

# Course Contents

## Unit-9 Fundamentals of Database(4 Hrs.)

- Introduction to Database
- Advantages or benefits of database
- Disadvantages of database
- File–Oriented approach and Database approach
- Data Model
- Database System
- Database Management System
- Database System Architectures
- Database Applications
- Introduction to Data Warehousing, Data Mining and Big Data

# Fundamentals of Database

## Introduction to Database

Database is basically nothing more than an computer based record keeping system its overall purpose is to maintain & control the records. DBMS manages the database. E.g. banks, hospitals, universities etc.

# Database

A database consists of data, entity, attribute, information, relationships, constraints & schema.

## Entity

- Entities are things which can be distinguishable. As for example, bank ,customer,student,teacher,account etc.

## Attribute

- Attributes are the properties of these entities.

## Relationship

- The association between two or more entities are called relationship.

# Database

## Constraints

are predicates that define the correct database states where the schema describes the organization of data & relationships within database.

# Database

## Schema

Schema defines the various views of the database for DBMS components use & for application security.

Schema is two types

Internal schema

- Defines how & where data's are organized in a physical data storage.

External schema

- Defines a views of the database for particular users

# Database

## Advantages or benefits of database using DBMS

- Redundancy can be reduced.
- Inconsistency can be avoided(removed).
- Data can be shared.
- Standards can be enforced.
- Security restrictions can be applied.
- Integrity(accuracy) can be maintained.

# Database

## Disadvantages of database using DBMS

- DBMS system is costly.
- Specialized person required.
- When one system failure then another system is inactive.
- Backup & recovery operations are fairly complex in a DBMS environment.

# DBMS VS File System

DBMS	File System
<b>DBMS is a collection of data. In DBMS, the user is not required to write the procedures.</b>	File system is a collection of data. In this system, the user has to write the procedures for managing the database.
<b>DBMS gives an abstract view of data that hides the details.</b>	File system provides the detail of the data representation and storage of data.
<b>DBMS provides a crash recovery mechanism, i.e., DBMS protects the user from the system failure.</b>	File system doesn't have a crash mechanism, i.e., if the system crashes while entering some data, then the content of the file will be lost.
<b>DBMS provides a good protection mechanism.</b>	It is very difficult to protect a file under the file system.
<b>DBMS contains a wide variety of sophisticated techniques to store and retrieve the data.</b>	File system can't efficiently store and retrieve the data.
<b>DBMS takes care of Concurrent access of data using some form of locking.</b>	In the File system, concurrent access has many problems like redirecting the file while other deleting some information or updating some information.

# Database

## Data Models:

- The basic structure or design of the database is the data model. A data model is a collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints. Some data models are given below:

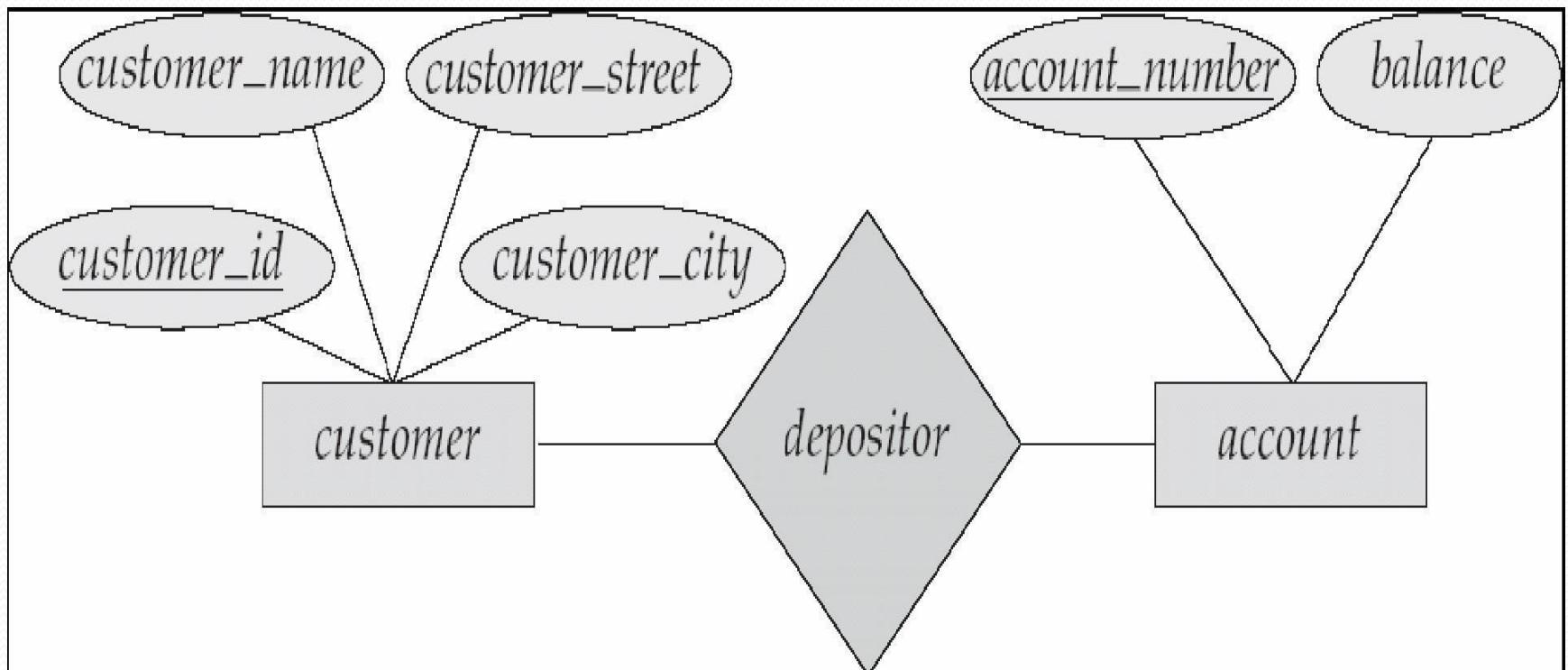
# Entity-relationship(E-R) Model

- An Entity-Relationship Model represents the structure of the database with the help of a diagram. ER Modelling is a systematic process to design a database as it would require you to analyze all data requirements before implementing your database.
- ER diagrams or ERD's are composed of three main elements: entities, attributes and relationships.

# Entity-relationship(E-R) Model

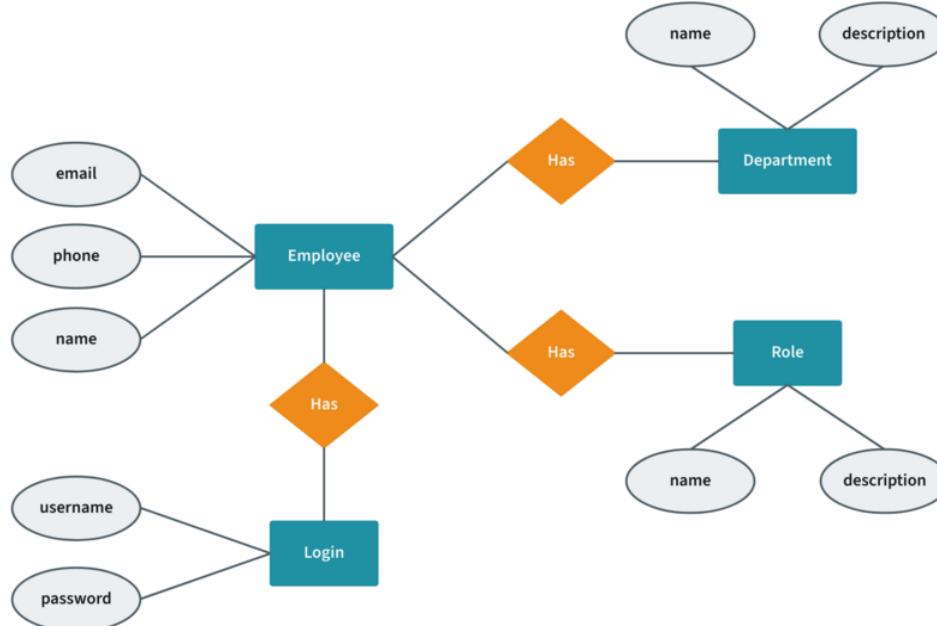
- **Entities** - typically displayed in a rectangle, entities can be represented by objects, persons, concepts or events that contain data.
- **Attributes** - displayed in a circle or an oval, the attributes refer to the characteristics of an entity. They can be categorized as simple, composite or derived and an object can have one or multiple attributes.
- **Relationships** - illustrate how two or more entities interact with each other. They are displayed as labels placed on the lines connecting the objects.

# Entity-relationship(E-R) Model



# Entity-relationship(E-R) Model

**Employee Management System ERD**



# Relational Model

- It is the current favorite model. The relational model is a *lower level model that uses a* collection of tables to represent both data and relationships among those data. Each table has multiple columns, and each column has a unique name. Each table corresponds to an entity set or relationship set, and each row represents an instance of that entity set or relationship set.

<i>customer_id</i>	<i>customer_name</i>	<i>customer_street</i>	<i>customer_city</i>
192-83-7465	Johnson	12 Alma St.	Palo Alto
677-89-9011	Hayes	3 Main St.	Harrison
182-73-6091	Turner	123 Putnam Ave.	Stamford
321-12-3123	Jones	100 Main St.	Harrison
336-66-9999	Lindsay	175 Park Ave.	Pittsfield
019-28-3746	Smith	72 North St.	Rye

(a) The *customer* table

<i>account_number</i>	<i>balance</i>
A-101	500
A-215	700
A-102	400
A-305	350
A-201	900
A-217	750
A-222	700

(b) The *account* table

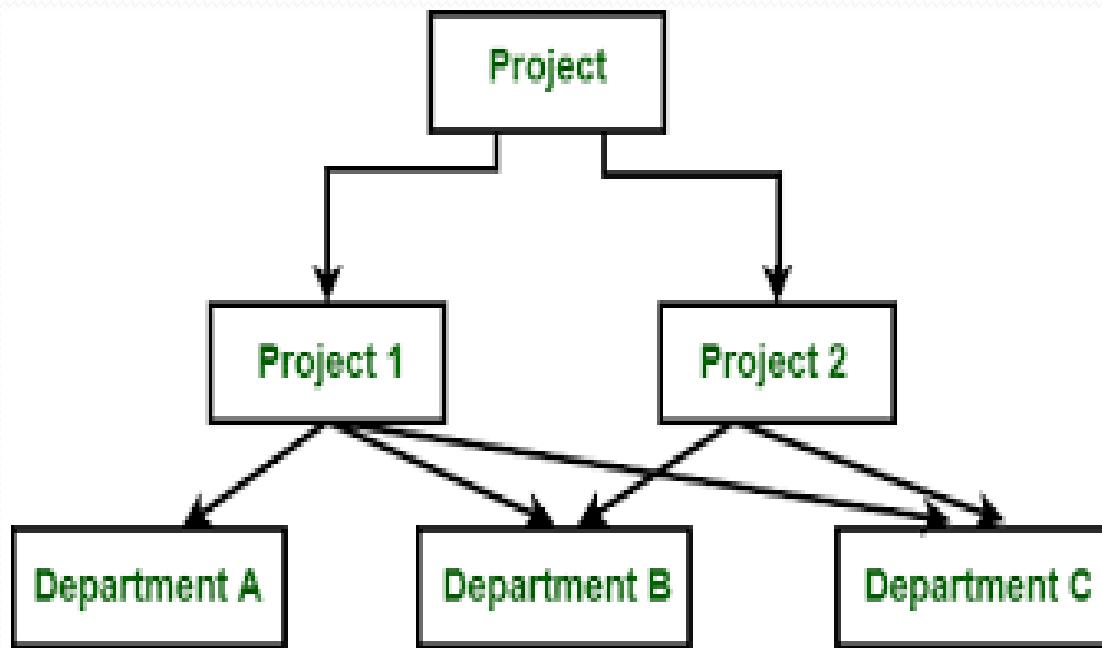
<i>customer_id</i>	<i>account_number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

(c) The *depositor* table

## Network Model

In network model, data are represented by records and relationships among data are represented by links. Network Database was developed with the views to eliminate the problems faced by hierarchical database during 1980's. A network database consists of a collection of records connected to one another through links. A relationship in this context is known as a set a link is an association between two records.

# Network Model

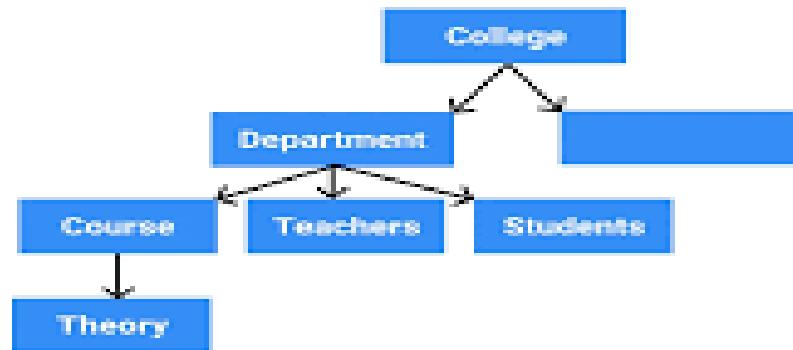


## Hierarchical Model

A hierarchical database model is a **data model in which the data are organized into a tree-like structure**. The data are stored as records which are connected to one another through links. A record is a collection of fields, with each field containing only one value.

# Hierarchical Model

## Hierarchical Model



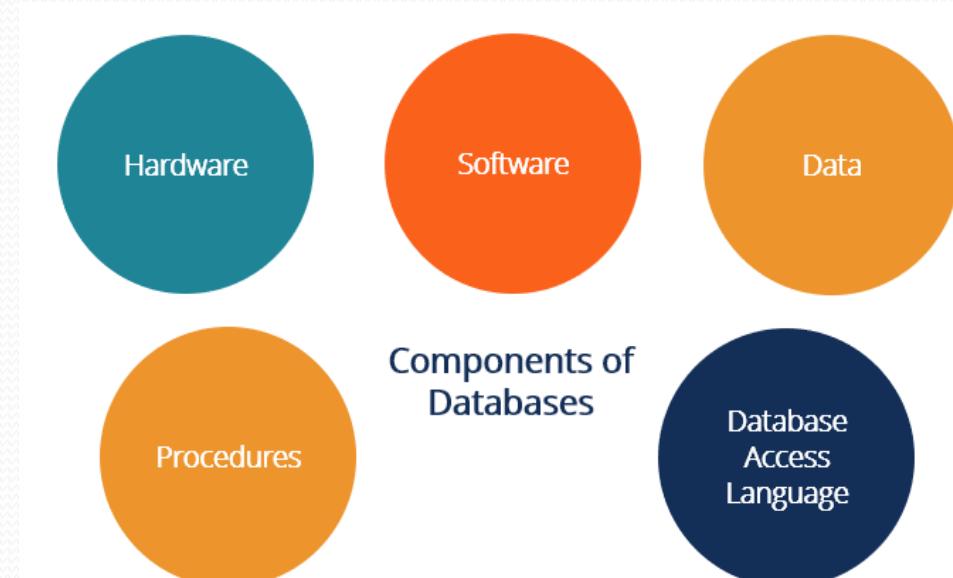
# Database system

## Introduction

A database is an organized collection of structured information or data, typically stored electronically in a computer system. A database is usually controlled by a database management system(DBMS).

# Database system

## Components



# Database system

## Components

### 1. Hardware

- Hardware refers to the physical, electronic devices such as computers and hard disks that offer the interface between computers and real-world systems.

### 2. Software

- Software is a set of programs used to manage and control the database and includes the database software, operating system, network software used to share the data with other users, and the applications used to access the data.

# Database system

## Components

### 3. Data

- Data are raw facts and information that need to be organized and processed to make it more meaningful. Database dictionaries are used to centralize, document, control, and coordinate the use of data within an organization. A database is a repository of information about a database (also called metadata).

# Database system

## Components

### 4. Procedures

- Procedures refer to the instructions used in a database management system and encompass everything from instructions to setup and install, login and logout, manage the day-to-day operations, take backups of data and generate reports.

### 5. Database Access Language

- Database Access Language is a language used to write commands to access, update, and delete data stored in a database. Users can write commands using Database Access Language before submitting them to the database for execution. Through utilizing the language, users can create new databases, tables, insert data and delete data.

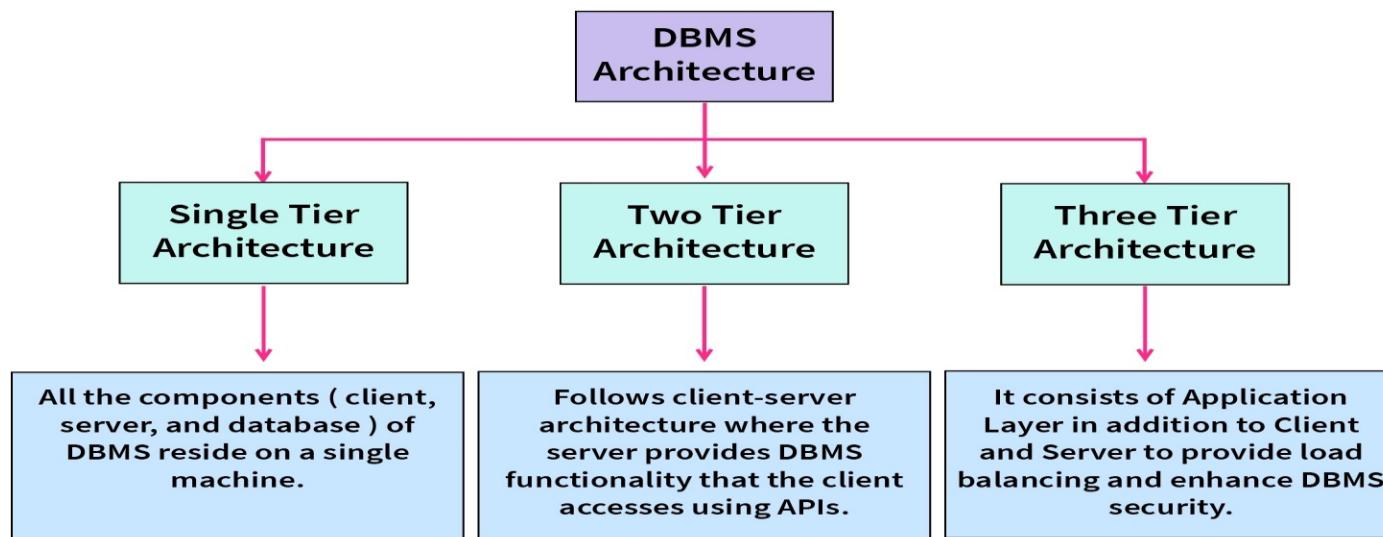
# Database system

## Architecture

- DBMS architecture describes the structure and how the users are connected to a specific database system. Architecture affects the performance of the database. In the single-tier architecture, all the components of the DBMS reside on a single server. It is the most straightforward approach.

# Database system

## Architecture



# Database Management System(DBMS)

A DBMS is a software that provides service for accessing a database , where maintaining the data. Some examples of DBMS are MS - Access, Oracle, MY-SQL, sys base etc.

# Database Languages

DBMS provides two languages.

## 1. Data Definition Language(DDL)

Database language that is used to create, delete or modify database schema is called data definition language(DDL).

- Create statement
- Drop statement
- Alter statement

# Database Language

As for example,

CREATE TABLE account

(

Account-number CHAR(10),

Balance INTEGER,

Branch CHAR(10)

)

# Database Language

## 2. Data Manipulation Language(DML)

Database language that enables insert, delete, update and retrieve data from database is called data definition language(DML).

- Insert statement
- Delete statement
- Update statement
- Select statement

As for example,

Insert into Employee values("A001",1000,"Patan");

# Database Administrators

**Database administrators (DBAs)** use specialized software to store and organize data. The role may include capacity planning, installation, configuration, database design, migration, performance monitoring, security, troubleshooting as well as backup and data recovery.

## Roles and Responsibility of DBA

- **Installing and upgrading the DBMS Servers:** – DBA is responsible for installing a new DBMS server for the new projects. He is also responsible for upgrading these servers as there are new versions that come into the market or requirement. If there is any failure in the up-gradation of the existing servers, he should be able to revert the new changes back to the older version, thus maintaining the DBMS working. He is also responsible for updating the service packs to the DBMS servers.

# Roles and Responsibility of DBA

- **Design and implementation:** – Designing the database and implementing is also DBA's responsibility. He should be able to decide on proper memory management, file organizations, error handling, log maintenance etc. for the database.
- **Security:** – DBA is responsible for creating various database users and roles and giving them different levels of access rights.
- **Migrate database servers:** – Sometimes, users using oracle would like to shift to SQL server. It is the responsibility of DBA to make sure that migration happens without any failure and there is no data loss.

# Roles and Responsibility of DBA

- **Backup and Recovery:-** Backup and recovery maintained by DBA and has to be maintained him. This is one of the main responsibilities of DBA. Data/objects should be backed up regularly so that if there is any crash, it should be recovered without much effort and data loss.
- **Performance tuning:** – Since the database is huge and it will have lots of tables, data, constraints, and indices, there will be variations in the performance from time to time. Also, because of some designing issues or data growth, the database will not work as expected. It is the responsibility of the DBA to tune the database performance. He is responsible to make sure all the queries and programs work in a fraction of seconds.

# Roles and Responsibility of DBA

- **Documentation:** – DBA should be properly documenting all his activities so that if he quits or any new DBA comes in, he should be able to understand the database without any effort. He should basically maintain all his installation, backup, recovery, security methods. He should keep various reports about database performance.

# Database Design

1. Requirements collections & analysis
2. Conceptual database design
  - E-R diagram , DBMS-independent
3. Choice of DBMS
  - MS - Access, Oracle , Sys base
4. Data model mapping(logical database design)
  - To convert E-R diagram to relations or tables & define relationship, enforce referential integrity.
5. Physical design
  - To define the storage structure. E.g. physical storage of data
6. Implementation

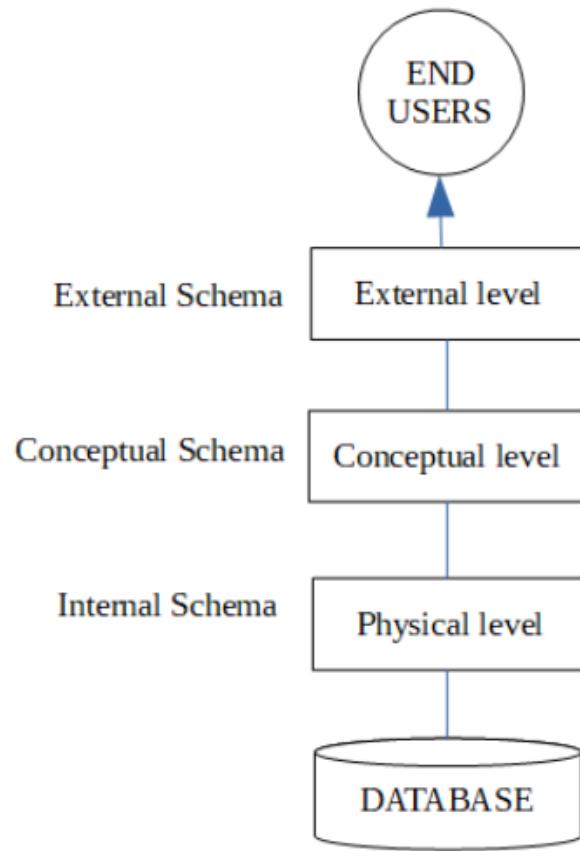
# Database Design

OR

Database design or database levels are

- Conceptual database design(External level)
- Logical database design(Internal level)
- Physical database design

# Database Design



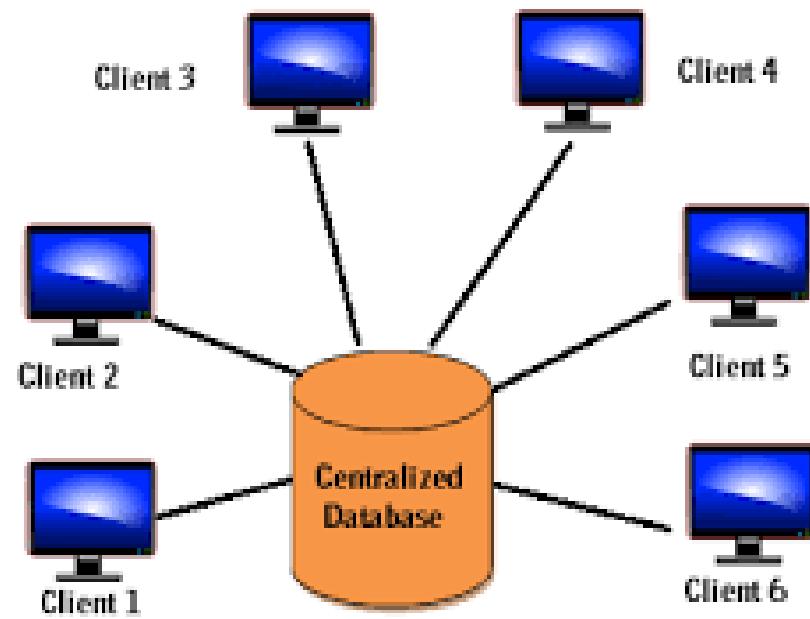
# Database System Architectures

- Centralized database
- Distributed database
- Client/Server database

# Centralized Database

- A centralized database is a **database that is located, stored, and maintained in a single location**. This location is most often a central computer or database system, for example a desktop or server CPU, or a mainframe computer.

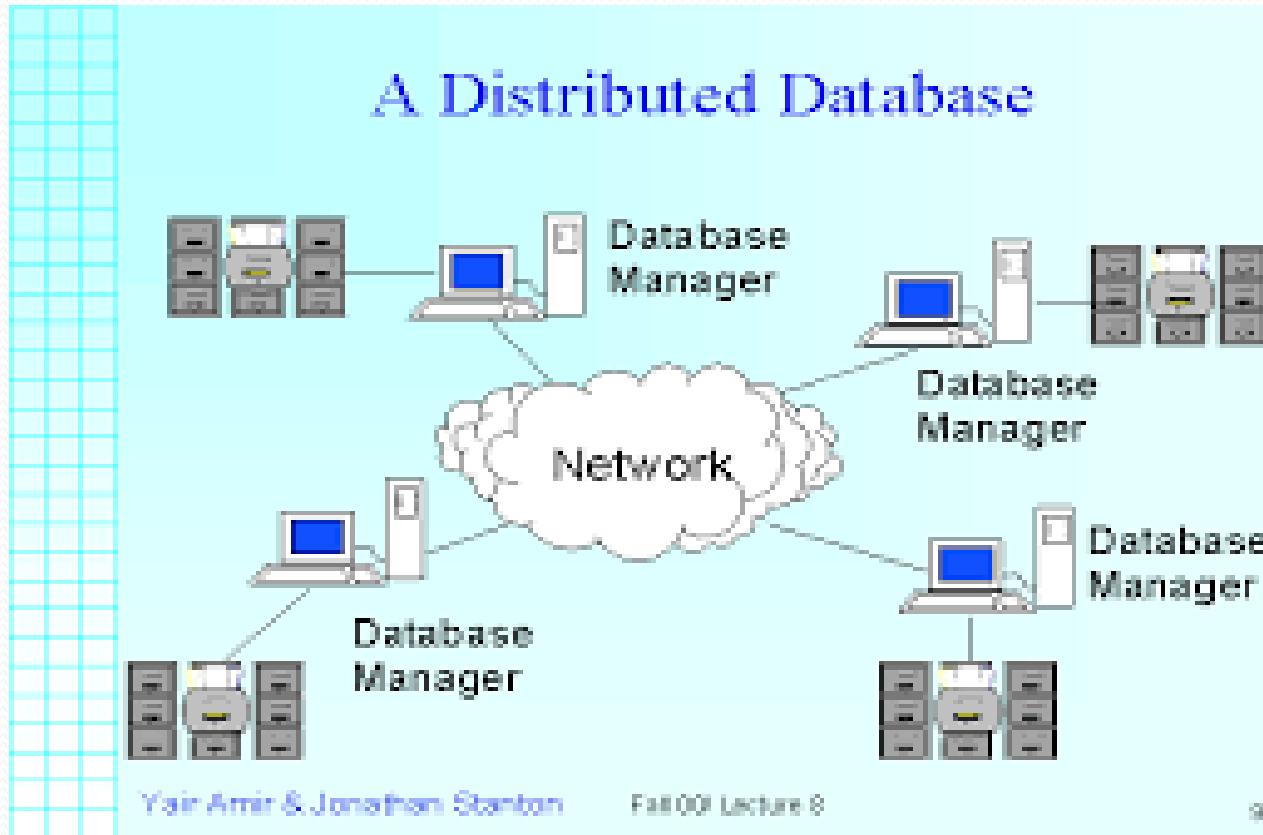
# Centralized Database



# Distributed Database

- A distributed database is a **database that consists of two or more files located in different sites either on the same network or on entirely different networks**. Portions of the database are stored in multiple physical locations and processing is distributed among multiple database nodes.

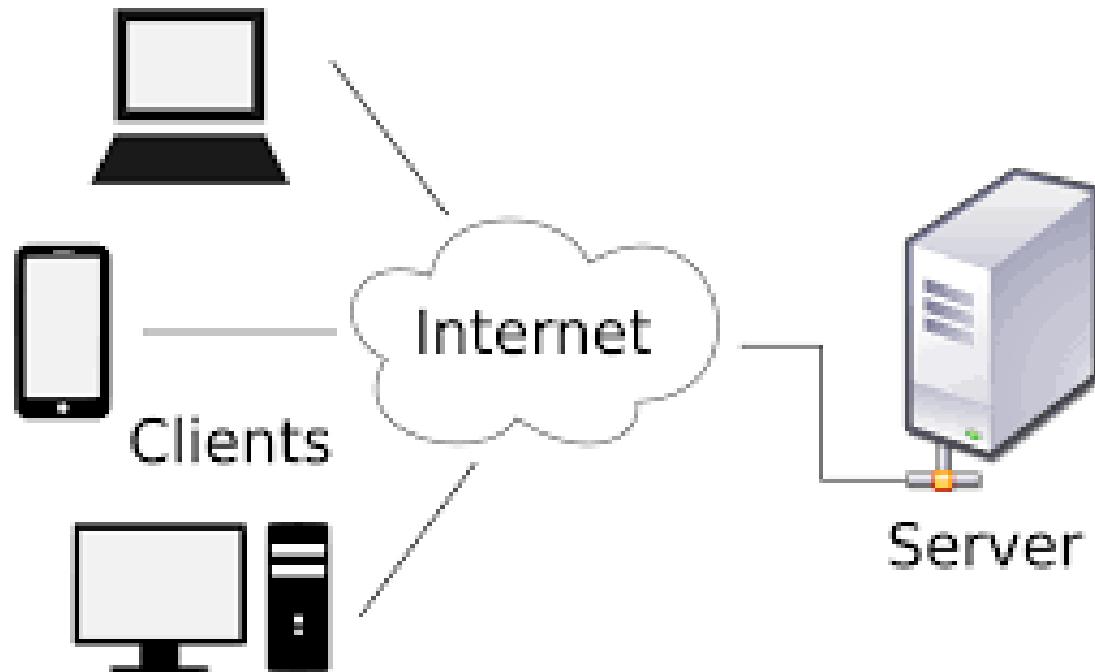
# Distributed Database



# Client/Server Database

- Client-server DBMS (database management system) refers to an architectural paradigm that separates database functionality between client machines and servers.

# Client/Server Database



# Database Applications

Different database applications are given below:

- Banking: For customer information, accounts, and loans and banking transactions.
- Airlines: For reservation and schedule information.
- Universities: For student information, course registrations, and grades.
- Credit card transactions: For purchase on credit cards and generation of monthly statements.

# Database Applications

- Telecommunication: For keeping records of call made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.
- Finance and Accounting: For storing information about holdings, sales, purchase of financial instruments such as stocks and bonds, debits and credits.
- Sales: For customer, product and purchase information.

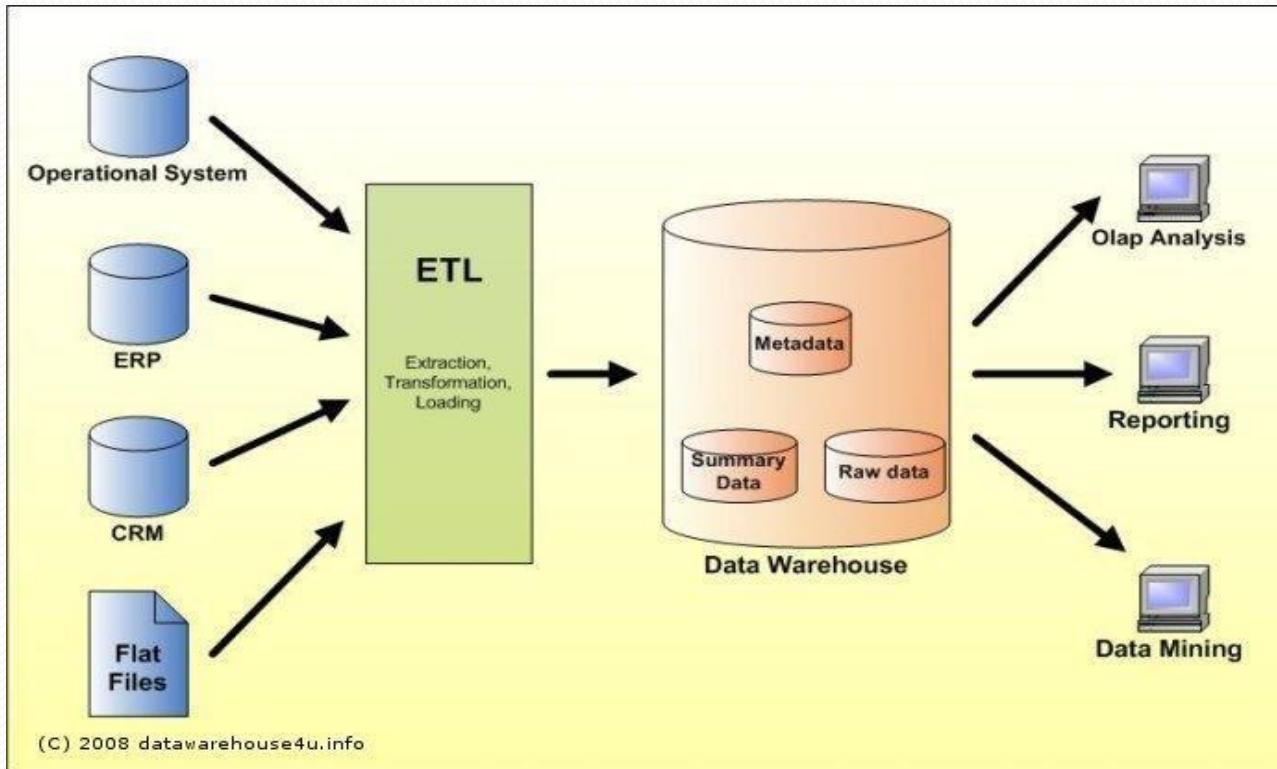
# Database Applications

- Manufacturing: For management of supply chain and for tracking production of items in factories, inventories of items in warehouses/stores and orders for items.
- Human resources: For information about employees, salaries, payroll taxes and benefits and for generation of paychecks.

## Data warehouse

- Data warehouse is defined as a process of centralized data management & retrieval. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access & analysis. Dramatic advances in data capture ,processing power ,data transmission & storage capabilities are enabling organizations to integrate their various databases into data warehouses.

# Data warehouse



## Data Marts

is a subset of an enterprise –wide data warehouse , which typically supports an enterprise element( dept, region ,functions etc.). A data mart is a simple form of data warehouse focused on a single subject or line of business. With a data mart, teams can access data and gain insights faster, because they don't have to spend time searching within a more complex data warehouse or manually aggregating data from different sources.

## Advantages of data warehouse

- More cost – effective decision – making.
- Better enterprise intelligence.
- Enhanced customer service.
- Business reengineering.
- Information system reengineering.

## Disadvantages of data warehouse

- Underestimation of data loading resources. Often, we fail to estimate the time needed to retrieve, clean and upload the data to the warehouse.
- Hidden problems in source systems.
- Data homogenization.

# Uses of data warehouse

- Standard reports & queries.
- Queries against summarized data.
- Data analysis

# Data Mining

Generally data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives & summarizing it into useful information (information that can be used to increase revenue, cuts costs or both). Data mining software is one of the number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles. Data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

# Evolution of Data Mining

From the user point of view the four steps are necessary.

- Data collection(1960)
- Data access(1980) by using SQL ,ODBC
- Data warehousing & Decision support(1990) by using OLAP(Online Analytical Processing).

## Advantages of data mining

- Automated prediction of trends & behaviors.
- Automated discovery of previously unknown patterns.
- Databases can be larger in both depth & breadth.

# Technologies used in data mining

1. Neural network
2. Rule induction
3. Evolutionary programming
4. Case based reasoning(CBR)
5. Decision tree
6. Genetic algorithms
7. Non linear regression methods

# Technologies used in data mining

## 1. Neural network

Non – linear predictive models that learn through training & resemble biological neural networks in structure.

## 2. Rule induction

The extraction of useful information by using rules from data based on statistical significance.

## 3. Evolutionary programming

The underlying idea of the method is that the system automatically formulates hypothesis about the dependence of the target form of programs expressed in an internal programming language

# Technologies used in data mining

## 4. Case based reasoning(CBR)

To forecast a future situation or to make a correct decision such systems find the closest past analogs of the present situation.

## 5. Decision tree

Decision tree is like a tree structure which makes a set of decisions.

## 6. Genetic algorithms

Genetic algorithm is a techniques that use processes such as genetic combination, mutation & natural selection during evolution.

# Technologies used in data mining

## 7. Nonlinear regression methods

These methods are based on searching for a dependency of the target variable on the other variable. The obtained formula is more suitable for analysis & interpreting in principle. In this way , data mining is the natural evolution of query & reporting tools. Every one , who creates queries & reports & then analyzing the data.

## Big Data Analysis

The Next Frontier Nii Ayi Armah, Canadian Economic Analysis current analysis is heavily dependent on data. The more timely, accurate and relevant the data, the better our assessment of the current state of economic activity. f Technological advancements have provided an opportunity to exploit digital data from business transactions, social media and networked computers. The combination of all of these data is called “big data.” Analysis of the vast quantities of digital information contained in big data can offer fresh insight for the monitoring of economic activity and inflation. Moreover, the timeliness of big data could improve real-time decision making for monetary policy.

# Big Data Analysis

- The potential of big data is, however, limited by challenges related to methodological constraints, a lack of easy access to the data and privacy concerns. By providing an assessment of the present state of the economy, current analysis<sup>1</sup> contributes to the Bank of Canada's long-term macroeconomic projections, which in turn help to inform monetary policy decisions. Immediate and complete information about every economic and financial transaction within a country would improve current analysis by facilitating accurate and timely measurement of important macro economic indicators. Unfortunately, this ideal data set does not exist.