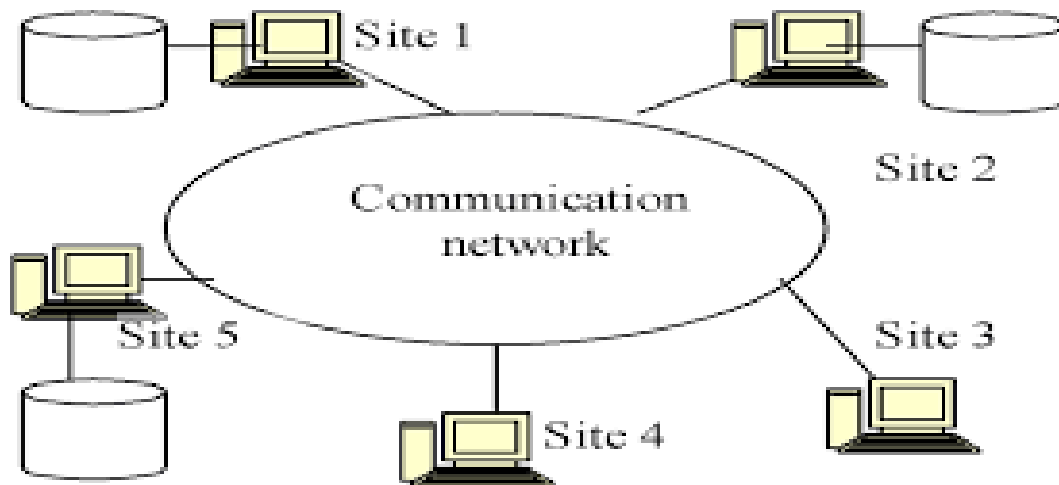


Distribute database

- A **distributed database** is a collection of multiple interconnected databases, which are spread physically across various locations that communicate via a computer network.



Feature

- Databases in the collection are logically interrelated with each other. Often they represent a single logical database.
- Data is physically stored across multiple sites. Data in each site can be managed by a DBMS independent of the other sites.

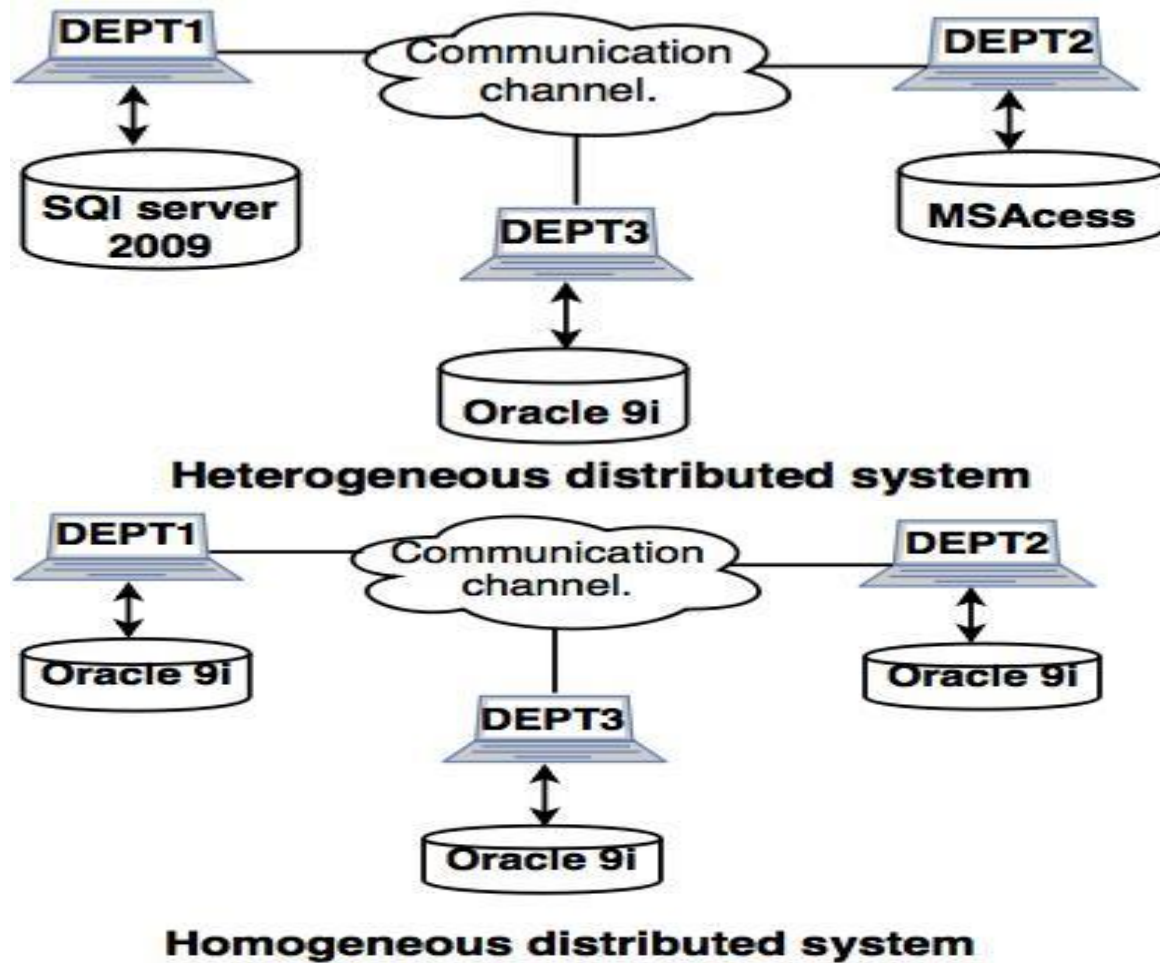
Types

1. Homogeneous Database:

- Homogeneous distributed database system is a network of two or more databases (With same type of DBMS software) which can be stored on one or more machines.

2. Heterogeneous Database:

- Heterogeneous distributed database system is a network of two or more databases with different types of DBMS software, which can be stored on one or more machines.



Advantages

- Easier Expansion
- Improved Performance
- Management of data with different level of transparency
- Increased Reliability and availability
- Sharing Availability and reliability
- Fast data data processing
- Reduce operating cost
- Sharing data

Disadvantages

- Deadlock handling
- Security Complexity management and control
- Lack of standard

Distributed Data Storage

There are 2 ways in which data can be stored on different sites. These are:

1. Replication

In this approach, the entire relation is stored redundantly at 2 or more sites. If the entire database is available at all sites, it is a fully redundant database. Hence, in replication, systems maintain copies of data.

This is advantageous as it increases the availability of data at different sites. Also, now query requests can be processed in parallel.

However, it has certain disadvantages as well. Data needs to be constantly updated. Any change made at one site needs to be recorded at every site that relation is stored or else it may lead to inconsistency. This is a lot of overhead. Also, concurrency control becomes way more complex as concurrent access now needs to be checked over a number of sites.

2. Fragmentation

In this approach, the relations are fragmented (i.e., they're divided into smaller parts) and each of the fragments is stored in different sites where they're required. It must be made sure that the fragments are such that they can be used to reconstruct the original relation (i.e, there isn't any loss of data).

Fragmentation is advantageous as it doesn't create copies of data, consistency is not a problem.

Fragmentation of relations can be done in two ways:

- ✓ Horizontal fragmentation – Splitting by rows – The relation is fragmented into groups of tuples so that each tuple is assigned to at least one fragment.

- ✓ Vertical fragmentation – Splitting by columns – The schema of the relation is divided into smaller schemas. Each fragment must contain a common candidate key so as to ensure lossless join

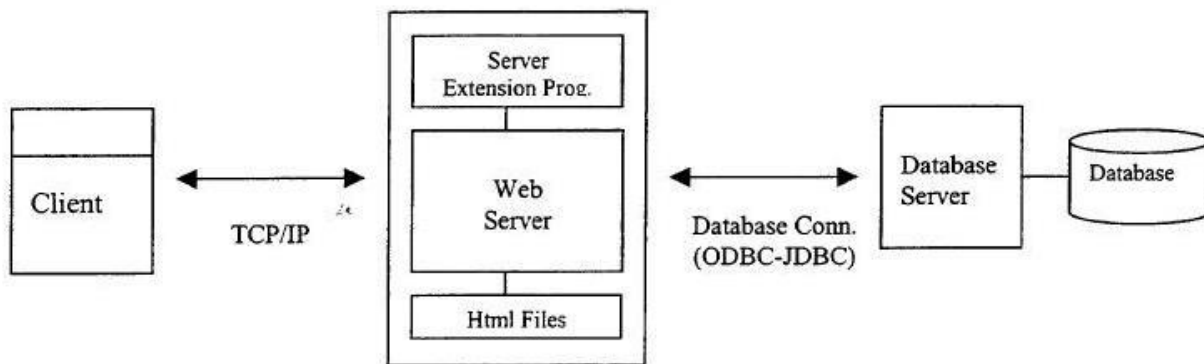
Web Database

A Web database is a database application designed to be managed and accessed through the Internet

Website operators can manage this collection of data and present analytical results based on the data in the Web database application.

A web-based Database management system is used to handle those databases that are having data regarding E-commerce, E-business, blogs, e-mail, and other online applications

Three Tier Web Data Base



Advantages

- ❖ Web-DBMS is Platform independence
- ❖ Provides Graphical User Interface (GUI)
- ❖ Provides Cross-platform support
- ❖ Facilitates transparent network access

Object oriented and Object relational database

- What Is Object Relational Database?
 - An object-relational database (ORD) is a database management system (DBMS) that's composed of both a relational database (RDBMS) and an object-oriented database (OODBMS). An object-relational database acts as an interface between relational and object-oriented databases because it contains aspects and characteristics from both models.
- Ex: PostgreSQL

Object-oriented database serves two main purposes:

- ✓ It connects the divide between relational databases and the object-oriented modeling techniques that are usually used in programming languages like C#, Java and C++.
- ✓ It bridges the gap between conceptual data modeling techniques for relational and object-oriented databases like entity-relationship diagram (ERD) and object-relational mapping (ORM).

What Is Object Oriented Database?

- An object-oriented database management system (OODBMS), sometimes shortened to ODBMS for object database management system, is a database management system (DBMS) that supports the modelling and creation of data as objects.
- Usually, when OODBMS is integrated with an object programming language, there is a much greater consistency between the database and the programming language because both use the same model of data representation.
- When compared to a relational database management system, an object-oriented database stores complex data and relationships between data directly, without mapping to relational rows and columns whereas a relational database stores information in tables with rows and columns.
- OODBMS also includes support for classes of objects and the inheritance of class properties, and incorporates methods, subclasses and their objects.
- E.g. InterSystems Caché, Versant Object Database, Db4o, ObjectStore, Matisse

Object-Oriented Model

Object 1: Maintenance Report

Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

Object 1 Instance

01-12-01
24
I-95
2.5
6.0
6.0

Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

Features Of Object Oriented Database (OODBMS)

- ✓ In object oriented database, relationships are represented by references via the object identifier (OID).
- ✓ Handles larger and complex data than RDBMS.
- ✓ In object oriented systems, the data management language is typically incorporated into a programming language such as C#, C++ etc.
- ✓ Object oriented database can handle different types of data.

Data Mining

- ✓ Refers to extracting or mining knowledge from large amount of data stored in databases, data warehouse, or other repository. i.e. extraction of small valuable information from huge data.
- ✓ It Is the process of discovering interesting patterns & knowledge from large amount of data.
- ✓ Data archeology, data dredging, data/pattern analysis are other terms for data mining. Another popular term Knowledge Discovery From Data(KDD).
- ✓ Data mining is looking for hidden, valid, and potentially useful patterns in huge data sets. Data Mining is all about discovering unsuspected/ previously unknown relationships amongst the data.

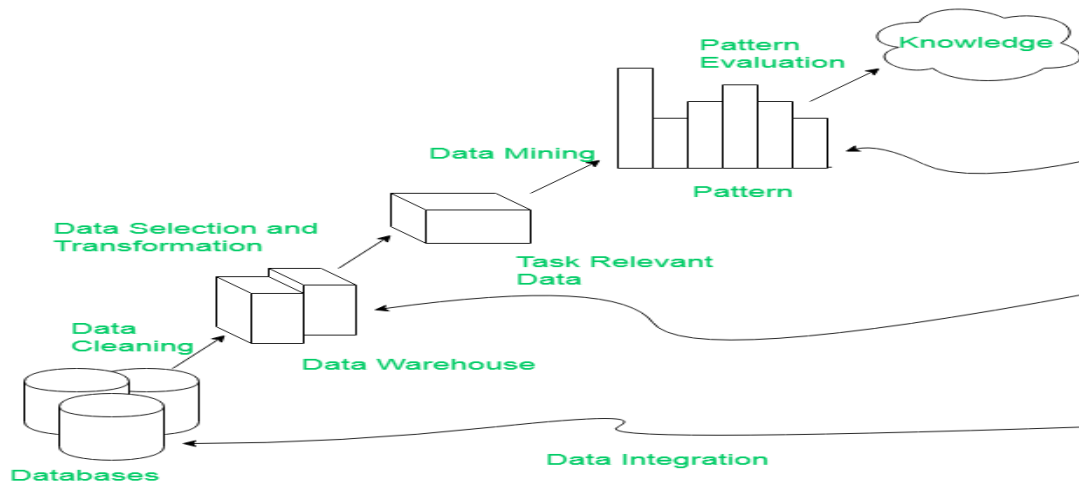
- ✓ It is a multi-disciplinary skill that uses machine learning, statistics, AI and database technology.
- ✓ The insights derived via Data Mining can be used for marketing, fraud detection, and scientific discovery, etc.
- ✓ Data mining is also called as Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, etc.
- ✓

Why Data Mining is important?

Huge data is generated and there is need to turn into useful information and knowledge. This information & knowledge is used for various applications like Market analysis(consumer buying pattern), Fraud detection(fraud account detection, fraud credit card holders), Science exploration(hidden facts in data), telecommunication

Steps in data mining

- ✓ Data Cleaning: Remove noise & inconsistent data.
- ✓ Data Integration: Multiple data sources are
- ✓ Data Selection: Only relevant data are retrieved from database
- ✓ Data Transformation: Data is consolidated into a form which is appropriate for mining
- ✓ Data Mining: Intelligent methods are applied to extract data pattern
- ✓ Pattern Evaluation: To identify the truly interesting patterns representing knowledge based on some interesting measures
- ✓ Knowledge Presentation: Visualizing(graphic) & knowledge representation technique are used to present the mined knowledge to the user



What kinds of patterns can be mined?

- ❖ Mining Frequent Pattern, Association & correlation: Frequent Pattern refers to pattern that occur frequently in data. Mining frequent pattern leads to discovery of interesting association & correlation with data
- ❖ Classification & Prediction: is a process of building a model that describes the class & then predicting the objects into different classes using the model
- ❖ Cluster Analysis: Clustering groups data without any model.
- ❖ Outlier Analysis: Finding out data which differ drastically from others.
- ❖ Evolution Analysis: Describe and models trends for objects whose behavior changes over time

Benefits of Data Mining Or Advantages:

- ❖ Data mining technique helps companies to get knowledge-based information.
- ❖ Data mining helps organizations to make the profitable adjustments in operation and production.
- ❖ The data mining is a cost-effective and efficient solution compared to other statistical data applications.
- ❖ Data mining helps with the decision-making process.

- ❖ Facilitates automated prediction of trends and behaviors as well as automated discovery of hidden patterns.
- ❖ It can be implemented in new systems as well as existing platforms
- ❖ It is the speedy process which makes it easy for the users to analyze huge amount of data in less time.

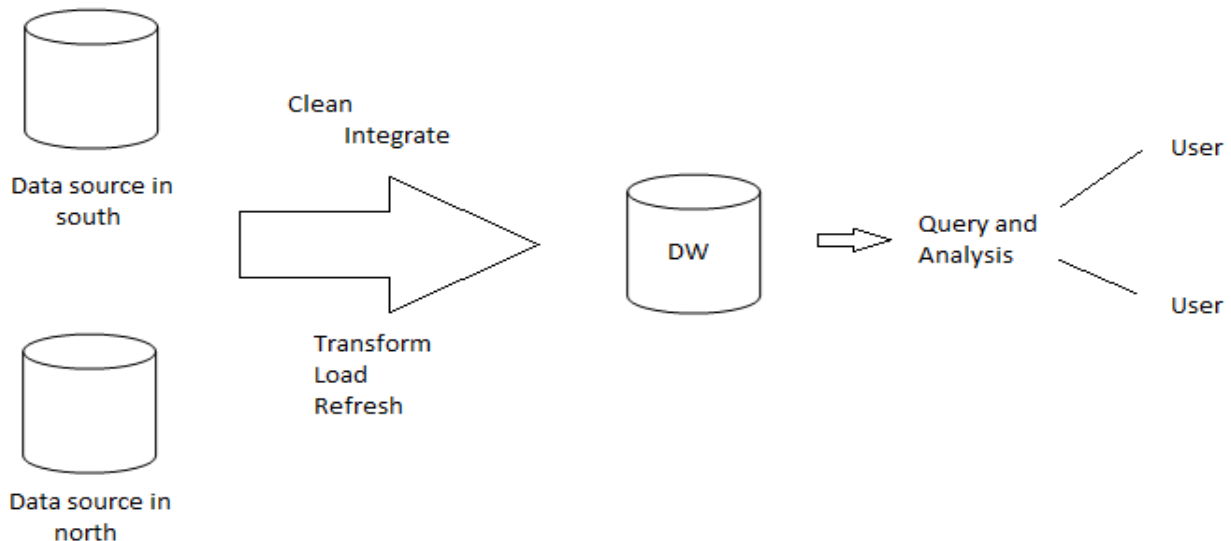
Disadvantages of Data Mining

- There are chances of companies may sell useful information of their customers to other companies for money. For example, American Express has sold credit card purchases of their customers to the other companies.
- Many data mining analytics software is difficult to operate and requires advance training to work on.
- Different data mining tools work in different manners due to different algorithms employed in their design. Therefore, the selection of correct data mining tool is a very difficult task.
- The data mining techniques are not accurate, and so it can cause serious consequences in certain conditions.

Data Warehouse

- A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process.”—W. H. Inmon
- Centralized data location for multiple sources of data.
- A Data Warehousing (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.
- It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

The process of constructing and using data warehouses as shown the following figure



Airline:

In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

Banking:

It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.

Healthcare:

Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.

Public sector:

In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.

Investment and Insurance sector:

In this sector, the warehouses are primarily used to analyze data patterns, customer trends, and to track market movements.

Retain chain:

In retail chains, Data warehouse is widely used for distribution and marketing. It also helps to track items, customer buying pattern, promotions and also used for determining pricing policy.

Telecommunication:

A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.

Hospitality Industry:

This Industry utilizes warehouse services to design as well as estimate their advertising and promotion campaigns where they want to target clients based on their feedback and travel patterns.

Data Mining Vs Data Warehouse: Key Differences

Data Mining	Data Warehouse
Data mining is the process of analyzing unknown patterns of data.	A data warehouse is database system which is designed for analytical instead of transactional work.
Data mining is a method of comparing large amounts of data to finding right patterns.	Data warehousing is a method of centralizing data from different sources into one common repository.
Data mining is usually done by business users with the assistance of engineers.	Data warehousing is a process which needs to occur before any data mining can take place.
Data mining is the considered as a process of extracting data from large data sets.	On the other hand, Data warehousing is the process of pooling all relevant data together.

One of the most important benefits of data mining techniques is the detection and identification of errors in the system.

One of the pros of Data Warehouse is its ability to update consistently. That's why it is ideal for the business owner who wants the best and latest features.

Data mining helps to create suggestive patterns of important factors. Like the buying habits of customers, products, sales. So that, companies can make the necessary adjustments in operation and production.

Data Warehouse adds an extra value to operational business systems like CRM systems when the warehouse is integrated.

The Data mining techniques are never 100% accurate and may cause serious consequences in certain conditions.

In the data warehouse, there is great chance that the data which was required for analysis by the organization may not be integrated into the warehouse. It can easily lead to loss of information.

The information gathered based on Data Mining by organizations can be misused against a group of people.

Data warehouses are created for a huge IT project. Therefore, it involves high maintenance system which can impact the revenue of medium to small-scale organizations.

After successful initial queries, users may ask more complicated queries which would increase the workload.

Data Warehouse is complicated to implement and maintain.

Organisations can benefit from this

Data warehouse stores a large amount of

analytical tool by equipping pertinent and usable knowledge-based information.

historical data which helps users to analyze different time periods and trends for making future predictions.

Organisations need to spend lots of their resources for training and Implementation purpose. Moreover, data mining tools work in different manners due to different algorithms employed in their design.

In Data warehouse, data is pooled from multiple sources. The data needs to be cleaned and transformed. This could be a challenge.

The data mining methods are cost-effective and efficient compares to other statistical data applications.

Data warehouse's responsibility is to simplify every type of business data. Most of the work that will be done on user's part is inputting the raw data.

Another critical benefit of data mining techniques is the identification of errors which can lead to losses. Generated data could be used to detect a drop-in sale.

Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.

Data mining helps to generate actionable strategies built on data insights.

Once you input any information into Data warehouse system, you will unlikely to lose track of this data again. You need to conduct a quick search, helps you to find the right statistic information.