Unit 1

Database Management System

Introduction to data

- Data is the plural of datum.
- Data can be defined as any raw fact, figures about person, place, event or thing which can be recorded in computer as in the form of number, text, picture, audio, video etc.
- It provides information when processed.
- It may or may not have meaning.
- e.g. "Computer" "22" "Ram" etc.

Information

- The refined or processed data to fulfill any objective of particular user is called information.
- Information gives value to the data.
- It is very meaningful and useful. It enables us to make right decision at right time.
- Data becomes information when processed.

Database

- A database is an organized collection of inter-related data.
- It can be considered as set of logically related data files organized that can be used or accessed by one or more application.
- Database is a file composed of records with a set of operations for searching, sorting, recombining and other functions.
- Database contains only data not information, so it should be analyzed or processed to get required information.

Database

- A database consists of 4 major elements:
  - o **data element**: stored logical entities or binary representation
  - o **relationship**: relation between various data
  - o **constraints**: predicates that define correct database states
  - o **schema**: outline, diagram, or model that describes the structure of database

DBMS

- DBMS stands for Database Management System
- It is collection of programs that manages the database structure and controls access to the database.
- It is a application software that manages, organizes and adds various functionality to database
- DBMS serves as interface between user and database.
- It hides much of the database's internal complexity from application programs and users
- e.g. MySQL, Microsoft Access, PostgreSQL, SQL Server, FileMaker, Oracle, dBASE etc.

Advantages of DBMS

- Data Sharing
  - multiple users have better access to more and better managed data
- Data Security
  - provides framework for better enforcement of data privacy and security policies
- Data Integration
  - combining different sources and providing unified view
- Minimized data inconsistency
  - data redundancy is much reduced. always makes consistent state

Advantages of DBMS

- Improved Data Access
  - quick data access through SQL queries
- Decision Making
  - generate better quality information on which better decisions are based.
- Increase end user productivity
  - transform data into usable information and empowers users to make better decision
- Backup and Recovery
  - data can be backed up and recover easily

Dis-advantages of DBMS

- Complexity of Backup and Recovery
  - need complex networks, RAID technologies
- Increased installation and management cost
  - need skilled manpower, annual maintenance cost, security costs
- Additional Hardware and Software cost
  - need extra hardware and software resources
- Requirement of new and specialized manpower
  - changing technology requires new skilled manpower
- Increased complexity
  - need database designs, administration, complex functionality

Structured Query Language (SQL)

- It is the universal standard database language for accessing and managing data in database.
- A query is the request to DBMS for the retrieval, modification, insertion and deletion of the data from the database.
- SQL was first known as 'SEQUEL' Structured English Query Language and was implemented at IBM research for relational database.
- Different companies have used SQL. e.g. Microsoft used T-SQL for SQL server, Oracle used P-SQL etc.
- SQL is non-procedural language. i.e. it just specifies what is to be done rather than how it is to be done.
- SQL has statements for data definition, data manipulation and data control.

Features of SQL

- It is non-procedural language.
- It is English like language
- It can process a single record as well as sets of records.
- It is 4GL.
- It hides the underlying complexity and structures of database.
- It has 3 different types of statements: DDL, DML, DCL

Some SQL commands

- CREATE TABLE command
  - Syntax: CREATE TABLE table_name;
  - Example: CREATE TABLE student_info (roll number(4), Name char (20), city char (20));

Some SQL Commands

- INSERT INTO command
  - Syntax: INSERT INTO table_name;
  - Example: INSERT INTO student_info (roll, name, city) VALUES (1, "Ram Bahadur", "Kathmandu");

Some SQL

- UPDATE command
  - Syntax: UPDATE table_name SET column_name = expression;
  - Example: UPDATE student_info SET address='Bkt' WHERE roll_no = 3;

Some SQL

- SELECT command
  - Syntax: SELECT * FROM table_name;
  - Example: SELECT * FROM student_info;
  - SELECT roll, name FROM student_info;

Some SQL

- Sort Data
  - Syntax: SELECT * FROM table_name ORDER BY column1, column2;
  - Example: SELECT * FROM student_info ORDER BY roll_no;

Some SQL

- DELETE command
  - Syntax: DELETE FROM table_name WHERE column_name = expression
  - Example: DELETE FROM student_info WHERE roll_no = 5;

DCL

- DCL stands for Data Control Language.
- They are used to change the permission or role of database user.
- Data privileges are passed to users by Administrator user.
- e.g. GRANT, DENY, REVOKE etc.

DDL

- DDL stands for Data Definition Language.
- They are used to create, modify or delete databases and database objects like table, data types, views etc.
- e.g. CREATE, ALTER, DROP etc.

DML

- DML stands for Data Manipulation Language.
- They are used to insert, update, delete, and retrieve the data.
- e.g. SELECT, INSERT, UPDATE, DELETE etc.

Some Related Terms

• Domain

o It is a set of permissible values for an attribute or field.

o Every field has specific domain

o Domain can be defined by assigning a data type format, range of values through rules or constraints.

Some Related Terms

- Attributes

o They are properties that describes the entity.

o The column in relational table represents attributes.

o Simple: limited to single attribute

o composite: contains sub attributes for the same parent attribute

o Single valued: that can contain only one value

o Multivalued: that can contain multiple values

Some Related Terms

- Tuples

o Any row in relational table is known as tuple.
o Each tuple in table represents instance or record of particular entity.
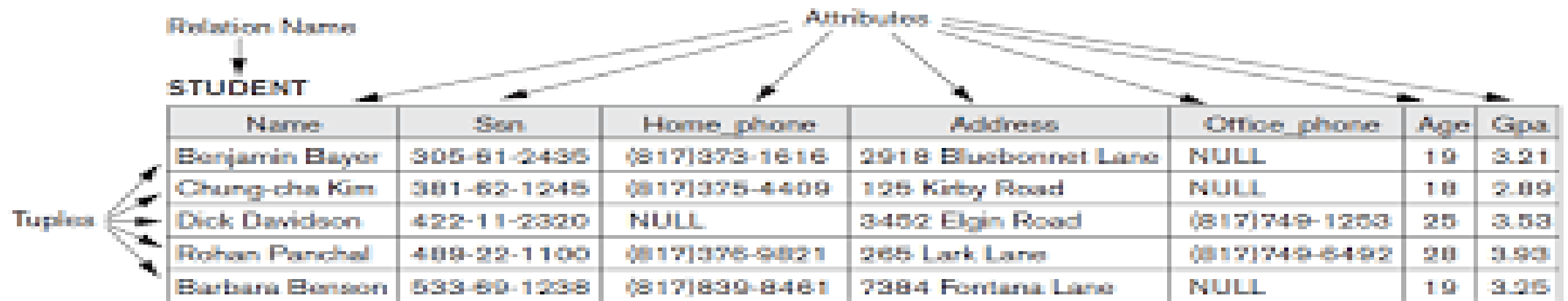
Figure 3.1
The attributes and tuples of a relation STUDENT.

File Processing System

- In this system, data is stored in separate files.
- In file based system, a set of programs is provided to facilitate the users in organizing, creating, deleting, updating and manipulating their files which forms a file management system.
- Commonly used two systems are
  - Sequential Files
    - Records and information are stored and accessed in sequence.
  - Direct/Random Files
    - Records are physically organized in such a way that make it possible for users to directly access a record based on its key field value.

Database Model

- The description of database is called database schema and it is specified during database design.
- The process of designing the database schema is called data modeling.
- Data modeling describes structure of database such as no. of tables, data types, relationship between them, constraints etc.

Database Model

- Database model defines the manner in which the various files of database are linked together.
- It is also called the structure of database
- It is method of organizing data and represents logical relationships among data elements in the database.

Purpose of data modeling

o Design fully normalized database
o Reduce data redundancy/repeat
o Improve faster sorting and indexing
o Improve performance of database
o Establish proper relationship between the tables
o Make database reliable and secure

Types Of Database Model

- Hierarchical Model
- Network Model
- Relational Model
- Entity Relation Model
- Object Oriented Model

Hierarchical Model

- It is the oldest data model. It uses one-to-many relationship for data elements.

- It arranges the files used in the database in a top-down structure which is similar to upside down tree.

- It links number of separate elements to one parent element or record.

Hierarchical Model
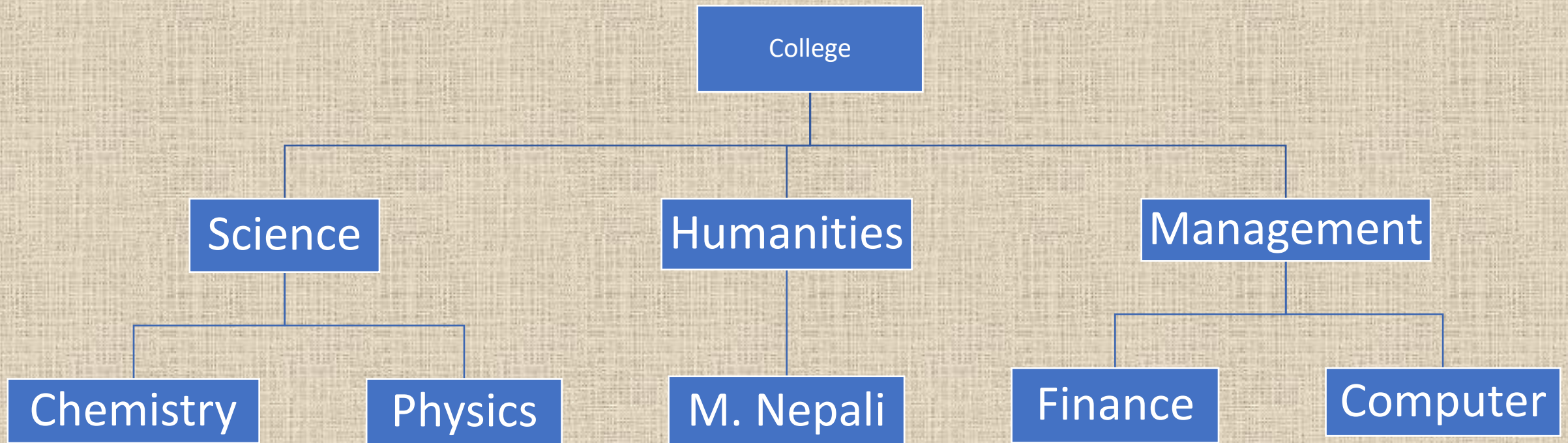
Hierarchical Model

Advantages

- It is the easiest model of database
- Searching is fast and easy if parent is known
- It is very efficient in handling one-to-many relationship

Disadvantages

- It is old and outdated database model
- Modification and addition of the child node is very hard i.e. it is non-flexible.
- It can't handle many-to-many relationship.
- Redundancy is increased

Network Model

- It is similar to hierarchical model but is more flexible to access the data.
- In this model a child element can have more than one parent or no parent at all.
- The model is viewed as a graph where relationship types are arcs and object types are nodes.
- Charles Bachman was the original inventor of the network model.

Network Model
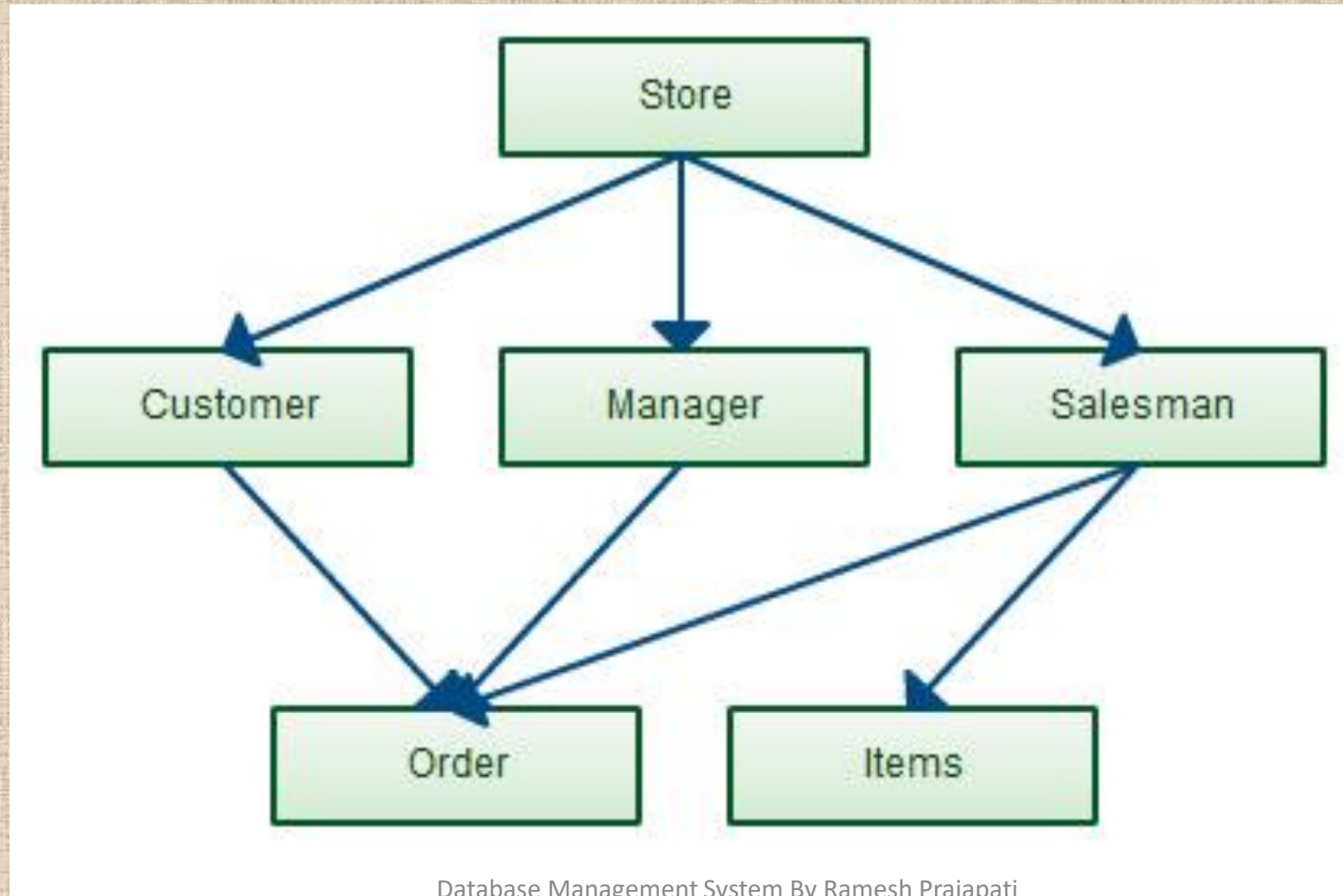
Network Model

Advantages

- More flexible than hierarchical model because it accepts many to many relationships.

- Reduces data redundancy because data is not stored in more than one location.

- Searching is faster because of multi-dimensional pointers.

Disadvantages

- Complex to design

- Needs long program to handle the relationships

- Less secure than hierarchical model because data can be accessed from any parents

Relational Model

- It is most widely used model nowadays
- data is maintained in the form of two dimensional table
- information is stored in the form of row and columns
- tables are also called relations

Relational Model

TABLE

Column ↓

Roll_No	Name	Age	GPA
1	Arya	21	4
2	Bran	19	3
3	John	24	4.3
4	Max	24	1

Attributes →

Tuple →

Relational Model

Advantages

- structural independence: changes in table structure do not affect data access or application program

- Tabular view improves simplicity, easier design, implementation and use

- hides or isolates end user from physical level design complexity
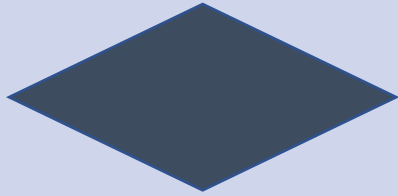
Disadvantages

- RDBMS requires hardware and software overhead

- skilled or trained man power is required

- may promote "islands of information" : individuals and departments can easily develop their own application

ER-Model

- Entity Relationship models are normally represented in an entity relationship diagram (ERD)
- Graphical representations to model database
- It can be easily changed to relational tables later.
- Entities are described by set of attributes. which are represented by OVAL shape.

ER-Model

- Entity:

- It is defined as anything about which data are to be collected and stored. It is represented by RECTANGLE

- Relationship:

- They describe associations among data or entities. There are 3 types of relationships
  - one-to-one (1:1)
  - one-to-many(1:M)
  - many-to-many(M:N)

Symbol	Representation	Description
	Rectangle	Represents entity
	Ellipse/Oval	Represents attributes
	Diamond	Represents relationship
	Line	Represents link

ER model

Advantages

- Visual modeling results conceptual simplicity
- effective communication tool
- can be easily integrated with relational model

Disadvantages

- limited constraint representation
- limited relationship representation
- no data manipulation language
- more space required to represent entities and attributes graphically

E-R Diagram for Library Management System
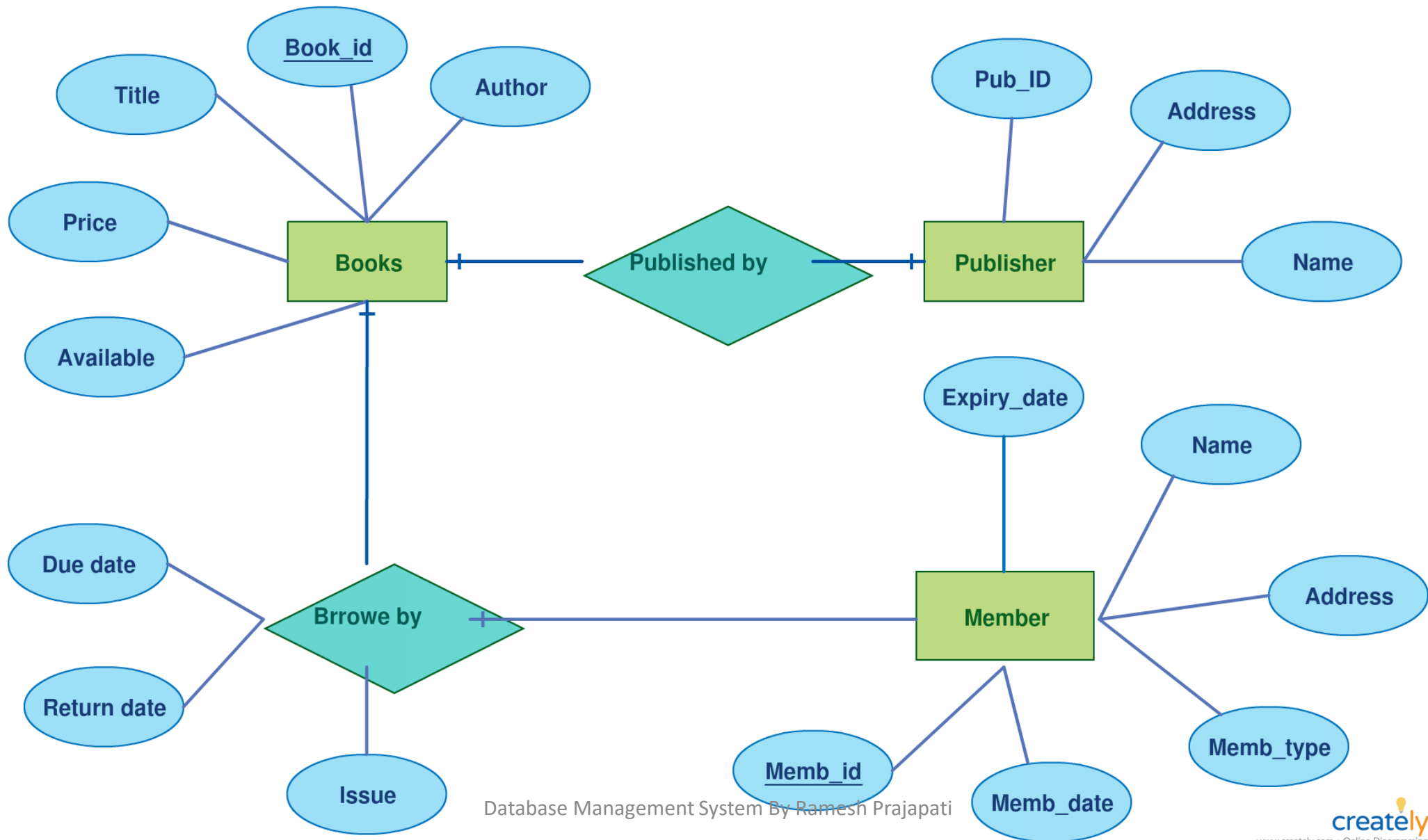
Object Oriented Model

- Both data and relationships are contained in a single structure known as an object.
- Object includes information about relationships between the facts within the object and with other objects.
- Includes data, various relationships, operational procedures
- Basic building blocks for autonomous structures.

Object Oriented model

Advantages

- Semantic content is added (proper meaning is added)
- Inheritance promotes data integrity

Disadvantages

- Complex navigation system
- Steep learning curve
- High system overhead slows transactions

keys

- Key is a field that uniquely identifies the records, tables or data.
- It allows us to establish relation between multiple tables.
- Useful for finding unique records or combination of records from large records

Keys

- Primary Key
- A Primary Key (PK) is one or more columns in a table used to uniquely identify each row in the table.
- Primary key cannot be null.
- There must be one and only primary key in table.
- Foreign Key
- Foreign key represent relationships between tables.
- Column whose value is derived from the primary key of some other table.

Keys

- Unique key

- It is one or more columns in a table used to uniquely identify each row in table.

- Unique key con not have duplicate values.

- Super key

- Super key is set of one or more attributes that allows us to identify uniquely a tuple in the relation.

Keys

- Candidate Key
- If a relation has more than one key, each key is called candidate key.
- All key which satisfy the condition of primary key.
- They can be used as PK if required
- Alternate Key
- All the remaining candidate key beside the primary key are known as alternate key
- Alternate key is also known as secondary key

Primary Key · **Candidate Key** · **Foreign Key**

TABLE-2

TABLE-1

EmpId	EmpName	EmpLicence	EmpPassport	DId
1001	Matt	LC1201	MA100LC	1
1002	Maxy	LC2078	XY100LC2	2
1003	Roy	LK00928	LK100RO	3

DId	Designation
1	BPO
2	Account
3	IT

Primary Key · **Candidate Key** · **Alternate Key**

Alternate Key

Super Key · **Unique Key**
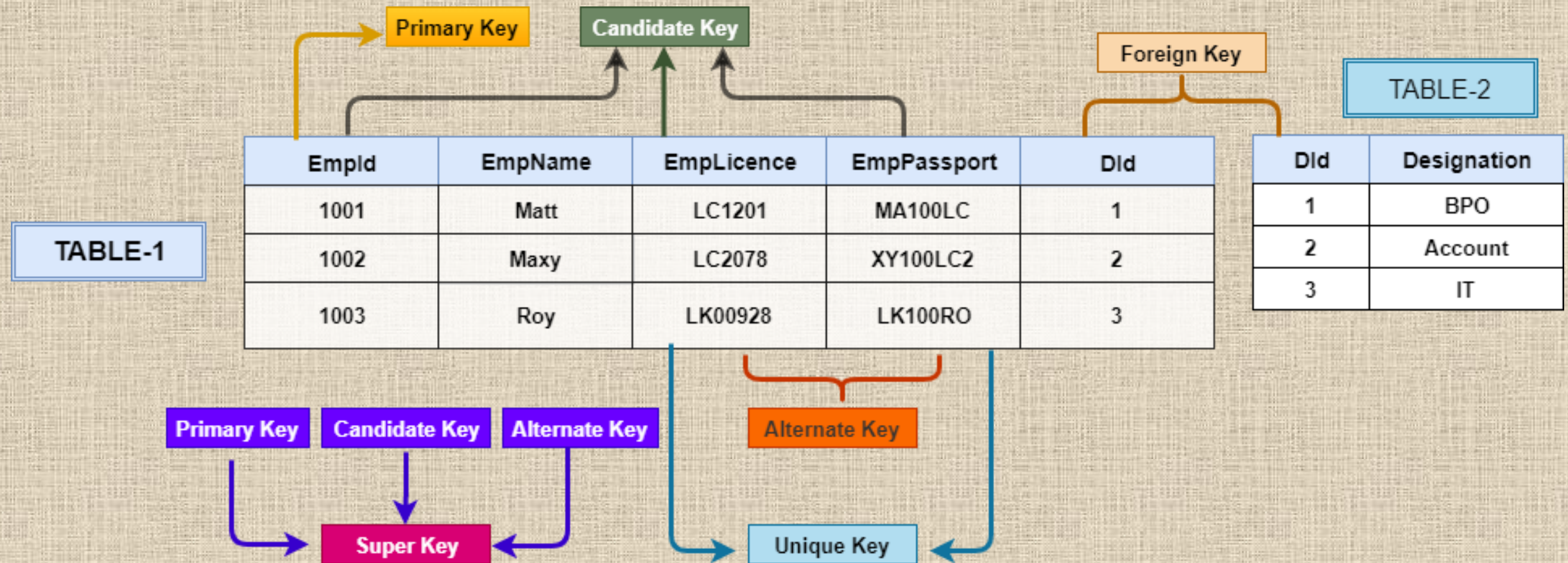
Normalization

- It is the process of organizing data in a database to reduce the redundancy and inconsistency.

- Normalization is the process of breaking down a single big table into many smaller tables with fewer fields.

- Edgar F. Codd originally defined the first three normal forms.

- A fully normalized record consists of:
  - A primary key that identifies an entity
  - A set of attributes that describes the entity

Advantages and disadvantages

Advantages:

- It reduces the data redundancy
- It improves faster sorting and index creation.
- It improves performance of database
- It simplifies the structures of tables
- It avoids data loss.

Disadvantages:

- It is complex to design relationship between tables
- It requires more joins to get desired results
- More maintenance needed due to larger number of tables.

Denormalization

- It is the reverse process of normalization.
- It is the process of designing fewer tables with more fields in a database.

Rno	Name	Address	Age	Sub_id	Subjects	Teacher	TAddress
1.	Ram	Bkt	19	S1, S2, S3	C, VB, Java	C-sir, VB-sir, Java-sir	C-zone, VB-zone, Java-zone
2.	Shyam	Ktm	18	S1,S3	C, Java	C-sir, Java-sir	C-zone, Java-zone
3.	Hari	Ptn	20	S3	Java	Java-sir	Java-zone

First Normal Form (1NF)

- A relation is said to be in First Normal Form only if each attribute of the relation is atomic. i.e. each column contain only single value and each row must contain same columns.

- Rules for 1NF:
  - It should only have single (atomic) valued attributes/columns
  - Values stored in a column should be of the same domain
  - All the columns in a table should have unique names.
  - The order in which data is stored does not matter.

First Normal Form (1NF)

Rno	Name	Address	Age	Sub_id	Subjects	Teacher	TAddress
1	Ram	Bkt	19	S1	C	C-sir	C-zone
1	Ram	Bkt	19	S2	VB	VB-sir	VB-zone
1	Ram	Bkt	19	S3	Java	Java-sir	Java-zone
2	Shyam	Ktm	18	S1	C	C-sir	C-zone
2	Shyam	Ktm	18	S3	Java	Java-sir	Java-zone
3	Hari	Ptn	20	S3	Java	Java-sir	Java-zone

Second Normal Form (2NF)

- A table is said to be in Second Normal Form (2NF) if both the following conditions hold:

- Table must be in First Normal Form (1NF).

- There should be no Partial Dependency.

Second Normal Form (2NF)

Rno	Name	Address	Age	Sub_id
1	Ram	Bkt	19	S1
1	Ram	Bkt	19	S2
1	Ram	Bkt	19	S3
2	Shyam	Ktm	18	S1
2	Shyam	Ktm	18	S3
3	Hari	Ptn	20	S3

Sub_id	Subjects	Teacher	TAddress
S1	C	C-sir	C-zone
S2	VB	VB-sir	VB-zone
S3	Java	Java-sir	Java-zone

Third Normal Form (3NF)

- A database is in third normal form if it satisfies the following conditions:

- It is in second normal form

- There is no transitive functional dependency

By transitive functional dependency, we mean we have the following relationships in the table: A is functionally dependent on B, and B is functionally dependent on C. In this case, C is transitively dependent on A via B.

Third Normal Form (3NF)

Rno	Name	Address	Age
1	Ram	Bkt	19
2	Shyam	Ktm	18
3	Hari	Ptn	20

tid	Teacher	TAddress
I	C-sir	C-zone
II	VB-sir	VB-zone
III	Java-sir	Java-zone

Rno	Sub_id
1	S1
1	S2
1	S3
2	S1
2	S3
3	S3

Sub_id	Subjects
S1	C
S2	VB
S3	Java

Sub_id	tid
S1	I
S2	II
S3	III

Normalize the following

bk_id	title	ISBN	price	auth_id	author	auth_email	pub_id	Pub_name	pub_address
201	Computer Science-XII	987-654-321	450	3001	Shankar Ghimire	shankar@gmail.com	5001	Kriti Publication	Putalisadak, Kathmandu
202	Principal of Accounting	963-852-741	325	3005	Tej Aryal	tej@gmail.com	5001	Kriti Publication	Putalisadak, Kathmandu
203	Artificial Intelligence	951-632-874	350	3001	Shankar Ghimire	shankar@gmail.com	5003	Ekta Publication	Dillibazar, Kathmandu
204	Database Management System	789-456-123	425	3002	Devendra Chapagain	Devendra@gmail.com	5002	Buddha Publication	Byasi, Bhaktapur
205	Cyber Security	852-846-357	425	3002	Devendra Chapagain	Devendra@gmail.com	5003	Ekta Publication	Dillibazar, Kathmandu

Centralized Database

- Centralized database works in a client-server system
- Centralized database server computer hosts the database service and all other client devices in network access them
- Only central server is responsible for all the processing
- Data retrieval is easy and secure
- All the rules and security mechanism can be applied only to server
- Clients never process data
- Suitable for small organization

Centralized Database

Advantages

- Low Setup cost and high performance
- Data centralization
- Easy to manage and manipulate
- Highly secure as single DBA have all control
- Suitable for small organization
- Easy access of data

Disadvantages

- Not suitable for larger organization, only limited coverage area
- Location dependent; cannot be accessed from other place
- Does not support globalized connection

Centralized Database
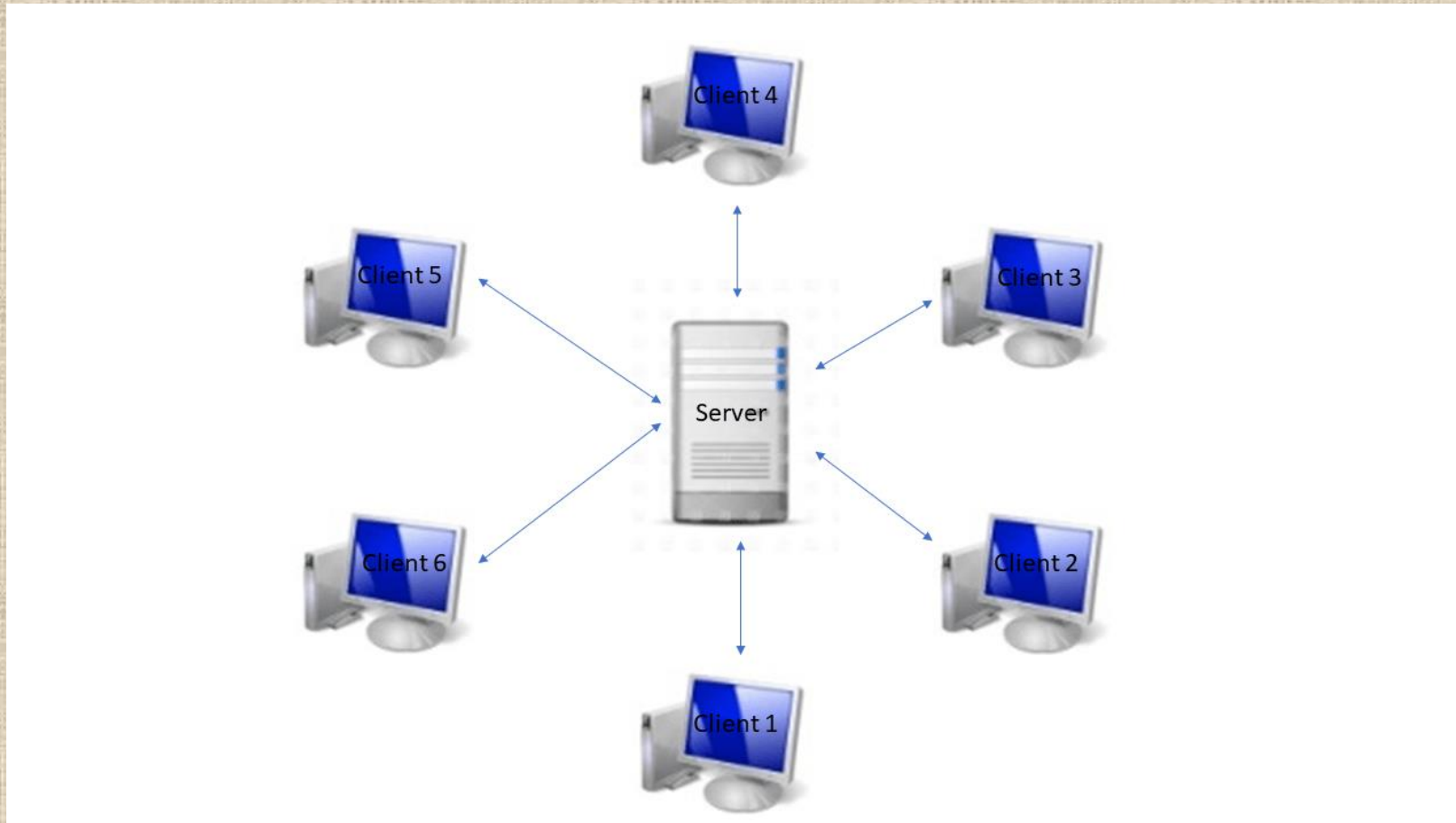
Distributed Database

- Distributed database system is a set of databases stored on multiple computers that appears to applications as a single database
- User can simultaneously access and modify the data in several databases in a network
- Computers in network communicate through high speed bus or lines
- Data is stored in several set of computer under the control of distributed DBMS
- Geographically separate locations but connected through internet
- Interconnects heterogeneous systems and useful for large organizations
- Data security is much more crucial
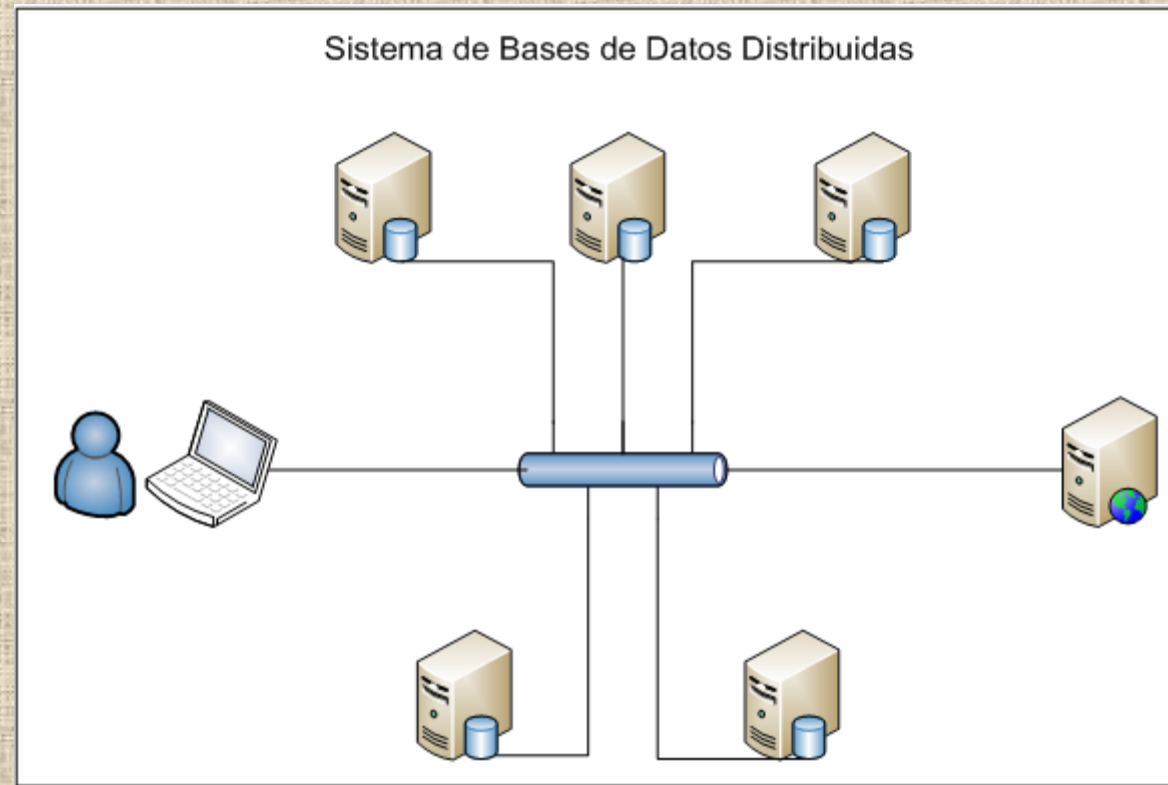
Distributed Database

Advantages

- Data sharing and distributed control all over the world
- Improved reliability for users
- Improved availability of data
- Low cost on operation and data sharing
- Scalable: Modular growth can be supported

Disadvantages

- High software development cost
- Threats for bugs and hacking of data
- Processing overhead
- More complex database design required

Distributed Database

Sistema de Bases de Datos Distribuidas

Centralized Vs Distributed Database

Centralized database

- It is managed by single machine or system

- It is easy to maintain

- It is less efficient

- The entire data gets lost if server crashes

- It has slow response

- It has high communication cost

Distributed Database

- It is spread and split up across various storage devices

- It is difficult to maintain

- It is more efficient

- User are still able to access other databases if one fails

- It has fast response

- It has low communication cost

Database Security

- Data stored in database should always remain safe, correct and authentic.
- Information security is the protection of information against unauthorized disclosure, alteration or destruction
- It ensures only 'right people' get the right to access the 'right data'
- It ensures confidentiality of data
- Data security also covers physical security such as data files protection from heat, dust, power failure, theft etc.
- The important objective of data security is to prevent data loss, misuse of data and disclosure of secret data.

Methods of Data Security

- Regular backup of database in disks, tapes, optical disks etc
- Using password login system to prevent data from unauthorized access
- Specifying the roles to every user and granting appropriate permission to them
- Making physical prevention by using voltage stabilizer, UPS to supply regular power
- Keeping system in safe room with proper security and supervision
- Implementing software protections like antivirus, firewalls, antispyware software

Thank you

End of Chapter 1