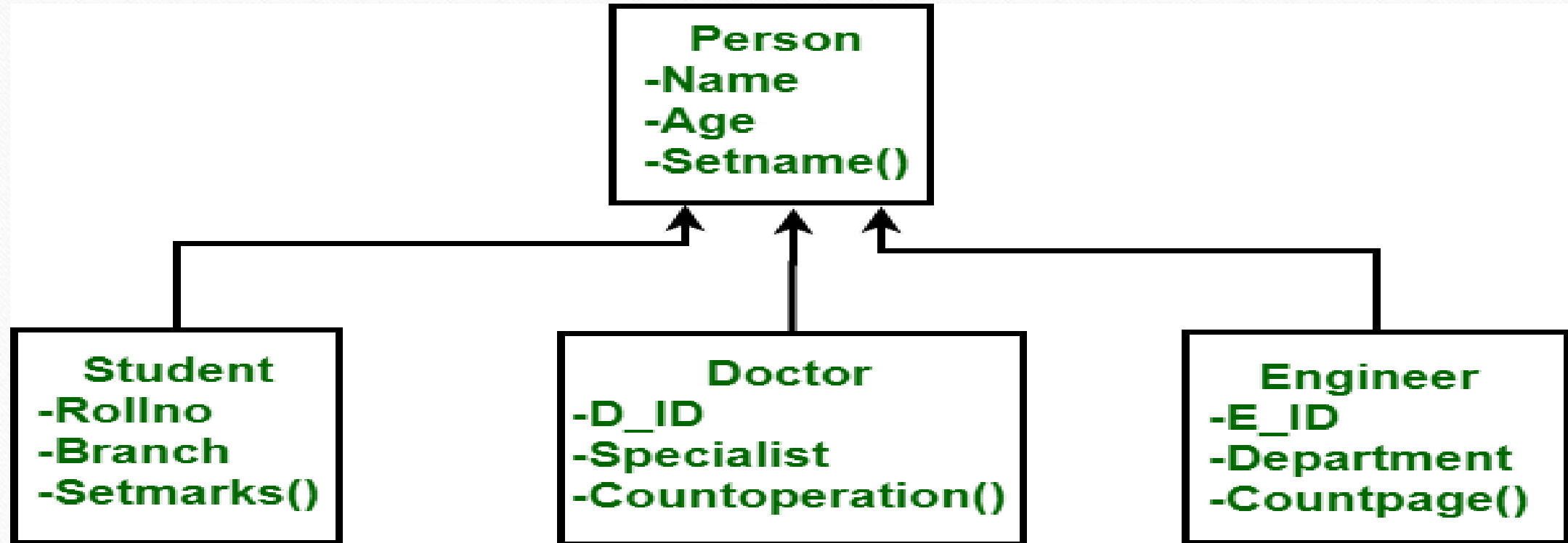
Dis-advantages of Relational Database Model

- It is more complex than other models.
- Too many rules makes database non-user-friendly. (Hard to understand)

Object Oriented Model

- To represent the complex real world problems there was a need for a data model that is closely related to real world. Object Oriented Data Model represents the real world problems easily.
- In Object Oriented Data Model, data and their relationships are contained in a single structure which is referred as object in this data model. In this, real world problems are represented as objects with different attributes. All objects have multiple relationships between them. Basically, it is combination of Object Oriented programming and Relational Database Model.

Object Oriented Model



Relational Database Management System(RDBMS)

- RDBMS stands for Relational Database Management System.
- RDBMS is a program used to maintain a relational database.
- RDBMS is the basis for all modern database systems such as MySQL, Microsoft SQL Server, Oracle, and Microsoft Access.
- RDBMS uses **SQL queries** to access the data in the database.

Some terms related to RDBMS

- **Domain:** A domain describes the set of possible values for a given attribute. Eg, Age must be INT.
- **Attributes:** Attribute is the name of a column in a relation. The attributes of the student may be Roll-no , Name, Age etc.
- **Tuples:** A tuple is a complete row of the table . It is also called a record.
- **Null Value:** It means value for the corresponding attribute is not available (empty) or not assigned.

Guiding Principles Used in database design.

- Reduced Redundancy(Data duplication)
- **Openness:** RDBMS uses SQL for its operation. SQL(Structured Query Language) is a common database language used in most databases.
- Sharing Of Data
- Security Of the Data

Homework : Explain these given points and try to understand it.

Flat File Database

- In Flat file database we could access and manipulate only one table at a time. Each table was stored in its own file. so, these table was simply called a file or just the database. This type of database using a single file for a table is called **flat file database**.
- It was used early.

Entity, Attribute and Relationship

- **Entity:** Entity is a name of things or a person or object in the real world that is distinguishable from all other objects. Example: student, book, movie.
- **Entity Set:** An entity set in DBMS is a set that collectively represents a group of entities of a similar type. Example: set of Loan means house loan, car loan , personal loan etc.
- **Attribute:** The properties that characterize an entity set are called its attributes. Example: The possible attributes of an entity customer can be customer_name, customer_age, customer_email etc

continue

- **Relationship:** The association between entities is called relationship.

example: A teacher **Teaches** students.

Types of Cardinality(Binary relationship)

- **One-to-one** : An entity in A is associated with at most one entity in B and an entity in B is associated with at most one entity in A.
- Example Husband and Wife

Types of Cardinality(Binary relationship)continue

- **One-to-many** : An entity in A is associated with any number of entities in B . and an entity in B , however can be associated with at most one entity in A .
- Example Teacher and Course

Types of Cardinality(Binary relationship)continue

- **many-to-one:** An entity in A is associated with at most one entity in B. and an entity in B, however can be associated with any number of entities in A.
- Example minster relation with mayor , Police Head, Local bodies.

Types of Cardinality(Binary relationship)continue

- **many-to-many:** An entity in A is associated with any number of entities in B. and an entity in B can be associated with any number of entities in A.
- Example students relation with subjects(course). Many student can study many subjects

Keys

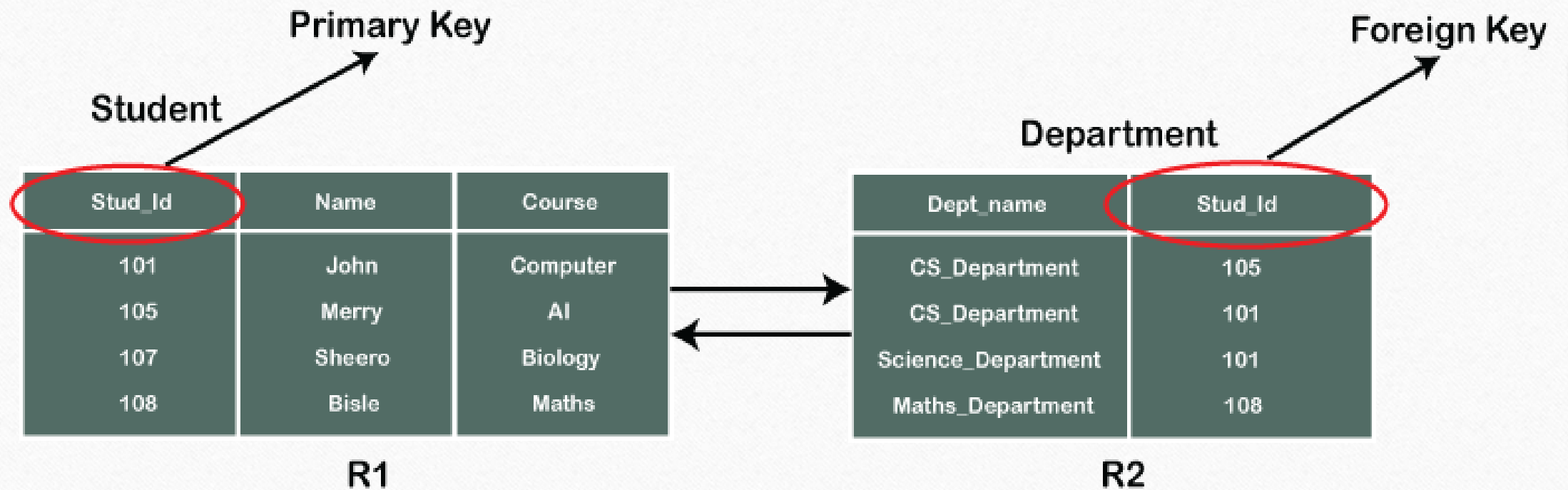
- A **Key** is an attribute that is used to identify a particular record in a database.
- Types of Keys:
 1. **Candidate Key:** All attributes combinations inside a relation that can be used to uniquely identify required record are candidate keys. A relation can have multiple candidate keys.
 2. **Primary Key:** A primary key is a candidate key chosen by a database designer to identify entity from entity set. Although several candidate key exists, one of the candidate key is selected to be primary key.

Keys

Types of Keys:

- 3. **Super Key:** A super key is a set of one or more attributes that taken collectively, allows us to identify uniquely an entity in the entity set. Example: Super key for students can be {student-id + student-name} also.
- 4. **Alternate Key:** In the case of two or more candidate keys , only one key is used as a primary key. An **alternate key** is the candidate key which is not used as a primary key.
- 5. **Foreign Key :** A foreign key is used to represent the relationship between two tables. A foreign key is a key attribute whose value is derived from the primary key of another table.




Foreign Key Example



E-R Diagram

- An ER diagram is a graphical representation of different entities and their relationship.

Components used in ER diagram are as follows:

- **Rectangle:**  It represents the entity.
- **Oval or ellipse:**  It represents attributes of entities.
- **Diamond:**  It represents the relationship among entities.

E-R Diagram

Components used in ER diagram are as follows:

- **Line:** — It is used to link attributes to entity sets and entity set to relationship and vice-versa.

Advantages Of ER diagram in system design

- It provides a visual representation of the database design.
- It is simple and easy to understand.
- It is highly flexible to use as additional relationships can be added as per need.
- It provides higher level description of the system.

Normalization

- Normalization divides the larger table into smaller and links them using relationships.
- Normalization is the process of breaking down the larger tables into smaller ones to reduce data duplication and different anomalies.
- Normalization is used to minimize the redundancy from a relation or set of relations. It is also used to eliminate undesirable characteristics like Insertion, Update, and Deletion Anomalies

Advantages of Normalization

- Normalization helps to minimize data redundancy.
- Much more flexible database design.
- It avoids the loss of information.
- It improves faster sorting and index creation.
- It improves the performance of the system.
- It simplifies the structure of tables.

Why do we need Normalization?

- Dependence between the data is identified.
- Redundancy in the database is minimized.
- The data model is more flexible and easier to maintain.
- Improves database design.
- Removes anomalies from database.
- Normalization consists of a series of guidelines that helps to guide you in creating a good database structure.

Types of Normal Forms

- **Normalization** works through a series of stages called Normal forms.
- **De-normalization** produces a lower normal form that is in, a 3rd Normal Form would be converted to a 2nd normal form through de-normalization.

Types of Normal Forms

- First Normal Form (1NF)
- Second Normal Form (2NF)
- Third Normal Form (3NF)
- Fourth Normal Form (4NF)
- Boyce Codd Normal Form (BCNF)

• First Normal Form (1NF)

- A relation is in 1NF if it contains an atomic value. There should not be any repeating groups of an attribute.
- It eliminate repeating groups.

- First Normal Form (1NF)

-
- **Example:** Relation EMPLOYEE is not in 1NF because of multi-valued attribute EMP_PHONE.

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

First Normal Form (1NF)

- The decomposition of the EMPLOYEE table into 1NF has been shown below:

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab
12	Sam	8589830302	Punjab

• Second Normal Form (2NF)

- A relation will be in 2NF if it is in 1NF and all non-key attributes are fully functional dependent on the primary key
- It eliminate partial functional dependency.

- Second Normal Form (2NF)

-
- **Example:** Let's assume, a school can store the data of teachers and the subjects they teach. In a school, a teacher can teach more than one subject.

TEACHER_ID	SUBJECT	TEACHER_AGE
25	Chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	Computer	38

• Second Normal Form (2NF)continue

- In the given table, non-prime attribute TEACHER_AGE is dependent on TEACHER_ID which is a proper subset of a candidate key. That's why it violates the rule for 2NF

Second Normal Form (2NF)continue

- To convert the given table into 2NF, we decompose it into two tables:
- Teacher-Detail Table

TEACHER_ID	TEACHER_AGE
25	30
47	35
83	38

Second Normal Form (2NF)continue

- Teacher-Subject Table

TEACHER_ID	SUBJECT
25	Chemistry
25	Biology
47	English
83	Math
83	Computer

• Third Normal Form (3NF)

- A relation will be in 3NF if it is in 2NF and no transitive dependency exists.
- It eliminates transitive dependency.
- Mathematically, among 3 attributes A, B & C. if $A \rightarrow B$, $B \rightarrow C$
- then $C \rightarrow A$.

• Third Normal Form (3NF) continue

Employee Detail Table:

EMP_ID	EMP_NAME	EMP_ZIP	EMP_STATE	EMP_CITY
222	Harry	201010	UP	Noida
333	Stephan	02228	US	Boston
444	Lan	60007	US	Chicago
555	Katharine	06389	UK	Norwich
666	John	462007	MP	Bhopal

• Third Normal Form (3NF) continue

- **Super key in the table above:**
- {EMP_ID}, {EMP_ID, EMP_NAME}, {EMP_ID, EMP_NAME, EMP_Z IP}....so on
- **Candidate key:** {EMP_ID}
- **Non-prime attributes:** In the given table, all attributes except EMP_ID are non-prime.

• Third Normal Form (3NF) continue

- Here, EMP_STATE & EMP_CITY dependent on EMP_ZIP and EMP_ZIP dependent on EMP_ID. The non-prime attributes (EMP_STATE, EMP_CITY) transitively dependent on super key(EMP_ID). It violates the rule of third normal form.
- That's why we need to move the EMP_CITY and EMP_STATE to the new <EMPLOYEE_ZIP> table, with EMP_ZIP as a Primary key.

Third Normal Form (3NF) continue

- **EMPLOYEE** table:

EMP_ID	EMP_NAME	EMP_ZIP
222	Harry	201010
333	Stephan	02228
444	Lan	60007
555	Katharine	06389
666	John	462007

Third Normal Form (3NF) continue

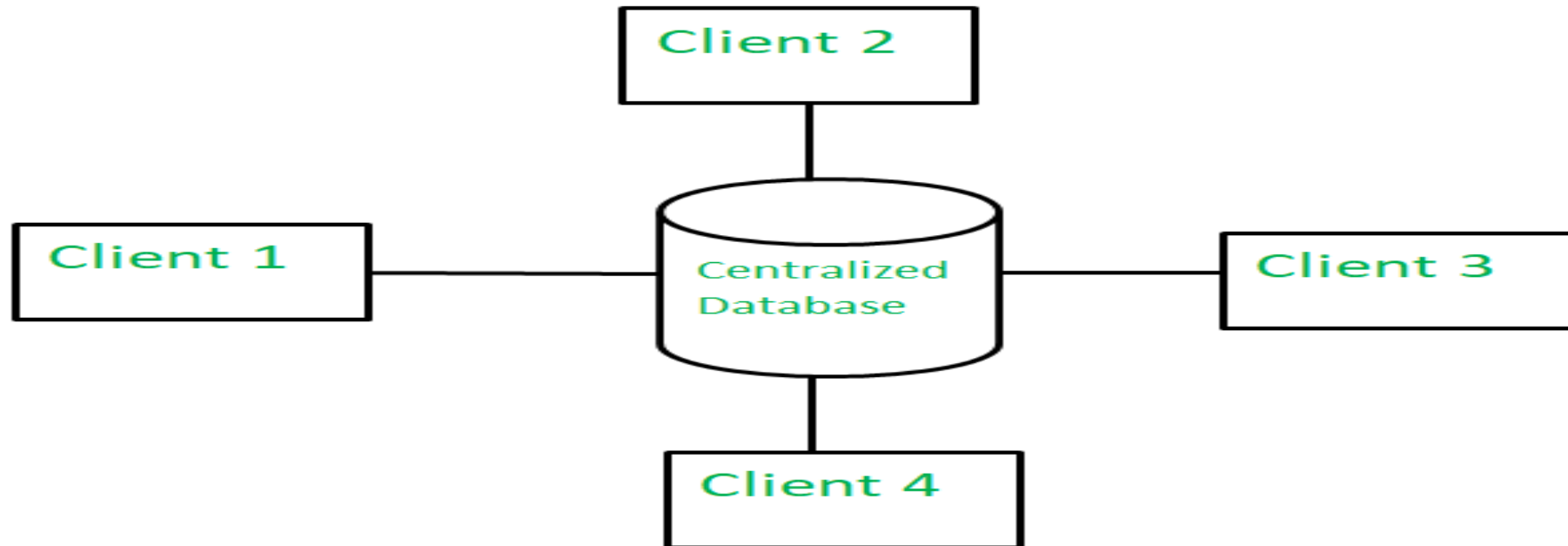
- **EMPLOYEE_ZIP** table:

EMP_ZIP	EMP_STATE	EMP_CITY
201010	UP	Noida
02228	US	Boston
60007	US	Chicago
06389	UK	Norwich
462007	MP	Bhopal

Centralized vs Distributed Database

- **Centralized**
- A centralized database is basically a type of database that is stored, located as well as maintained at a single location only. This type of database is modified and managed from that location itself.

- Centralized



Advantages Of centralized Database

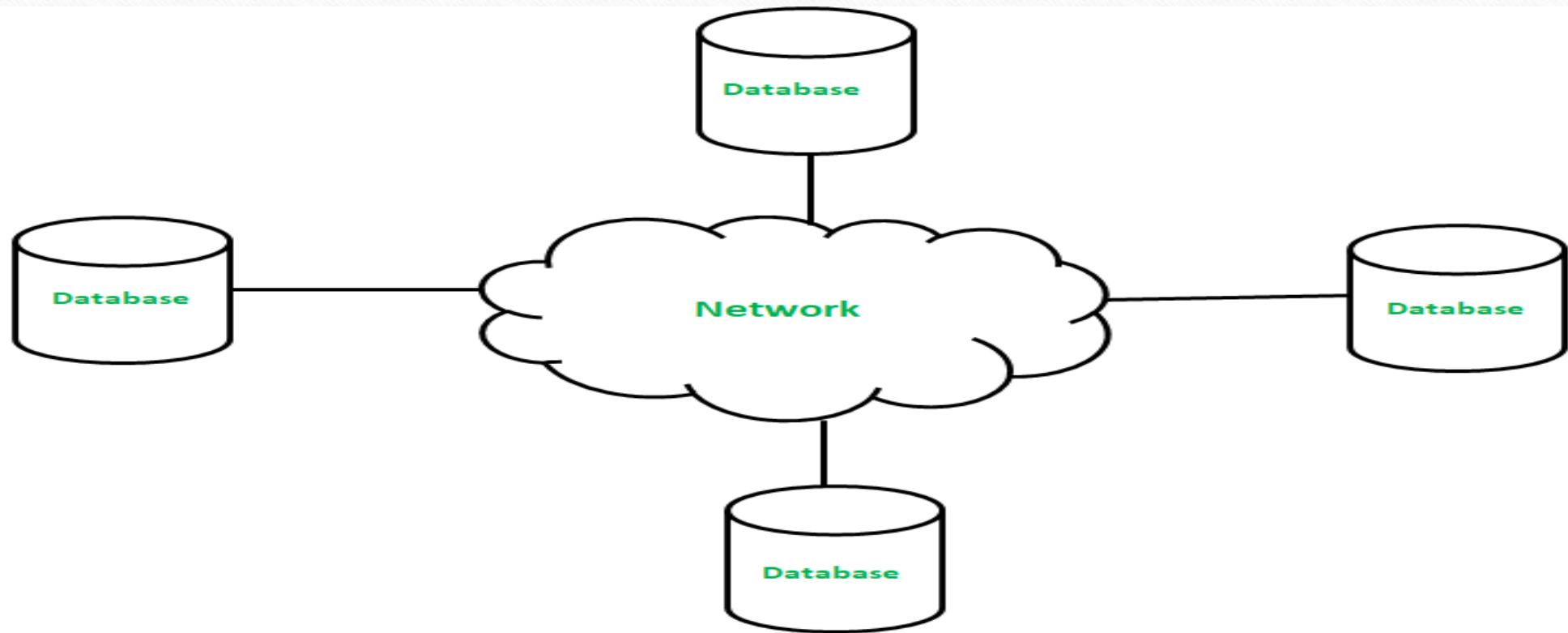
- Since all data is stored at a single location only thus it is easier to access and coordinate data.
- The centralized database has very minimal data redundancy since all data is stored in a single place.
- It is cheaper in comparison to all other databases available.

Dis-advantages Of centralized Database

- If any kind of system failure occurs in the centralized system then the entire data will be destroyed.
- It is not suitable for big organization.

Distributed Database

- A distributed database is basically a type of database which consists of multiple databases that are connected with each other and are spread across different physical locations.



Advantages of Distributed Database

- This database can be easily expanded as data is already spread across different physical locations.
- The distributed database can easily be accessed from different networks.
- This database is more secure in comparison to a centralized database.
- If any kind of system failure occurs in the server then the entire data is not affected.

Dis-advantages of Distributed Database

- This database is very costly and is difficult to maintain because of its complexity.
- It has high risk of data theft , hacking.

Difference between Centralized and Distributed Database

- Do it by yourself.

Data Security

- Database security(Data Security) is the technique that protects and secures the database against intentional or accidental threats. It is concerned with the protection of data from the unauthorized access, alteration or deletion of data or destruction of data.

Misuse of data

- **1.Accidental loss of data**
- Crashes during transaction processing
- Logical errors in the programs
- Due to the distribution of data over several computers.

Misuse of data

- **2.Intentional loss of data**
- Unauthorized reading of data.
- Unauthorized modification of data
- Unauthorized destruction of data.

Security Measures to protect the database.

- To protect the database, following security measures should maintained at different levels
- 1. Physical level
- 2. Human
- 3. Operating System
- 4. Network
- 5. Database system.

The main guidelines for data security.

- Always maintain backup schedule
- Spread your database across several folder or disks
- Verify the data integrity of your database regularly.
- **Some techniques used for different level of security in computer systems.**
- **Cryptographic** tools can be used.(to reduce the chance of data from modifying while sending the data)
- Use strong authentication or password, and Use antivirus software etc.

Some other terms related to DBMS

- **Data Abstraction:** It is a process of hiding unwanted or irrelevant details from the end user. System hides certain details of how the data are stored and maintained.

Mainly there are three levels of **data abstraction** for DBMS, which are as follows

- **Physical or Internal Level:** It is the lowest level of abstraction for DBMS which defines how the data is actually stored, it defines data-structures to store data and access methods used by the database.
- **Logical or Conceptual Level:** Logical level is the intermediate level or next higher level. It describes what data is stored in the database and what relationship exists among those data.
- **View or External Level:** It is the highest level and describes only a part of the entire database. Users interact with the system through view level.

Database Administrator(DBA)

- The **data administrator** is the person who makes the strategic and policy decisions regarding the data of the enterprise , and the **DBA** is the person who provides the necessary technical support for implementing those decisions.

Role or functions of DBA

- Defining the internal schema.
- Defining the external schema.
- Defining security and integrity.
- Defining system Backup
- Monitoring performance and overall system.

Data Integrity

- Data integrity or integrity constraint is a rule that restricts the values that may be present in the database.
- The different types of data integrity are as follows:
- 1. **Entity integrity:** it is the rule that no column that is a part of a primary key may accept null values.
- 2. **Referential integrity:** it says that essentially all the information in the two fields should match. Referential integrity is a series of processes that ensure data remains stored and used in a uniform manner. Database structures are embedded with rules that define how foreign keys are used. This can prevent data duplication and guarantee data accuracy. Foreign key of table A should match with the primary key of another table.

Data Integrity

- **3.Domain Integrity:** Domain integrity is a series of processes that guarantee the accuracy of pieces of data within a domain. A domain is classified by a set of values that a table's columns are allowed to contain, along with constraints and measures that limit the amount, format, and type of data that can be entered. Eg: INT for Age, VARCHAR for name, etc

Thank You ...

...