B.C.A. Semester – VI
US06CBCA25: Current Trends in IT (Syllabus Effective from June 2020)

Unit : 1 Data Analytics

Data Warehouse
- Definition, Key Features
- Data Warehousing: A Multitiered Architecture
- Data Warehouse Models
- Data Warehouse functions
- Data warehouse applications

Data Mining
- Introduction to data mining
- Process of knowledge discovery in databases(KDD)
- Data Mining Tasks
- Technologies used for data mining
- Advantages , Disadvantages and Applications

Business Intelligence
- Introduction, Meaning, Importance, Benefits

Data Analytics
- What is Data Analytics
- Importance of Data Analytics
- Types of Data Analytics

Data Warehouse

Introduction

Data warehousing provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions.
Data warehousing is the latest must-have marketing weapon—a way to retain customers by learning more about their needs.

Data warehouses have been defined in many ways. Loosely speaking, a data warehouse refers to a data repository that is maintained separately from an organization's operational databases. Data warehouse systems allow for integration of a variety of application systems. They support information processing by providing a solid platform of combined historic data for analysis.

According to William H. Inmon, a leading architect in the construction of data warehouse systems, "A data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management's decision making process".

This short but comprehensive definition presents the major features of a data warehouse. The four keywords—subject-oriented, integrated, time-variant, and nonvolatile - distinguish data

warehouses from other data repository systems, such as relational database systems, transaction processing systems, and file systems.

Key Features of Data warehouse

Subject-oriented: A data warehouse is organized around major subjects such as customer, supplier, product, and sales. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and short view of particular subject issues by excluding data that are not useful in the decision support process.

Integrated: A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and online transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on.

Time-variant: Data are stored to provide information from an historic perspective (e.g., the past 5–10 years). Every key structure in the data warehouse contains, either implicitly or explicitly, a time element.

Nonvolatile: A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a data warehouse does not require transaction processing, recovery and concurrency control mechanisms. It usually requires only two operations in data accessing: initial loading of data and access of data.

Data Warehousing: A Multitiered Architecture

Data Warehouses usually have a three-level (tier) architecture that includes:

1. Bottom Tier (Data Warehouse Server)
2. Middle Tier (OLAP Server)
3. Top Tier (Front end Tools).

The bottom tier is a **warehouse database server** that is almost always a relational database system. Back-end tools and utilities are used to feed data into the bottom tier from operational databases or other external sources. These tools and utilities perform data extraction, cleaning, and transformation, as well as load and refresh functions to update the data warehouse. The data are extracted using application program interfaces known as gateways. A gateway is supported by the underlying DBMS and allows client programs to generate SQL code to be executed at a server. Examples of gateways include ODBC (Open Database Connection) and OLEDB (Object Linking and Embedding Database) by Microsoft and JDBC (Java Database Connection).

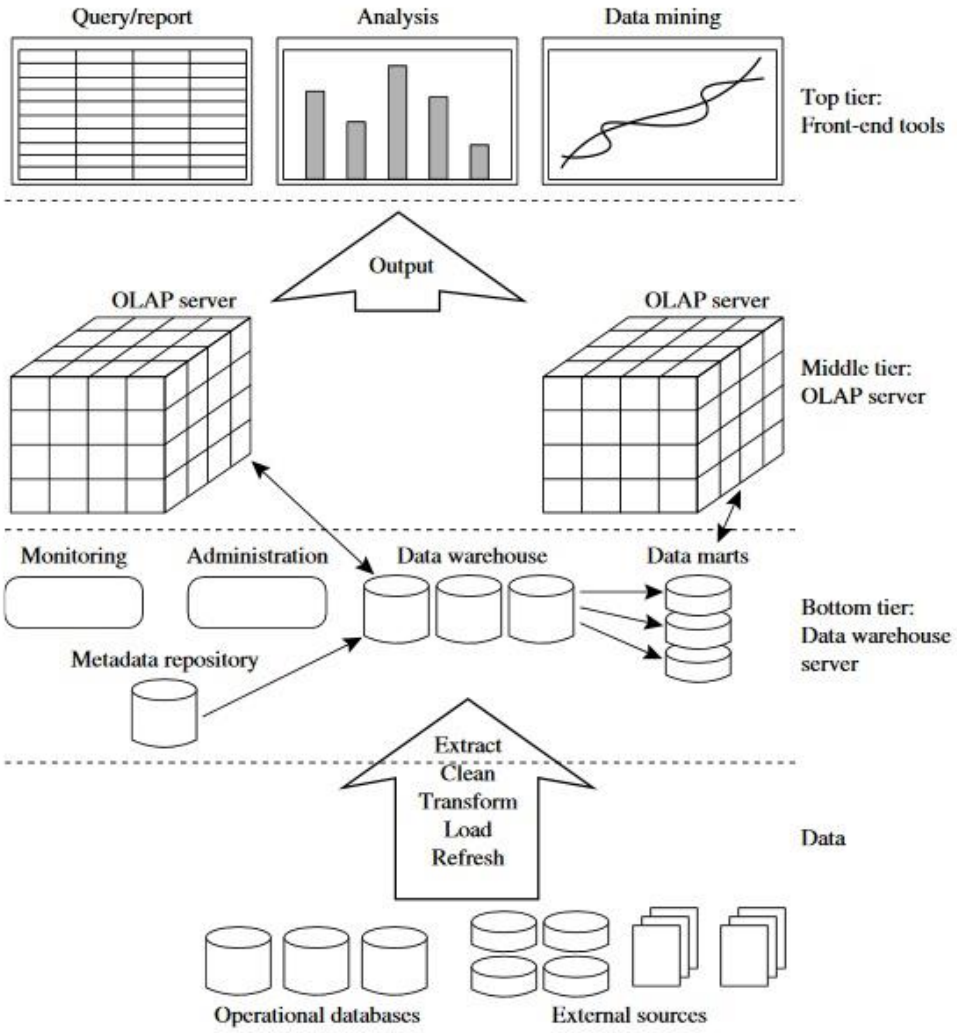


Figure 1.1 A three-tier data warehousing architecture.

The middle tier is an **OLAP server** that is typically implemented using either (1) a relational OLAP (ROLAP) model (i.e., an extended relational DBMS that maps operations on multidimensional data to standard relational operations); or (2) a multidimensional OLAP (MOLAP) model (i.e., a special-purpose server that directly implements multidimensional data and operations).

The top tier is a **front-end client layer**, which contains query and reporting tools, analysis tools, and/or data mining tools (e.g., trend analysis, prediction, and so on).

Data Warehouse Models

From the architecture point of view, there are **three data warehouse models**: the enterprise warehouse, the data mart, and the virtual warehouse.

Enterprise warehouse:

An enterprise warehouse collects all of the information about subjects across the entire organization. It provides corporate-wide data integration, usually from one or more operational systems or external information providers, and is cross-functional in scope. It typically contains detailed data as well as summarized data, and can range in size from a few gigabytes to hundreds of gigabytes, terabytes, or beyond. An enterprise data warehouse may be implemented on traditional mainframes, computer super servers, or parallel architecture platforms. It requires extensive business modeling and may take years to design and build.

Data mart:

A data mart contains a subset of corporate-wide data that is of value to a specific group of users. The scope is confined to specific selected subjects. The data contained in data marts tend to be summarized. Data marts are usually implemented on low-cost departmental servers that are Unix/Linux or Windows based. The implementation cycle of a data mart is more likely to be measured in weeks rather than months or years.

Virtual warehouse:

A virtual warehouse is a set of views over operational databases. For efficient query processing, only some of the possible summary views may be materialized. A virtual warehouse is easy to build but requires excess capacity on operational database servers.

Data Warehouse functions

Data warehouse systems use back-end tools and utilities to populate and refresh their data (Figure 1.1). These tools and utilities include the following functions:

Data extraction: This typically gathers data from multiple, heterogeneous, and external sources.

Data cleaning: This detects errors in the data and rectifies them when possible.

Data transformation: This converts data from legacy or host format to warehouse format.

Load: This sorts, summarizes, consolidates, computes views, checks integrity, and builds indices and partitions.

Refresh: This propagates the updates from the data sources to the warehouse.

Data warehouse applications

Business users need to have the means to know what exists in the data warehouse, how to access the contents of the data warehouse, how to examine the contents using analysis tools, and how to present the results of such analysis.

There are three kinds of data warehouse applications:

1. Information processing
2. Analytical processing
3. Data mining

Information processing supports querying, basic statistical analysis, and reporting using crosstabs, tables, charts, or graphs. A current trend in data warehouse information processing is to construct low-cost web-based accessing tools that are then integrated with web browsers.

Analytical processing supports basic OLAP operations, including slice-and-dice, drill-down, roll-up, and pivoting. It generally operates on historic data in both summarized and detailed forms. The major strength of online analytical processing over information processing is the multidimensional data analysis of data warehouse data.

Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools.

Data Mining

Introduction to data mining

Rapid advances in data collection and storage technology have enabled organizations to accumulate vast amounts of data. However, extracting useful information has proven extremely challenging. Often, traditional data analysis tools and techniques cannot be used because of the massive size of a dataset. Data mining is a technology that blends traditional data analysis methods with sophisticated algorithms for processing large volumes of data. It has also opened up exciting opportunities for exploring and analyzing new types of data and for analyzing old types of data in new ways.

Data mining techniques can be used to support a wide range of business intelligence applications such as customer profiling, targeted marketing, work-flow management, store layout, and fraud detection. Researchers in medicine, science, and engineering are rapidly accumulating data that is key to important new discoveries.

What Is Data Mining?

Data mining is the process of automatically discovering useful information in large data repositories. Data mining techniques are deployed to scour large databases in order to find novel and useful patterns that might otherwise remain unknown. They also provide capabilities to predict the outcome of a future observation. Data mining utilizes complex mathematical algorithms for data segments and evaluates the probability of future events. Data Mining is also called Knowledge Discovery of Data (KDD).

Process of knowledge discovery in databases (KDD)

Data mining is an integral part of knowledge discovery in databases (KDD), which is the overall process of converting raw data into useful in-formation, as shown in Figure 1.2. This process consists of a series of transformation steps, from data preprocessing to post processing of data mining results.

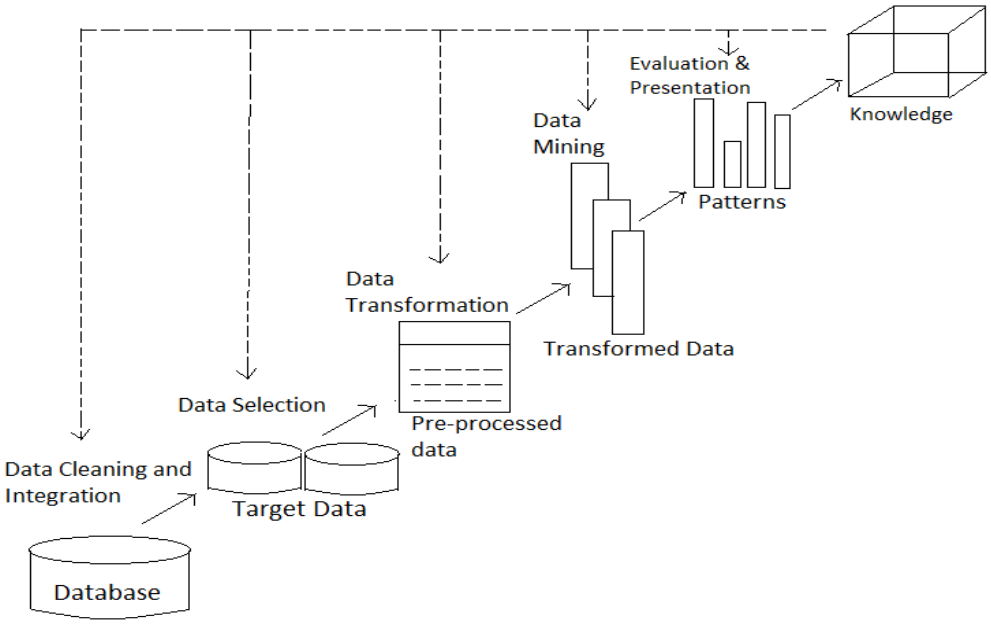


Figure 1.2 Process of knowledge discovery in databases (KDD)

The iterative process consists of the following steps:

- **Data cleaning**: also known as data cleansing, it is a phase in which noise data and irrelevant data are removed from the collection.
- **Data integration**: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.
- **Data selection**: at this step, the data relevant to the analysis is decided on and retrieved from the data collection.
- **Data transformation**: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- **Data mining**: it is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- **Pattern evaluation**: in this step, strictly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge representation**: is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

Data Mining Tasks

There are a number of data mining tasks such as classification, prediction, time-series analysis, association, clustering, summarization etc. All these tasks are either predictive data mining tasks or descriptive data mining tasks.

Data mining tasks are generally divided into two major categories:

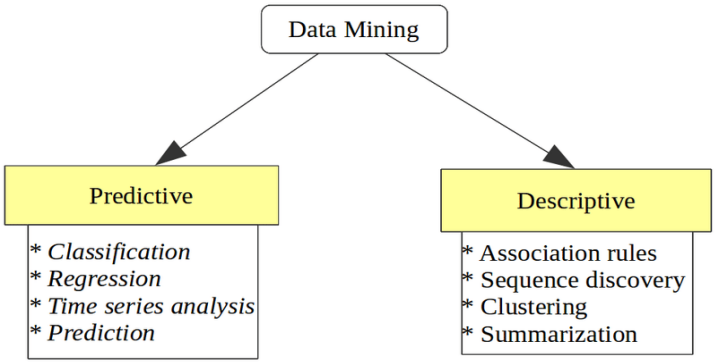


Figure 1.3 Data Mining Tasks

Predictive tasks: The objective of these tasks is to predict the value of a particular attribute based on the values of other attributes. The attribute to be predicted is commonly known as the target or dependent variable, while the attributes used for making the prediction are known as the explanatory or independent variables.

For e.g. a medical practitioner trying to diagnose a disease based on the medical test results of a patient can be considered as a predictive data mining task.

Descriptive tasks: Descriptive data mining tasks usually finds data describing patterns and comes up with new, significant information from the available data set. Descriptive data mining tasks are often exploratory in nature and frequently require post processing techniques to validate and explain the results.

For e.g. a retailer trying to identify products that are purchased together can be considered as a descriptive data mining task.

Technologies used for data mining

As a highly application-driven domain, data mining has incorporated many techniques from other domains such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application domains (Figure 1.4). The interdisciplinary nature of data mining research and development contributes significantly to the success of data mining and its extensive applications.

Statistics: A statistical model is a set of mathematical functions that describe the behavior of the objects in a target class.
Machine learning: Investigates how computers can learn based on data.
Database systems: Research focuses on the creation, maintenance, and use of databases for organizations and end users.
Data warehouse: Integrates data originating from multiple sources and various timeframes.
Information retrieval: is the science of searching for documents or information in documents.

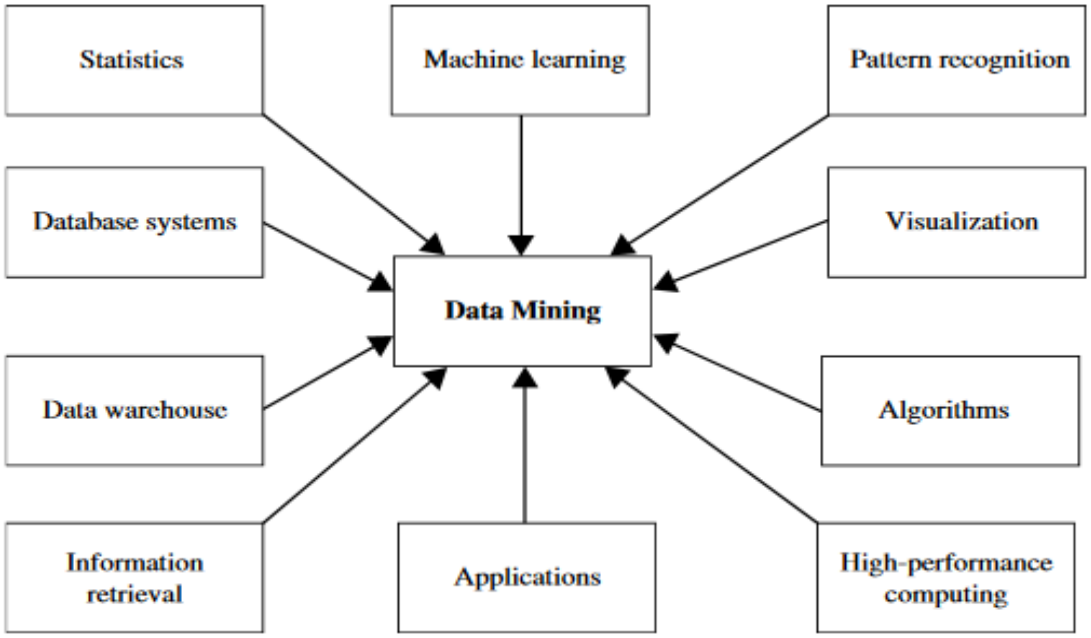


Figure 1.4 Data mining adopts techniques from many domains

Advantages of Data Mining

- The Data Mining technique enables organizations to obtain knowledge-based data.
- Data mining enables organizations to make profitable modifications in operation and production.
- Compared with other statistical data applications.
- Data Mining helps the decision-making process of an organization.
- It facilitates the automated discovery of hidden patterns as well as the prediction of trends and behaviors.
- It can be induced in the new system as well as the existing platforms.
- It is a quick process that makes it easy for new users to analyze enormous amounts of data in a short time.

Disadvantages of Data Mining

- There is a probability that the organizations may sell useful data of customers to other organizations for money.
- Many data mining analytics software is difficult to operate and needs advance training to work on.
- Different data mining instruments operate in distinct ways due to the different algorithms used in their design. Therefore, the selection of the right data mining tools is a very challenging task.

Applications

These are the following areas where data mining is widely used:

- Healthcare
- Market Basket Analysis
- Education
- Manufacturing Engineering
- CRM (Customer Relationship Management)
- Fraud detection
- Lie Detection
- Financial Banking

Business Intelligence (BI)

Introduction

The BI term refers to **Business Intelligence**. It is a data-driven decision support system (DSS), which helps you to analyze the data and provide actionable information. It helps the business manager, corporate executives, and other users in making their decisions easily. Data mining is the core of business intelligence.

It is critical for businesses to acquire a better understanding of the commercial context of their organization, such as their customers, the market, supply and resources, and competitors. Business intelligence (BI) technologies provide historical, current, and predictive views of business operations. Examples include reporting, online analytical processing, business performance management, competitive intelligence, benchmarking, and predictive analytics.

Importance of BI

Business intelligence is used to improve all parts of a company by improving access to the firm's data and then using that data to increase profitability. Companies that practices BI can translate their collected data into insights their business processors.

Then the insights can be used to create strategic business decisions that improve productivity and accelerate the growth.

Business intelligence are very important because,

- For powerful new revenues.
- To increases operational efficiency.
- To optimizes internal business processes.
- To improves decision making.
- To gain a competitive advantage over business rivals.
- To use for spotting business problems that need to be addressed.
- To use for assisting companies in the identification of market trends.

Benefits of Business intelligence

- Improves Decision making
- Fast answers to any business query
- Enables real time analysis with quick navigation
- Identify cross selling and Up- selling
- Reduce the risk of bottlenecks
- Helps you to know your business better

Data Analytics

Data has been the buzzword for ages now. Either the data being generated from large-scale enterprises or the data generated from an individual, each and every aspect of data needs to be analyzed to benefit from it. But how do we do it? That's where the term 'Data Analytics' comes in.

Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain. Data is extracted from various sources and is cleaned and categorized to analyze various behavioral patterns. The techniques and the tools used vary according to the organization or individual.

Importance of Data Analytics

Data Analytics has a key role in improving your business as it is used for,

- **Gather Hidden Insights** – Hidden insights from data are gathered and then analyzed with respect to business requirements.
- **Generate Reports** – Reports are generated from the data and are passed on to the respective teams and individuals to deal with further actions for a high rise in business.
- **Perform Market Analysis** – Market Analysis can be performed to understand the strengths and weaknesses of competitors.
- **Improve Business Requirement** – Analysis of Data allows improving Business to customer requirements and experience.

Types of Data Analytics

There are four primary types of data analytics: descriptive, diagnostic, predictive and prescriptive analytics. Each type has a different goal and a different place in the data analysis process.

- **Descriptive analytics** helps answer questions about what happened. These techniques summarize large datasets to describe outcomes. Specialized metrics are developed to track performance in specific industries. This process requires the collection of relevant data, processing of the data, data analysis and data visualization. This process provides essential insight into past performance.

- **Diagnostic analytics** helps answer questions about why things happened. These techniques supplement more basic descriptive analytics. The performance indicators are further investigated to discover why they got better or worse.

- **Predictive analytics** helps answer questions about what will happen in the future. These techniques use historical data to identify trends and determine if they are likely to recur. Predictive analytical tools provide valuable insight into what may happen in the future and its techniques include a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression.

- **Prescriptive analytics** helps answer questions about what should be done. By using insights from predictive analytics, data-driven decisions can be made. This allows businesses to make informed decisions in the face of uncertainty. By analyzing past decisions and events, the likelihood of different outcomes can be estimated.

Tools used in Data Analytics

With the increasing demand for Data Analytics in the market, many tools have emerged with various functionalities for this purpose. Either open-source or user-friendly, the top tools in the data analytics market are as follows.

- R programming
- Python
- Tableau Public
- QlikView
- SAS
- RapidMiner
- KNIME
- OpenRefine
- Apache Spark

TYBCA SEM-VI
Current Trends in IT
Unit – II Machine Learning

*INTRODUCTION TO AI

What is Artificial Intelligence?

In today's world, technology is growing very fast, and we are getting in touch with different new technologies day by day.

Here, one of the booming technologies of computer science is Artificial Intelligence which is ready to create a new revolution in the world by making intelligent machines.

The Artificial Intelligence is now all around us. It is currently working with a variety of subfields, ranging from general to specific, such as self-driving cars, playing chess, proving theorems, playing music, Painting, etc.

AI is one of the fascinating and universal fields of Computer science which has a great scope in future. AI holds a tendency to cause a machine to work as a human.

Artificial Intelligence is composed of two words **Artificial** and **Intelligence**, where Artificial defines "man-made," and intelligence defines "thinking power", hence AI means "a man-made thinking power."

So, we can define AI as:

"It is a branch of computer science by which we can create intelligent machines which can behave like a human, think like humans, and able to make decisions."

Artificial Intelligence exists when a machine can have human based skills such as learning, reasoning, and solving problems.

With Artificial Intelligence you do not need to preprogram a machine to do some work, despite that you can create a machine with programmed algorithms which can work with own intelligence, and that is the awesomeness of AI.

It is believed that AI is not a new technology, and some people says that as per Greek myth, there were Mechanical men in early days which can work and behave like humans.

Goals of Artificial Intelligence

Following are the main goals of Artificial Intelligence:

1. Replicate human intelligence
2. Solve Knowledge-intensive tasks
3. An intelligent connection of perception and action
4. Building a machine which can perform tasks that requires human intelligence such as:
   - Proving a theorem
   - Playing chess
   - Plan some surgical operation
   - Driving a car in traffic
5. Creating some system which can exhibit intelligent behavior, learn new things by itself, demonstrate, explain, and can advise to its user.

What Comprises to Artificial Intelligence?

Artificial Intelligence is not just a part of computer science even it's so vast and requires lots of other factors which can contribute to it.

To create the AI first we should know that how intelligence is composed, so the Intelligence is an intangible part of our brain which is a combination of

- **Reasoning,**
- **learning,**
- **problem-solving perception,**
- **language understanding, etc**.

To achieve the above factors for a machine or software Artificial Intelligence requires the following discipline:

- Mathematics
- Biology
- Psychology
- Sociology
- Computer Science
- Neurons Study
- Statistics

Advantages of Artificial Intelligence

- **High Accuracy with less errors:** AI machines or systems are prone to less errors and high accuracy as it takes decisions as per pre-experience or information.

- **High-Speed:** AI systems can be of very high-speed and fast-decision making, because of that AI systems can beat a chess champion in the Chess game.

- **High reliability:** AI machines are highly reliable and can perform the same action multiple times with high accuracy.

- **Useful for risky areas:** AI machines can be helpful in situations such as defusing a bomb, exploring the ocean floor, where to employ a human can be risky.

- **Digital Assistant:** AI can be very useful to provide digital assistant to the users such as AI technology is currently used by various E-commerce websites to show the products as per customer requirement.

- **Useful as a public utility:** AI can be very useful for public utilities such as a self-driving car which can make our journey safer and hassle-free, facial recognition for security purpose, Natural language processing to communicate with the human in human-language, etc.

Disadvantages of Artificial Intelligence

- **High Cost:** The hardware and software requirement of AI is very costly as it requires lots of maintenance to meet current world requirements.

- **Can't think out of the box:** Even we are making smarter machines with AI, but still they cannot work out of the box, as the robot will only do that work for which they are trained, or programmed.

- **No feelings and emotions:** AI machines can be an outstanding performer, but still it does not have the feeling so it cannot make any kind of emotional attachment with human, and may sometime be harmful for users if the proper care is not taken.

- **Increase dependency on machines:** With the increment of technology, people are getting more dependent on devices and hence they are losing their mental capabilities.

- **No Original Creativity:** As humans are so creative and can imagine some new ideas but still AI machines cannot beat this power of human intelligence and cannot be creative and imaginative.

*AI and RELATED FIELDS

Following are the most common subsets of AI:

1. Machine Learning
2. Deep Learning
3. Natural Language processing
4. Expert System
5. Robotics
6. Machine Vision
7. Speech Recognition

1. Machine Learning

Machine learning is a part of AI which provides intelligence to machines with the ability to automatically learn with experiences without being explicitly programmed.

- It is primarily concerned with the design and development of algorithms that allow the system to learn from historical data.
- Machine Learning is based on the idea that machines can learn from past data, identify patterns, and make decisions using algorithms.
- Machine learning algorithms are designed in such a way that they can learn and improve their performance automatically.
- Machine learning helps in discovering patterns in data.

Types of Machine Learning

Machine learning can be subdivided into the main three types:

Supervised learning:

Supervised learning is a type of machine learning in which machine learn from known datasets (set of training examples), and then predict the output. A supervised learning agent needs to find out the function that matches a given sample set.

Reinforcement learning:

Reinforcement learning is a type of learning in which an AI agent is trained by giving some commands, and on each action, an agent gets a reward as a feedback.

Using these feedbacks, agent improves its performance. Reward feedback can be positive or negative which means on each good action, agent receives a positive reward while for wrong action, it gets a negative reward.

Unsupervised learning:

Unsupervised learning is associated with learning without supervision or training.

In unsupervised learning, the algorithms are trained with data which is neither labeled nor classified. In unsupervised learning, the agent needs to learn from patterns without corresponding output values.

2. Deep Learning

Deep learning is a subset of machine learning which provides the ability to machine to perform human-like tasks without human involvement.

It provides the ability to an AI agent to mimic the human brain. Deep learning can use both supervised and unsupervised learning to train an AI agent.

- Deep learning is implemented through neural networks architecture hence also called a **deep neural network**.
- Deep learning is the primary technology behind self-driving cars, speech recognition, image recognition, automatic machine translation, etc.
- The main challenge for deep learning is that it requires lots of data with lots of computational power.

3. Natural Language processing

Natural language processing is a subfield of computer science and artificial intelligence. NLP enables a computer system to understand and process human language such as English.

NLP plays an important role in AI as without NLP, AI agent cannot work on human instructions, but with the help of NLP, we can instruct an AI system on our language.

Today we are all around AI, and as well as NLP, we can easily ask Siri, Google or Cortana to help us in our language.

Natural language processing application enables a user to communicate with the system in their own words directly.

The Input and output of NLP applications can be in two forms:

- Speech
- Text

4. Expert Systems

- An expert system is an application of artificial intelligence. In artificial intelligence, expert systems are the computer programs that rely on obtaining the knowledge of human experts and programming that knowledge into a system.
- Expert systems emulate the decision-making ability of human experts. These systems are designed to solve the complex problem through bodies of knowledge rather than conventional procedural code.
- One of the examples of an expert system is a Suggestion for the spelling error while typing in the Google search box.

5. Robotics

- Robotics is a branch of artificial intelligence and engineering which is used for designing and manufacturing of robots.
- Robots are the programmed machines which can perform a series of actions automatically or semi-automatically.
- AI can be applied to robots to make intelligent robots which can perform the task with their intelligence. AI algorithms are necessary to allow a robot to perform more complex tasks.
- Nowadays, AI and machine learning are being applied on robots to manufacture intelligent robots which can also interact socially like humans. One of the best examples of AI in robotics is Sophia robot.

6. Machine Vision

- Machine vision is an application of computer vision which enables a machine to recognize the object.
- Machine vision captures and analyses visual information using one or more video cameras, analog-to-digital conversations, and digital signal processing.
- Machine vision systems are programmed to perform narrowly defined tasks such as counting objects, reading the serial number, etc.
- Computer systems do not see in the same way as human eyes can see, but it is also not bounded by human limitations such as to see through the wall.
- With the help of machine learning and machine vision, an AI agent can be able to see through walls.

7. Speech Recognition

Speech recognition is a technology which enables a machine to understand the spoken language and translate into a machine-readable format. It can also be said as automatic Speech recognition and computer speech recognition.

It is a way to talk with a computer, and on the basis of that command, a computer can perform a specific task.

There is some speech recognition software which has a limited vocabulary of words and phrase. This software requires unambiguous spoken language to understand and perform specific task.

Today's there are various software or devices which contains speech recognition technology such as Cortana, Google virtual assistant, Apple Siri, etc.

We need to train our speech recognition system to understand our language. In previous days, these systems were only designed to convert the speech to text, but now there are various devices which can directly convert speech into commands.

Speech recognition systems can be used in the following areas:

- System control or navigation system
- Industrial application
- Voice dialing system

*EXPERT SYSTEMS

Expert systems (ES) are one of the prominent research domains of AI. It is introduced by the researchers at Stanford University, Computer Science Department.

What are Expert Systems?

The expert systems are the computer applications developed to solve complex problems in a particular domain, at the level of extra-ordinary human intelligence and expertise.

Characteristics of Expert Systems

- High performance
- Understandable
- Reliable
- Highly responsive

Capabilities of Expert Systems

The expert systems are capable of −

- Advising
- Instructing and assisting human in decision making
- Demonstrating
- Deriving a solution
- Diagnosing
- Explaining
- Interpreting input
- Predicting results
- Justifying the conclusion
- Suggesting alternative options to a problem

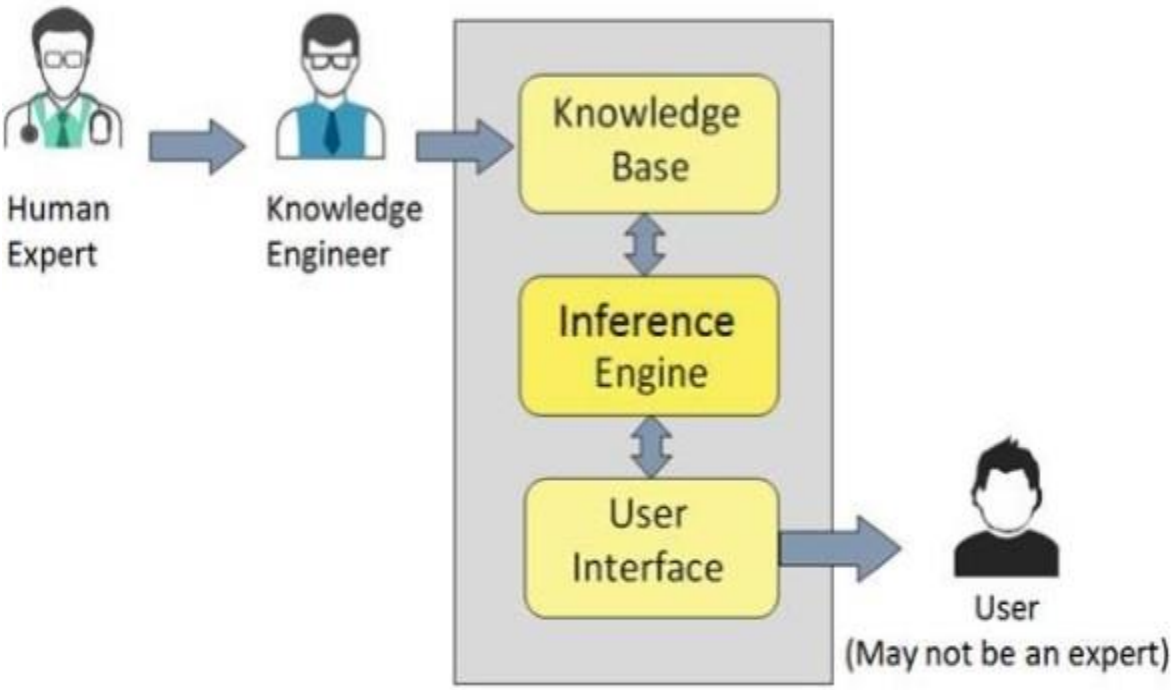
They are incapable of −

- Substituting human decision makers
- Possessing human capabilities
- Producing accurate output for inadequate knowledge base
- Refining their own knowledge

Components of Expert Systems

The components of ES include −

1. Knowledge Base
2. Inference Engine
3. User Interface



1. Knowledge Base

It contains domain-specific and high-quality knowledge.

Knowledge is required to exhibit intelligence. The success of any ES majorly depends upon the collection of highly accurate and precise knowledge.

What is Knowledge?

The data is collection of facts. The information is organized as data and facts about the task domain. **Data, information,** and **past experience** combined together are termed as knowledge.

Components of Knowledge Base

The knowledge base of an ES is a store of both, factual and heuristic knowledge.

- **Factual Knowledge** − It is the information widely accepted by the Knowledge Engineers and scholars in the task domain.
- **Heuristic Knowledge** − It is about practice, accurate judgement, one's ability of evaluation, and guessing.

Knowledge representation

It is the method used to organize and formalize the knowledge in the knowledge base. It is in the form of IF-THEN-ELSE rules.

Knowledge Acquisition

The success of any expert system majorly depends on the quality, completeness, and accuracy of the information stored in the knowledge base.

The knowledge base is formed by readings from various experts, scholars, and the **Knowledge Engineers**. The knowledge engineer is a person with the qualities of empathy, quick learning, and case analyzing skills.

He acquires information from subject expert by recording, interviewing, and observing him at work, etc. He then categorizes and organizes the information in a meaningful way, in the form of IF-THEN-ELSE rules, to be used by interference machine. The knowledge engineer also monitors the development of the ES.

2. Inference Engine

Use of efficient procedures and rules by the Inference Engine is essential in deducting a correct, flawless solution.

In case of knowledge-based ES, the Inference Engine acquires and manipulates the knowledge from the knowledge base to arrive at a particular solution.

In case of rule based ES, it −

- Applies rules repeatedly to the facts, which are obtained from earlier rule application.
- Adds new knowledge into the knowledge base if required.
- Resolves rules conflict when multiple rules are applicable to a particular case.

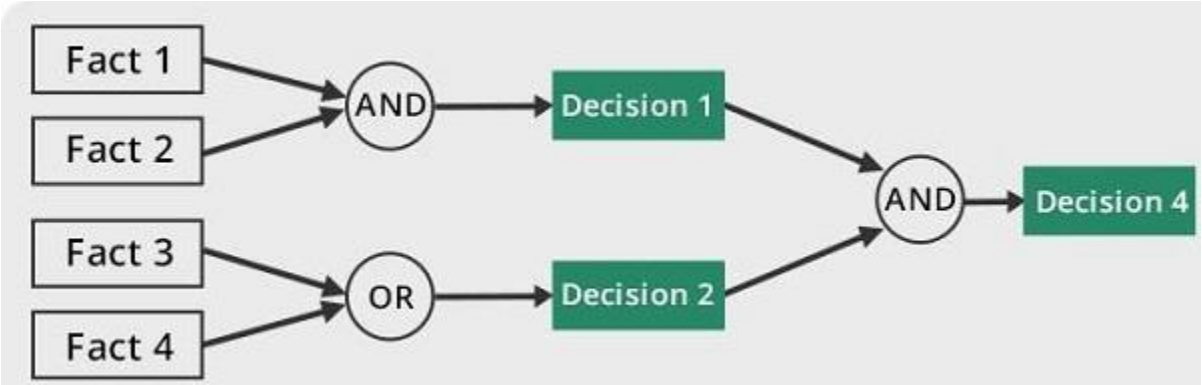
To recommend a solution, the Inference Engine uses the following strategies −

- Forward Chaining
- Backward Chaining

Forward Chaining

It is a strategy of an expert system to answer the question, **"What can happen next?"** Here, the Inference Engine follows the chain of conditions and derivations and finally deduces the outcome. It considers all the facts and rules, and sorts them before concluding to a solution.

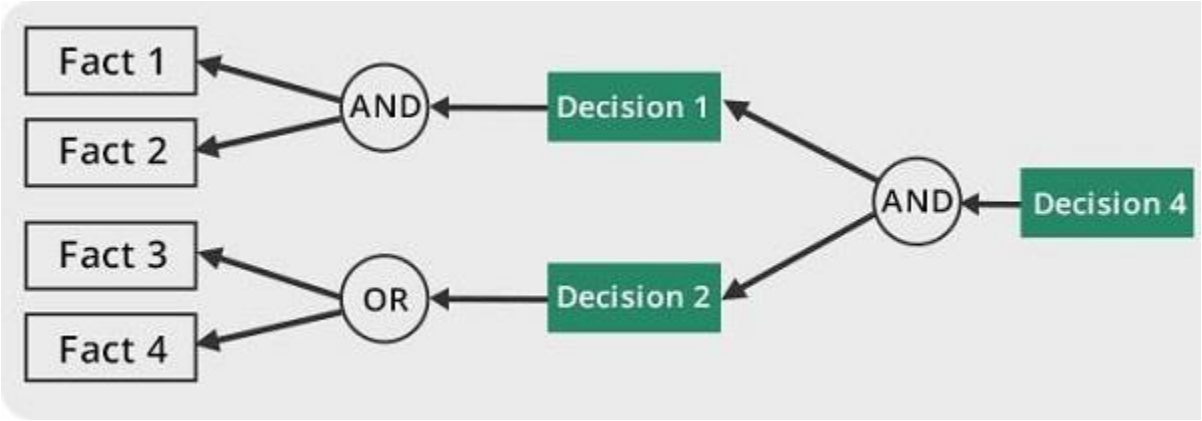
This strategy is followed for working on conclusion, result, or effect. For example, prediction of share market status as an effect of changes in interest rates.



Backward Chaining

With this strategy, an expert system finds out the answer to the question, **"Why this happened?"**

On the basis of what has already happened, the Inference Engine tries to find out which conditions could have happened in the past for this result. This strategy is followed for finding out cause or reason. For example, diagnosis of blood cancer in humans.



3. User Interface

User interface provides interaction between user of the ES and the ES itself. It is generally Natural Language Processing so as to be used by the user who is well-versed in the task domain. The user of the ES need not be necessarily an expert in Artificial Intelligence.

It explains how the ES has arrived at a particular recommendation. The explanation may appear in the following forms –

- Natural language displayed on screen.
- Verbal narrations in natural language.
- Listing of rule numbers displayed on the screen.

The user interface makes it easy to trace the credibility of the deductions.

Requirements of Efficient ES User Interface

- It should help users to accomplish their goals in shortest possible way.
- It should be designed to work for user's existing or desired work practices.
- Its technology should be adaptable to user's requirements; not the other way round.
- It should make efficient use of user input.

Expert Systems Limitations

No technology can offer easy and complete solution. Large systems are costly, require significant development time, and computer resources. ESs have their limitations which include −

- Limitations of the technology
- Difficult knowledge acquisition
- ES are difficult to maintain
- High development costs

Applications of Expert System

The following table shows where ES can be applied.

Application	Description
Design Domain	Camera lens design, automobile design.
Medical Domain	Diagnosis Systems to deduce cause of disease from observed data, conduction medical operations on humans.
Monitoring Systems	Comparing data continuously with observed system or with prescribed behavior such as leakage monitoring in long

	petroleum pipeline.
Process Control Systems	Controlling a physical process based on monitoring.
Knowledge Domain	Finding out faults in vehicles, computers.
Finance/Commerce	Detection of possible fraud, suspicious transactions, stock market trading, Airline scheduling, cargo scheduling.

Development of Expert Systems: General Steps

The process of ES development is iterative. Steps in developing the ES include –

Identify Problem Domain

- The problem must be suitable for an expert system to solve it.
- Find the experts in task domain for the ES project.
- Establish cost-effectiveness of the system.

Design the System

- Identify the ES Technology
- Know and establish the degree of integration with the other systems and databases.
- Realize how the concepts can represent the domain knowledge best.

Develop the Prototype

From Knowledge Base: The knowledge engineer works to −

- Acquire domain knowledge from the expert.
- Represent it in the form of If-THEN-ELSE rules.

Test and Refine the Prototype

- The knowledge engineer uses sample cases to test the prototype for any deficiencies in performance.
- End users test the prototypes of the ES.

Develop and Complete the ES

- Test and ensure the interaction of the ES with all elements of its environment, including end users, databases, and other information systems.
- Document the ES project well.
- Train the user to use ES.

Maintain the System

- Keep the knowledge base up-to-date by regular review and update.
- Cater for new interfaces with other information systems, as those systems evolve.

Benefits of Expert Systems

- Availability − They are easily available due to mass production of software.
- Less Production Cost − Production cost is reasonable. This makes them affordable.
- Speed − They offer great speed. They reduce the amount of work an individual puts in.
- Less Error Rate − Error rate is low as compared to human errors.
- Reducing Risk − They can work in the environment dangerous to humans.
- Steady response − They work steadily without getting motional, tensed or fatigued.

*FUZZY LOGIC

Fuzzy Logic Systems (FLS) produce acceptable but definite output in response to incomplete, ambiguous, distorted, or inaccurate (fuzzy) input.

What is Fuzzy Logic?

Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning. The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.

The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO.

The inventor of fuzzy logic, Lotfi Zadeh, observed that unlike computers, the human decision making includes a range of possibilities between YES and NO, such as –

CERTAINLY YES
POSSIBLY YES
CANNOT SAY
POSSIBLY NO
CERTAINLY NO

The fuzzy logic works on the levels of possibilities of input to achieve the definite output.

Implementation

- It can be implemented in systems with various sizes and capabilities ranging from small micro-controllers to large, networked, workstation-based control systems.
- It can be implemented in hardware, software, or a combination of both.

Why Fuzzy Logic?

Fuzzy logic is useful for commercial and practical purposes.

- It can control machines and consumer products.
- It may not give accurate reasoning, but acceptable reasoning.
- Fuzzy logic helps to deal with the uncertainty in engineering.

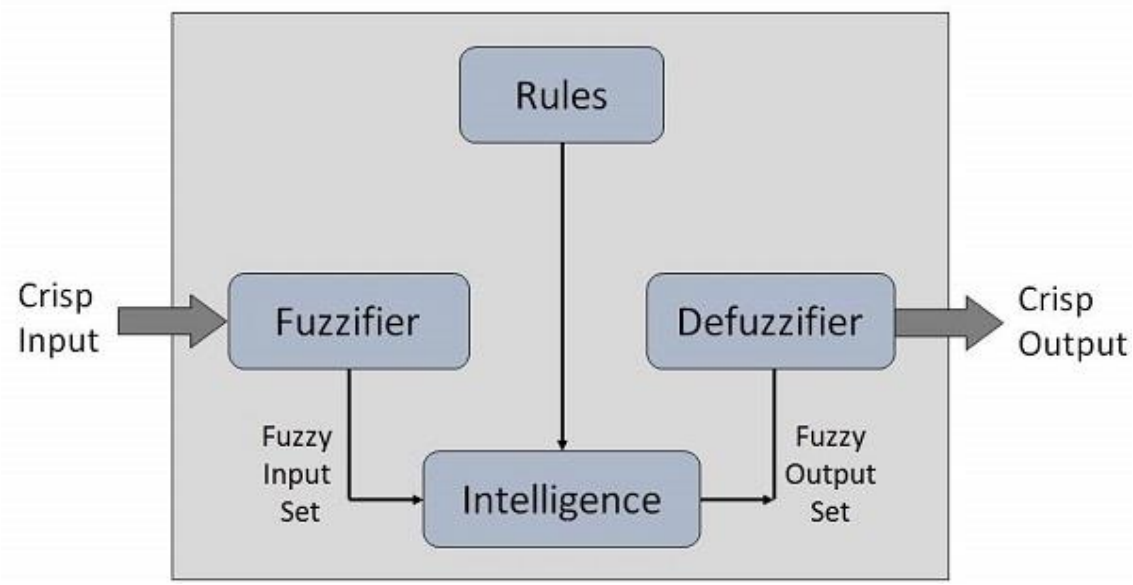
Fuzzy Logic Systems Architecture

It has four main parts as shown –

- **Fuzzification Module** − It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into five steps such as –

LP	x is Large Positive
MP	x is Medium Positive
S	x is Small
MN	x is Medium Negative
LN	x is Large Negative

- **Knowledge Base** − It stores IF-THEN rules provided by experts.
- **Inference Engine** − It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.
- **Defuzzification Module** − It transforms the fuzzy set obtained by the inference engine into a crisp value.



Application Areas of Fuzzy Logic

The key application areas of fuzzy logic are as given −

Automotive Systems

- Automatic Gearboxes
- Four-Wheel Steering
- Vehicle environment control

Consumer Electronic Goods

- Hi-Fi Systems
- Photocopiers
- Still and Video Cameras
- Television

Domestic Goods

- Microwave Ovens
- Refrigerators
- Toasters
- Vacuum Cleaners
- Washing Machines

Environment Control

- Air Conditioners/Dryers/Heaters
- Humidifiers

Advantages of FLSs

- Mathematical concepts within fuzzy reasoning are very simple.
- You can modify a FLS by just adding or deleting rules due to flexibility of fuzzy logic.
- Fuzzy logic Systems can take imprecise, distorted, noisy input information.
- FLSs are easy to construct and understand.
- Fuzzy logic is a solution to complex problems in all fields of life, including medicine, as it resembles human reasoning and decision making.

Disadvantages of FLSs

- There is no systematic approach to fuzzy system designing.
- They are understandable only when simple.
- They are suitable for the problems which do not need high accuracy.

APPLICATIONS of AI

AI has been dominant in various fields such as –

- **Gaming** − AI plays crucial role in strategic games such as chess, poker, tic-tac-toe, etc., where machine can think of large number of possible positions based on heuristic knowledge.

- **Natural Language Processing** − It is possible to interact with the computer that understands natural language spoken by humans.

- **Expert Systems** − There are some applications which integrate machine, software, and special information to impart reasoning and advising. They provide explanation and advice to the users.

- **Vision Systems** − These systems understand, interpret, and comprehend visual input on the computer. For example,

  - ❖ A spying aeroplane takes photographs, which are used to figure out spatial information or map of the areas.
  - ❖ Doctors use clinical expert system to diagnose the patient.
  - ❖ Police use computer software that can recognize the face of criminal with the stored portrait made by forensic artist.

- **Speech Recognition** − Some intelligent systems are capable of hearing and comprehending the language in terms of sentences and their meanings while a human talks to it. It can handle different accents, slang words, noise in the background, change in human's noise due to cold, etc.

- **Handwriting Recognition** −The handwriting recognition software reads the text written on paper by a pen or on screen by a stylus. It can recognize the shapes of the letters and convert it into editable text.

- **Intelligent Robots** − Robots are able to perform the tasks given by a human. They have sensors to detect physical data from the real world such as light, heat, temperature, movement, sound, bump, and pressure. They have efficient processors, multiple sensors and huge memory, to exhibit intelligence. In addition, they are capable of learning from their mistakes and they can adapt to the new environment.

SARDAR PATEL UNIVERSITY
B.C.A. SEMESTER – VI
US06CBCA25 : CURRENT TRENDS IN IT

Unit-III: Search Engine Optimization
Topics:
● Internet Basics
● Internet Marketing
● Search Engines Basics
● Search Engine Algorithm - Vector Space Model
● Using Search Engine
● Search Engine Optimization

INTERNET BASICS

What is World Wide Web?

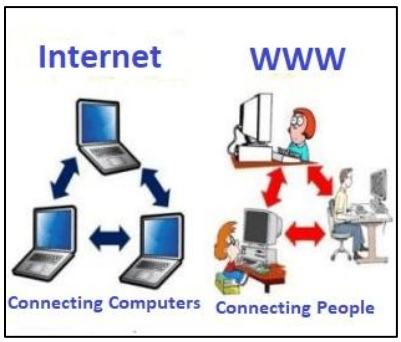
World Wide Web, which is also known as a Web, is a collection of websites or web pages stored in web servers and connected to local computers through the internet. These websites contain text pages, digital images, audios, videos, etc. Users can access the content of these sites from any part of the world over the internet using their devices such as computers, laptops, cell phones, etc. The WWW, along with internet, enables the retrieval and display of text and media to the device.

The building blocks of the Web are web pages which are formatted in HTML and connected by links called "hypertext" or hyperlinks and accessed by HTTP. These links are electronic connections that link related pieces of information so that users can access the desired information quickly. Hypertext offers the advantage to select a word or phrase from text and thus to access other pages that provide additional information related to that word or phrase.

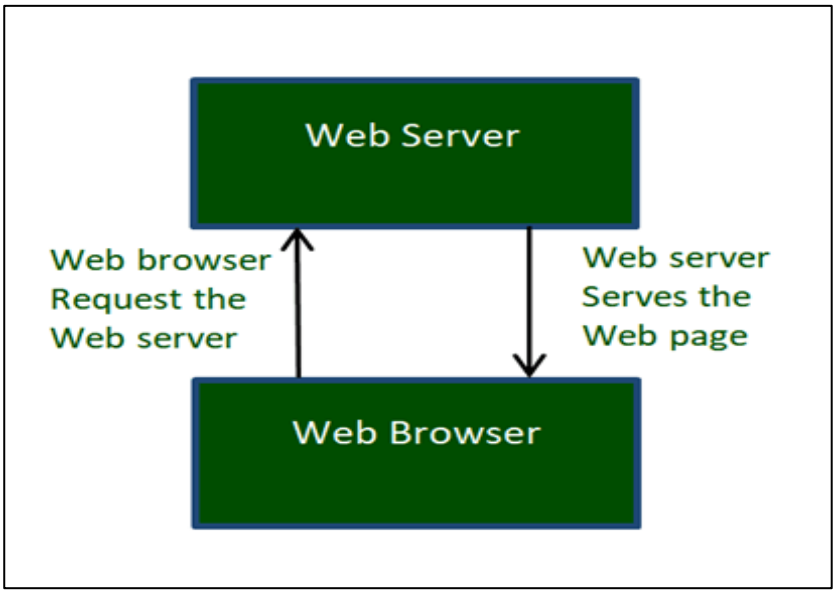
A web page is given an online address called a Uniform Resource Locator (URL). A particular collection of web pages that belong to a specific URL is called a website, e.g., *www.facebook.com*, *www.google.com*, etc. So, the World Wide Web is like a huge electronic book whose pages are stored on multiple servers across the world.

Small websites store all of their WebPages on a single server, but big websites or organizations place their WebPages on different servers in different countries so that when users of a country search their site they could get the information quickly from the nearest server.

So, the web provides a communication platform for users to retrieve and exchange information over the internet. Unlike a book, where we move from one page to another in a sequence, on World Wide Web we follow a web of hypertext links to visit a web page and from that web page to move to other web pages. You need a browser, which is installed on the computer, to access the Web.



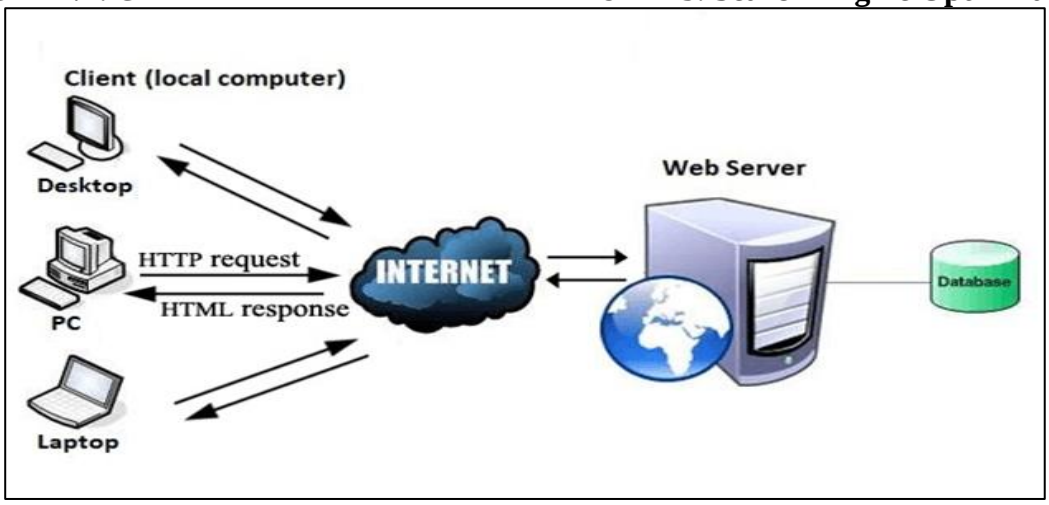
How the World Wide Web Works?



The servers store and transfer web pages or information to user's computers on the network when requested by the users. A web server is a software program, which serves the web pages requested by web users using a browser. The computer of a user who requests documents from a server is known as a client. Browser, which is installed on the user' computer, allows users to view the retrieved documents.

All the websites are stored in web servers. A website occupies a space in a server and remains stored in it. The server hosts the website whenever a user requests its Webpages, and the website owner has to pay the hosting price for the same.

There are three main technologies involved in transferring information (web pages) from servers to clients (computers of users). These technologies include Hypertext Markup Language (HTML), Hypertext Transfer Protocol (HTTP) and Web browsers.



What is a Website?

A website is the collection of web pages, which are interlinked together and are hosted from a single domain. A website is accessible via URL (Uniform Resource Locator), which organizes multiple web pages into a series that helps users to navigate through them. Mostly all websites have some important web pages like About us, Contact us, Home page, etc.

Types of Websites

- **Static Websites**

Static websites are the ones that have web pages with fixed content. The content remains the same throughout until any changes are made to it. These websites have simple plain text, or some may also have rich media content. Whenever a user visits any static website, the same content is seen, irrespective of the time of visit.

- **Dynamic Websites**

A dynamic website shows a variety of content, each time a user visits the website. These websites are updated automatically, depending on a set of parameters like templates, content, scripts, etc. So, every time a user visits a website, the content is renewed. These websites are created with the help of various software and languages like PHP, Python, JSP, etc.

- **Interactive Websites**

Websites that communicate and interact with users, using them, are known as interactive websites. These websites are rich with appealing graphics and engaging content which helps retains the users. Users can post their comments or leave feedback on these websites. Interactive websites are a great way to connect with your users, and most businesses prefer these types of websites. They not only help you gauge user's attention but also, are really helpful in boosting sales. The cost of developing these websites differs from that of static and dynamic ones.

What is a web portal?

Web portals are user-centric, and each user has its own password that can be used to access these portals. These portals have dynamic content, i.e. the content changes every time a user visits these portals. There are many web development companies that develop web portals based on your requirements.

Types of Web Portals

- **Horizontal Web Portals**

Also known as, Horizontal Enterprise Portals, these are the general portals that can be used as a common platform by several companies, to provide all kinds of information, which a user may require while looking for their services. Some of the common examples of horizontal web portals are government portals, educational portals, corporate portals, cultural portals, etc.

- **Vertical Web Portals**

Also known as Vertical Enterprise portals or vortals, these are the user-centric portals, that provide information related specific company, interest, or services. A user looking for information about any specific topic or about a particular company can find these types of portals helpful. If a business wants to share something specific related to their services, they can opt for expert and experienced web development services, for getting vertical portals developed, as per their business requirements. College portals, tender and bidding portals, are a few examples of vertical portals.

Difference between a Website and Web Portal

Both are quite different and can be differentiated in various aspects.

• A website focuses on attracting more users i.e. driving traffic whereas, a web portal is for specific users. It, in a way, restricts or limits the web traffic and allows only specific users to visit the portal.

• A website does not ask for a login ID and password from the users visiting them, but in a web portal, every time a user tries to open it, login credential is required. These credentials are user-specific and are different for every user.

• A website is publicly accessible i.e. anyone on the internet, who knows the URL, can visit the website and see its content, on the other hand, a web portal is privately accessible and only members of the portal can visit and see its content, using their login ID and password.

• The content on a website can be either dynamic or static, the content is dynamic and every time a user visits these portals, the content change.

INTERNET MARKETING

Definition

Internet marketing is the process of promoting a business or brand and its products or services over the internet using tools that help drive traffic, leads, and sales.

Internet marketing, a broad term that encompasses a range of marketing tactics and strategies – including content, email, search, paid media, and more.

- **e-Business**: All electronically mediated information exchanges, both within an organisation and with external stakeholders, supporting the range of business processes, e.g. e-Commerce, R&D, logistics, administration, etc.
- **Intranet**: A secure network within a single company that enables staff to access company information and communicate with collaborators using web tools, such as email and web browsers.
- **Extranet**: A secure network that enables internal and external stakeholders, such as staff, customers, suppliers, etc., to access information and communicate with collaborators using web tools, such as email and web browsers.

Types of internet marketing

There's a numerous of internet marketing types that encompass different tactics and strategies, and the few types are listed below.

- **Search Engine Optimization (SEO)**

SEO is the process of adjusting a website and digital content to improve its organic or "natural" placement in search rankings. The higher a webpage ranks, the more likely it is to be viewed by a potential customer. Search engines (specifically Google) use crawler bots (sometimes called spiders) to crawl the internet and build an index of the content available online. When a user searches a keyword, the search engine will provide the most relevant information.

There are two types of SEO: on-page and off-page. On-page SEO is the manipulations made directly to a web page to increase search engine ranking. It involves optimizing HTML code, content quality, and content structure. Off-page SEO is the SEO practices that take place outside of the website itself, such as backlinks, link relevancy, and social signals.

- **Content marketing**

Content marketing is the creation and distribution of relevant online content in a way that's strategically designed to attract and convert consumers. It focuses on communicating with customers rather than selling and is usually better received. Forms of content marketing include blog posts, infographics, ebooks, podcasts, case studies, and webinars.

- **Social media marketing**

Social media marketing is the use of social media platforms to improve customer engagement and promote a brand. While it doesn't necessarily drive sales, social media marketing increases engagement, builds links, and expands brand awareness. Popular social media platforms used for marketing include Facebook, Instagram, and Twitter.

- **Influencer marketing**

One of the newer types of internet marketing, influencer marketing uses influencers, or someone with a large social following, to promote their product or service for a price. This can be highly effective if the influencer is in line with a company's values and resonates with the company's customers.

- **Email marketing**

Email marketing is the process of using email to send direct marketing messages to consumers in an attempt to gain new customers and retain existing ones. It's one of the most cost effective types of marketing and can be used to reach both a wide network of customers or a very niche one.

- **Affiliate marketing**

Affiliate marketing describes any revenue-sharing plan where an online automated marketing program lets bloggers and website owners place an advertiser's banner ads, buttons, or other advertising media on their own website. This could also be in the form of promoting a product through a blog or video. A payment is received for every sale made through a link.

- **Paid advertising**

Paid advertising is when advertisers pay to show their advertisements of search engines and other online platforms. This is often referred to as Pay Per Click (PPC), meaning advertisers will pay a fee each time a user clicks on one of their ads. However, advertisers are now charged in different ways depending on their marketing objectives. Other means of charging include cost per thousand impressions, cost per view, and cost per action.

SEARCH ENGINES BASICS

What is SEO?

SEO stands for Search Engine Optimization. It is a process designed to optimize a website for search engines. It helps websites achieve a higher ranking in search engine results when people search for keywords related to their products and services. So, it is a practice of increasing the quantity and quality of traffic to a website through organic search engine results. The following image shows the basic activities involved in the SEO.

Search results are presented in the form of an ordered list, and the sites which are higher on the list tend to receive more traffic. For example, for a search query, the result which is at number one will receive 40 to 60% of the total traffic generated for that query. Only 2 to 3% visitors go beyond the first page of search results.

How do search engines work?

Search engines have three primary functions:

1. **Crawl:** Scour / search the Internet for content, looking over the code/content for each URL they find. Crawling is the discovery process in which search engines send out a team of robots (known as crawlers or spiders) to find new and updated content. Content can vary — it could be a webpage, an image, a video, a PDF, etc. — but regardless of the format, content is discovered by links.
2. **Index:** Store and organize the content found during the crawling process. Once a page is in the index, it is in the running to be displayed as a result to relevant queries. Search engines process and store information they find in an index, a huge database of all the content they've discovered and deem good enough to serve up to searchers.
3. **Rank:** Provide the pieces of content that will best answer a searcher's query, which means that results are ordered by most relevant to least relevant. When someone performs a search, search engines scour their index for highly relevant content and then orders that content in the hopes of solving the searcher's query. This ordering of search results by relevance is known as ranking. In general, you can assume that the higher a website is ranked, the more relevant the search engine believes that site is to the query.

Keyword Research & Keyword Targeting Best Practices

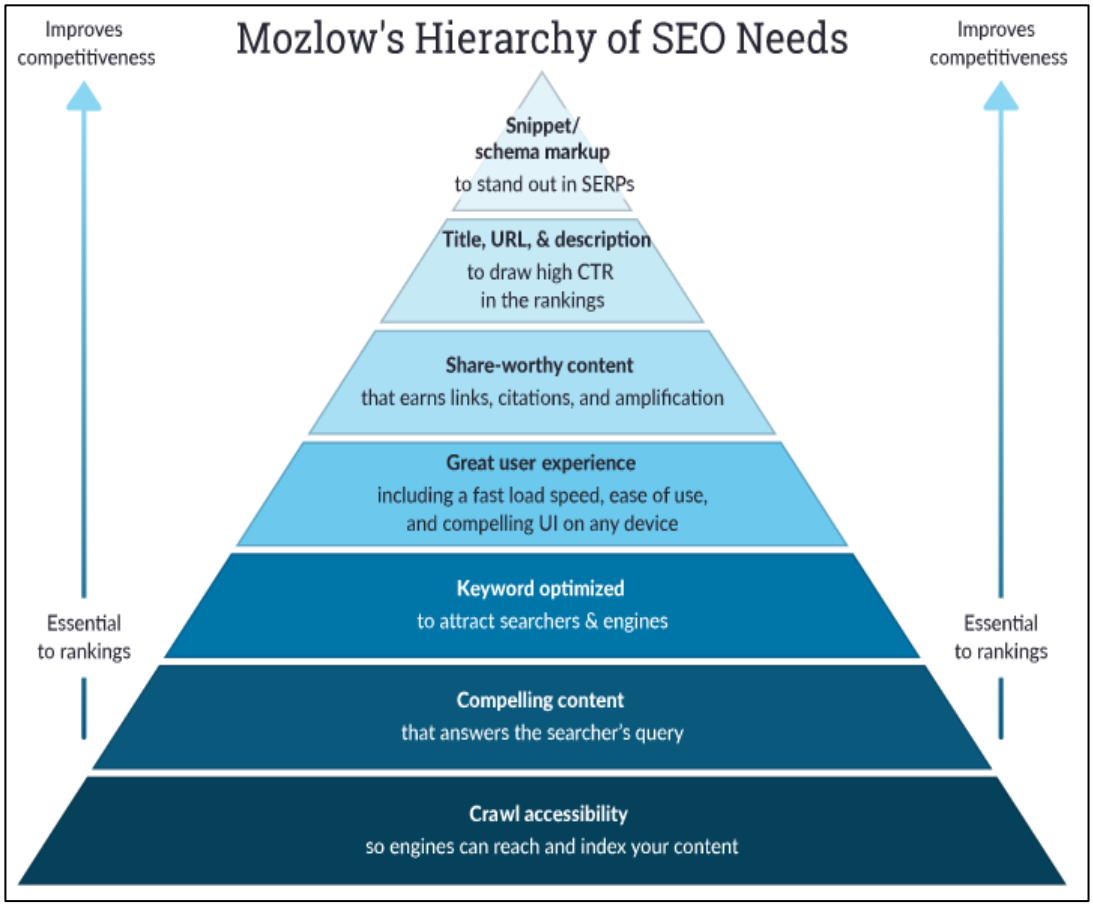
The first step in search engine optimization is to determine what it is actually optimizing for. This means identifying the terms people are searching for (also known as "keywords") that rank the website in search engines like Google.

There are a few key factors to take into account when determining the keywords you want to target on the site:

- **Search Volume** – The first factor to consider is how many people (if any) are actually searching for a given keyword. The more people there are searching for a keyword, the bigger the audience you stand to reach. Conversely, if no one is searching for a keyword, there is no audience available to find the content through search.
- **Relevance** – If a term is frequently searched for that's great: but what if it's not completely relevant to the prospects? Relevance seems straight-forward at first. For example, if you're selling enterprise email marketing automation software you don't want to show up for searches that don't have anything to do with the business, like "pet supplies." But what about terms like "email marketing software"? This might intuitively seem like a great description of what you do, but if you're selling to Fortune 100 companies, most of the traffic for this very competitive term will be searchers who don't have any interest in buying the software (and the person you do want to reach might never buy the expensive, complex solution based on a simple Google search). Conversely, you might think a tangential keyword like "best enterprise PPC marketing solutions" is totally irrelevant to the business since you don't sell PPC marketing software. But if the prospect is a CMO or marketing director, getting in front of them with a helpful resource on evaluating pay-per-click tools could be a great "first touch" and an excellent way to start a relationship with a prospective buyer.
- **Competition** – As with any business opportunity, in SEO you want to consider the potential costs and likelihood of success. For SEO, this means understanding the relative competition (and likelihood to rank) for specific terms.

SEARCH ENGINE OPTIMIZATION

Maslow's hierarchy of needs is a theory of psychology that prioritizes the most fundamental human needs (like air, water, and physical safety) over more advanced needs (like esteem and social belonging). Rand Fishkin made a similar pyramid to explain SEO, which is called "*Moz*low's hierarchy of SEO needs."



Using this beginner's guide, we can follow these seven steps to successful SEO:

1. Crawl accessibility so engines can read the website
2. Compelling content that answers the searcher's query
3. Keyword optimized to attract searchers & engines
4. Great user experience including a fast load speed and compelling UX
5. Share-worthy content that earns links, citations, and amplification
6. Title, URL, & description to draw high CTR in the rankings
7. Snippet/schema markup to stand out in SERPs

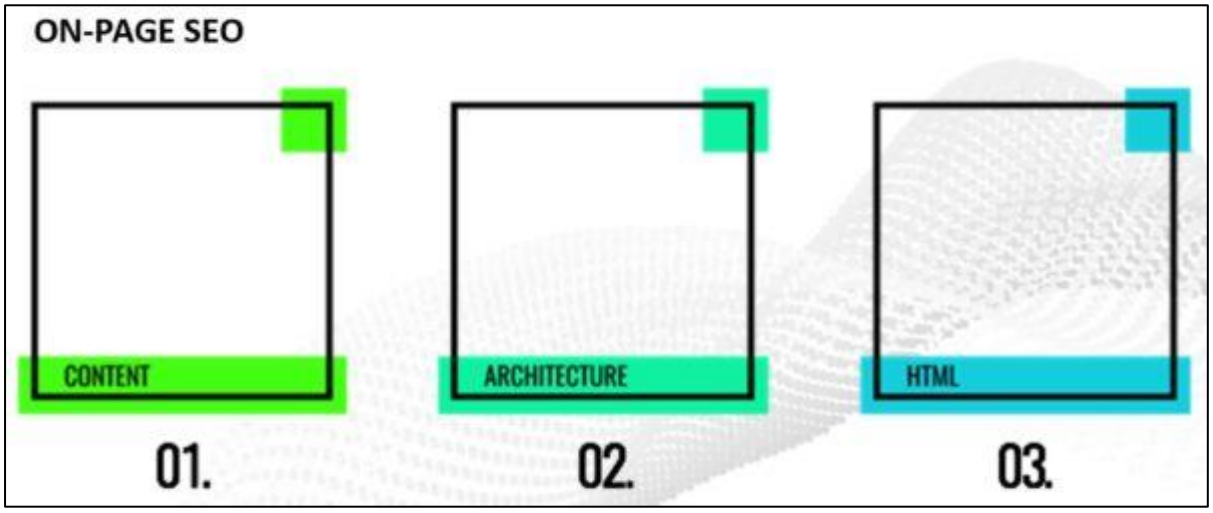
TYPES OF SEARCH ENGINE SEO FACTORS

There are four major groups covered by Search Engine Land's Periodic Table of SEO Factors:

- On-page SEO: Content, Architecture, HTML
- Off-page SEO: Trust, Links, User
- Toxins
- Emerging verticals

The elements within each group or subgroup are factors to consider to increase the site's organic visibility and rankings.

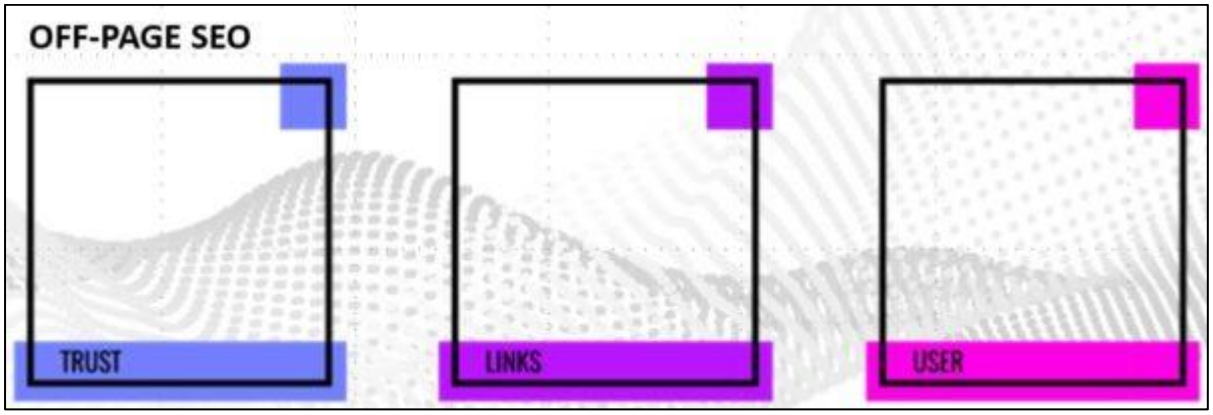
- **On-page SEO factors**



On-page search ranking factors are almost entirely within the publisher's control. This is also where it's critical to balance serving the needs of the audience with making the pages search engine friendly.

The title of the page or article, the depth of research, keywords used and so on should all be used with the specific audience's needs in mind. HTML headings, anchor text and more should provide clues for both search engines and the audience about the relevancy of the content. The site architecture should help search engine crawlers navigate the site and help users find what they're looking for.

- **Off-page SEO factors**



The search engines don't just evaluate what's on the page and visible to users. Off-page ranking factors are typically out of the creator or publisher's direct influence. Search engines evaluate reputation, the quality of a site's backlinks, the user's geographic location and many other factors to deliver the most relevant results.

Although these factors aren't as easy to control on a per-page basis, they must be taken into account when optimizing the site for search.

- **Toxins**

When done well, SEO benefits the search engines just as much as it benefits sites. SEO helps search engines provide users with better search results. However, using SEO techniques that aim to manipulate ranking signals to gain an unfair advantage over the competition can backfire.

We group spam and so-called "black hat" techniques into "toxins." Using them can result in the pages receiving a ranking penalty or even getting banned from the search results entirely.

- **Emerging verticals**

Voice, local, image and video search have their own special chapter in this guide because they represent different ways for users to find the content. Each of these emerging verticals has its own distinguishing features, nuances and opportunities for brands and publishers.

SEO ALGORITHM

A good search engine does not attempt to return the pages that best match the input query. A good search engine tries to answer the underlying question. Google search engines use a complex algorithm to determine what results they should return. The factors in the algorithm consist of "hard factors" as the number of backlinks to a page and perhaps some social recommendations through likes and +1' s. These are usually external influences. You also have the factors on the page itself. For this, the way a page is build and various page elements play a role in the algorithm. But only by analyzing the on-site and off-site factors is it possible for Google to determine which pages will answer is the question behind the query. For this Google will have to analyze the text on a page. Search engines have evolved tremendously in recent years, but at first they could only deal with Boolean operators. In simple terms, a term was included in a document or not. Something was true or false, 1 or 0. Additionally you could use the operators as AND, OR and NOT to search documents that contain multiple terms or to exclude terms.

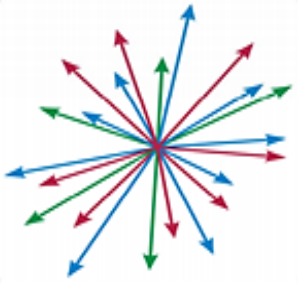
VECTOR SPACE MODEL

- Very popular model, even today
  - Simple, intuitive
  - Useful for weighting, ranking, and relevance feedback
- Documents and query represented by a vector of term weights
  - t is number of index terms (i e very large) is number of index terms (i.e., very large)

$$D_i = (d_{i1}, d_{i2}, \ldots, d_{it}) \qquad Q = (q_1, q_2, \ldots, q_t)$$

- Collection represented by a matrix of term weights

	$Term_1$	$Term_2$	$\ldots$	$Term_t$
Doc_1	d_{11}	d_{12}	$\ldots$	d_{1t}
Doc_2	d_{21}	d_{22}	$\ldots$	d_{2t}
$\vdots$	$\vdots$			
Doc_n	d_{n1}	d_{n2}	$\ldots$	d_{nt}



- D1: Tropical Freshwater Aquarium Fish.
- D2: Tropical Fish, Aquarium Care, Tank Setup.
- D3: Keeping Tropical Fish and Goldfish in Aquariums, and Fish Bowls.
- D4: The Tropical Tank Homepage - Tropical Fish and Aquariums.

Terms	Documents			
	D_1	D_2	D_3	D_4
aquarium	1	1	1	1
bowl	0	0	1	0
care	0	1	0	0
fish	1	1	2	1
freshwater	1	0	0	0
goldfish	0	0	1	0
homepage	0	0	0	1
keep	0	0	1	0
setup	0	1	0	0
tank	0	1	0	1
tropical	1	1	1	2

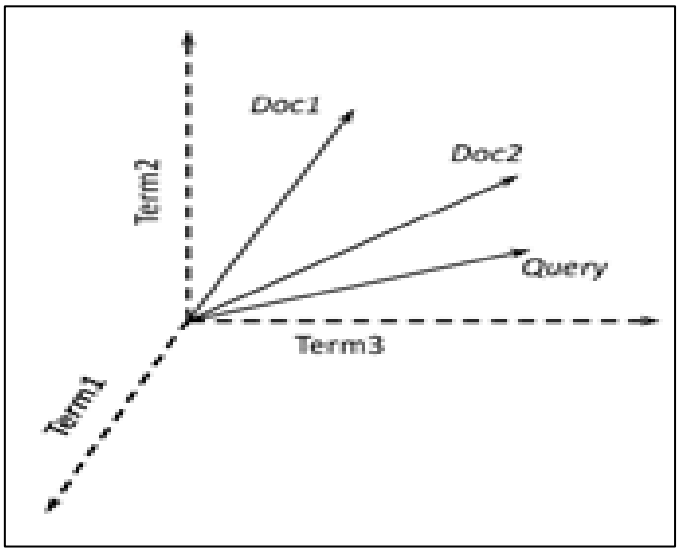
Rotated

Stopwords are removed

Weights are term counts

Query for „tropical fish" (0 0 0 1 0 0 0 0 0 0 1)

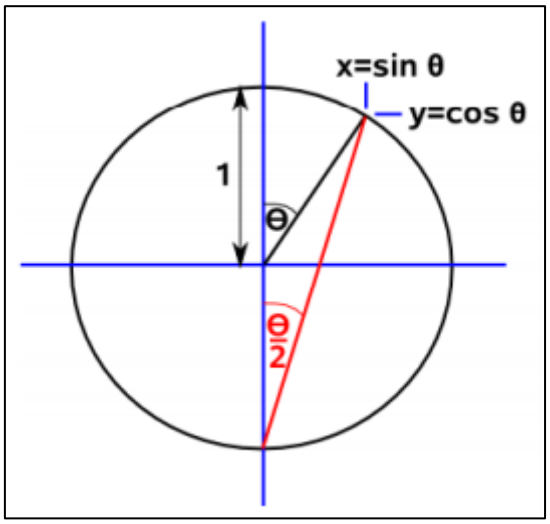
- 3-d pictures useful, but can be misleading for high-dimensional space
  - Intuition no longer necessarily correct
  - Millions of terms (and dimensions)



- Each document ranked by distance between points representing query and document
  - Similarity measure more common than a distance or dissimilarity measure
  - Popular: Cosine correlation
    - Cosine of angle between document and query vectors
    - Normalized dot-product

$$Cosine(D_i, Q) = \frac{\sum_{j=1}^{t} d_{ij} \cdot q_j}{\sqrt{\sum_{j=1}^{t} {d_{ij}}^2 \cdot \sum_{j=1}^{t} {q_j}^2}}$$

- As retrieval model: No explicit definition of relevance
    - Implicit: Closer documents are more relevant.



Unit- 4 IoT and Cloud Computing

- **Introduction to Internet of Things**
- **Introduction to Cloud computing & Evolution of Cloud Computing**
- **Applications of Cloud Computing**
- **Benefits –Limitations of Cloud Computing**
- **Cloud Services**
- **Cloud Computing hardware and infrastructure**

Introduction of Internet of Things

What is Internet of Things (IoT)?

Internet of Things (IoT) is a network of physical objects or people called "things" that are embedded with software, electronics, network, and sensors that allows these objects to collect and exchange data. The goal of IoT is to extend to internet connectivity from standard devices like computer, mobile, tablet to relatively dumb devices like a toaster.

IoT makes virtually everything "smart," by improving aspects of our life with the power of data collection, AI algorithm, and networks. The thing in IoT can also be a person with a diabetes monitor implant, an animal with tracking devices, etc.

Components of IoT

The entire IoT process starts with the devices themselves like smartphones, smartwatches, electronic appliances like TV, Washing Machine which helps you to communicate with the IoT platform. There are four fundamental components of IoT: Sensors/Devices, Connectivity, Data Processing and User Interface.

1. **Sensors/Devices:** Sensors or devices are a key component that helps you to collect live data from the surrounding environment. All this data may have various levels of complexities. It could be a simple temperature monitoring sensor, or it may be in the form of the video feed.

A device may have various types of sensors which performs multiple tasks apart from sensing. Example, A mobile phone is a device which has multiple sensors like GPS, camera but your smartphone is not able to sense these things.

2. **Connectivity:** All the collected data is sent to a cloud infrastructure. The sensors should be connected to the cloud using various mediums of communications. These communication mediums include mobile or satellite networks, Bluetooth, WI-FI, WAN, etc.

3. **Data Processing:** Once that data is collected, and it gets to the cloud, the software performs processing on the gathered data. This process can be just checking the temperature, reading on devices like AC or heaters. However, it can sometimes also be very complex like identifying objects, using computer vision on video.

4. **User Interface:** The information needs to be available to the end-user in some way which can be achieved by triggering alarms on their phones or sending them notification through email or text message. The user sometimes might need an interface which actively checks their IoT system. For example, the user has a camera installed in his home. He wants to access video recording and all the feeds with the help of a web server.

IoT Applications

IoT solutions are widely used in numerous companies across industries. Some most common IoT applications are given below:

Application Type	Description
Smart Thermostats	Helps you to save resource on heating bills by knowing your usage patterns.
Connected Cars	IoT helps automobile companies handle billing, parking, insurance, and other related stuff automatically.
Activity Trackers	Helps you to capture heart rate pattern, calorie expenditure, activity levels, and skin temperature on your wrist.
Smart Outlets	Remotely turn any device on or off. It also allows you to track a device's energy level and get custom notifications directly into your smartphone.
Parking Sensors	IoT technology helps users to identify the real-time availability of parking spaces on their phone.
Connect Health	The concept of a connected health care system facilitates real-time health monitoring and patient care. It helps in improved medical decision-making based on patient data.
Smart City	Smart city offers all types of use cases which include traffic management to water distribution, waste management, etc.

Smart home	Smart home encapsulates the connectivity inside your homes. It includes smoke detectors, home appliances, light bulbs, windows, door locks, etc.
Smart supply chain	Helps you in real time tracking of goods while they are on the road, or getting suppliers to exchange inventory information.

Challenges of IoT

Following are some challenges may be faced when implementing the IoT.

- Insufficient testing and updating
- Concern regarding data security and privacy
- Software complexity
- Data volumes and interpretation
- Integration with AI and automation
- Devices require a constant power supply which is difficult
- Interaction and short-range communication

Advantages of IoT

- **Technical Optimization:** IoT technology helps a lot in improving technologies and making them better. Example, with IoT, a manufacturer is able to collect data from various car sensors. The manufacturer analyzes them to improve its design and make them more efficient.

- **Improved Data Collection:** Traditional data collection has its limitations and its design for passive use. IoT facilitates immediate action on data.
- **Reduced Waste:** IoT offers real-time information leading to effective decision making & management of resources. For example, if a manufacturer finds an issue in multiple car engines, he can track the manufacturing plan of those engines and solves this issue with the manufacturing belt.

- **Improved Customer Engagement:** IoT allows you to improve customer experience by detecting problems and improving the process.

Disadvantages of IoT

- **Security:** IoT technology creates an ecosystem of connected devices. However, during this process, the system may offer little authentication control despite sufficient security measures.
- **Privacy:** The use of IoT, exposes a substantial amount of personal data, in extreme detail, without the user's active participation. This creates lots of privacy issues.

- **Flexibility:** There is a huge concern regarding the flexibility of an IoT system. It is mainly regarding integrating with another system as there are many diverse systems involved in the process.

- **Complexity:** The design of the IoT system is also quite complicated. Moreover, it's deployment and maintenance also not very easy.

- **Compliance:** IoT has its own set of rules and regulations. However, because of its complexity, the task of compliance is quite challenging.

Some Important Definitions

- **Raspberry Pi:** Raspberry Pi is a computer which is capable of doing all the operations like a conventional computer. It has other features such as onboard WiFi, GPIO pins, and Bluetooth in order to communicate with external things.

- **Arduino:** Arduino is a free electronics platform having easy to use hardware and software. It has a microcontroller capable of reading input from sensors to control the motors programmatically.

- **MicroPython:** MicroPython is a Python implementation, which includes a small subset of its standard library. It can be optimized to run on the ModeMCU microcontroller.

- **Thingworx:** Thingworx is a platform for the fast development and deployment of connected devices. It is a collection of integrated IoT development tools that support analysis, production, property, and alternative aspects of IoT development.

- **Contiki:** IoT Contiki is software that targets explicitly little devices connected with the Internet. It is used with process power bandwidth, power, and restricted memory. Contiki helps for the management of programs, resources, processes, communication, and memory.

- **Zigbee:** Zigbee is the same like Bluetooth. It used in a complex system for low power operation, robustness, and high security.

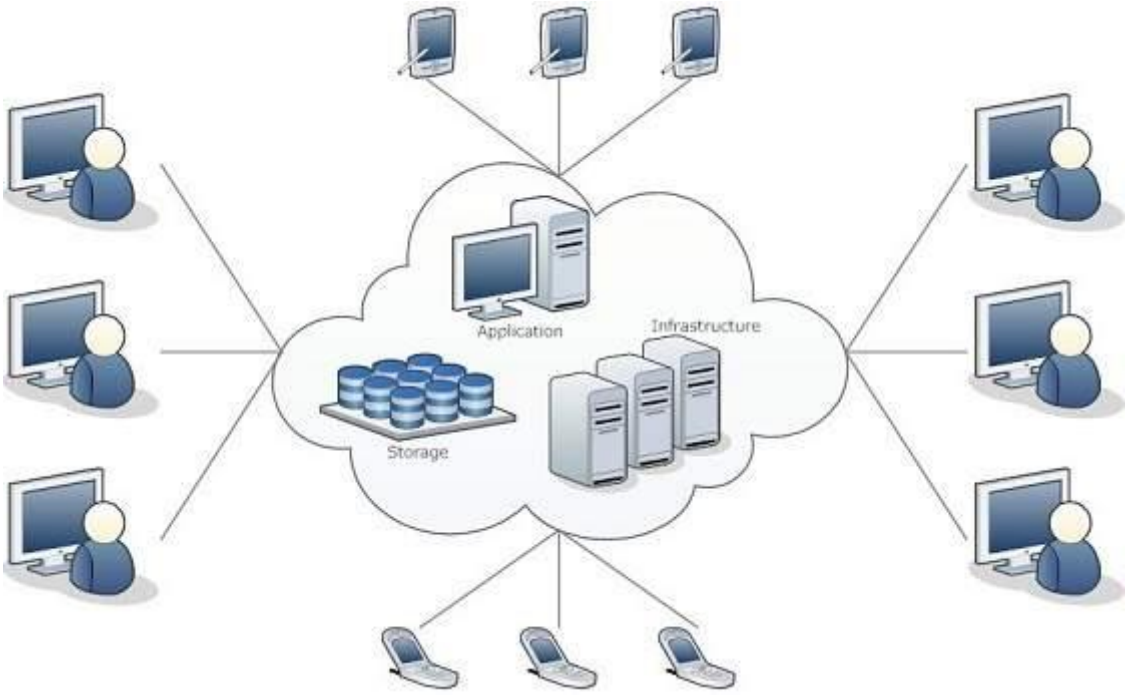
Introduction & Evolution of Cloud Computing

What is Cloud?

The term Cloud refers to a Network or Internet. In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over public and private networks, i.e., WAN, LAN or VPN. Applications such as e-mail, web conferencing, customer relationship management (CRM) execute on cloud.

What is Cloud Computing?

Cloud Computing refers to manipulating, configuring, and accessing the hardware and software resources remotely. It offers online data storage, infrastructure, and application.



Cloud computing offers platform independency, as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications mobile and collaborative.

History of Cloud Computing

Before emerging the cloud computing, there was Client/Server computing which is basically a centralized storage in which all the software applications, all the data and all the controls are resided on the server side. If a single user wants to access specific data or run a program, he/she need to connect to the server and then gain appropriate access, and then he/she can do his/her business.

Then after, distributed computing came into picture, where all the computers are networked together and share their resources when needed. On the basis of above computing, there was emerged of cloud computing concepts that later implemented.

At around in 1961, John MacCharty suggested in a speech at MIT that computing can be sold like a utility, just like a water or electricity. It was a brilliant idea, but like all brilliant ideas, it was ahead if its time, as for the next few decades, despite interest in the model, the technology simply was not ready for it.

In 1999, Salesforce.com started delivering of applications to users using a simple website. The applications were delivered to enterprises over the Internet, and this way the dream of computing sold as utility were true.

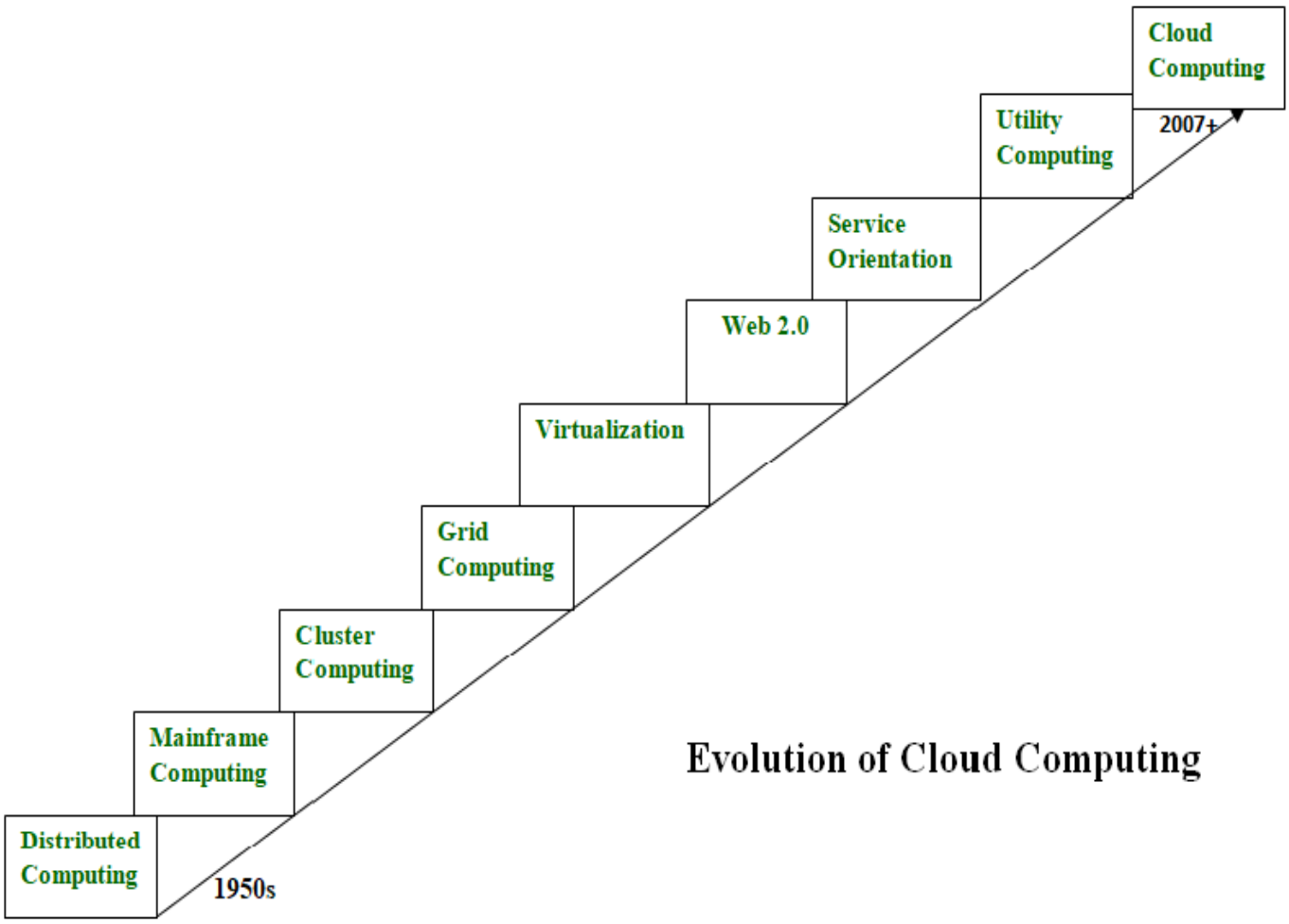
In 2002, Amazon started Amazon Web Services, providing services like storage, computation and even human intelligence. However, only starting with the launch of the Elastic Compute Cloud in 2006 a truly commercial service open to everybody existed.

In 2009, Google Apps also started to provide cloud computing enterprise applications.

In 2009, Microsoft launched Windows Azure, and companies like Oracle and HP have all joined the cloud computing and it has become mainstream.

Evolution of Cloud Computing

Cloud computing is all about renting computing services. This idea first came in the 1950s. In making cloud computing what it is today, five technologies played a vital role. These are distributed systems and its peripherals, virtualization, web 2.0, service orientation, and utility computing.



Evolution of Cloud Computing

- **Distributed Computing:** It is a composition of multiple independent systems but all of them are depicted as a single entity to the users. The purpose of distributed systems is to share resources and also use them effectively and efficiently. Distributed systems possess characteristics such as scalability, concurrency, continuous availability, heterogeneity, and independence in failures. But the main problem with this system was that all the systems were required to be present at the same geographical location. Thus, to solve this problem, distributed computing led to three more types of computing and they were-Mainframe computing, cluster computing, and grid computing.

- **Mainframe Computing:** Mainframes which first came into existence in 1951 are highly powerful and reliable computing machines. These are responsible for

handling large data such as massive input-output operations. Even today these are used for bulk processing tasks such as online transactions etc. These systems have almost no downtime with high fault tolerance. After distributed computing, these increased the processing capabilities of the system. But these were very expensive. To reduce this cost, cluster computing came as an alternative to mainframe technology.

- **Cluster Computing:** In 1980s, cluster computing came as an alternative to mainframe computing. Each machine in the cluster was connected to each other by a network with high bandwidth. These were way cheaper than those mainframe systems. These were equally capable of high computations. Also, new nodes could easily be added to the cluster if it was required. Thus, the problem of the cost was solved to some extent but the problem related to geographical restrictions still pertained. To solve this, the concept of grid computing was introduced.

- **Grid computing:** In 1990s, the concept of grid computing was introduced. It means that different systems were placed at entirely different geographical locations and these all were connected via the internet. These systems belonged to different organizations and thus the grid consisted of heterogeneous nodes. Although it solved some problems but new problems emerged as the distance between the nodes increased. The main problem which was encountered was the low availability of high bandwidth connectivity and with-it other network associated issues. Thus. cloud computing is often referred to as “Successor of grid computing”.

- **Virtualization:** It was introduced nearly 40 years back. It refers to the process of creating a virtual layer over the hardware which allows the user to run multiple instances simultaneously on the hardware. It is a key technology used in cloud computing. It is the base on which major cloud computing services such as Amazon EC2, VMware vCloud, etc work on. Hardware virtualization is still one of the most common types of virtualization.

- **Web 2.0:** It is the interface through which the cloud computing services interact with the clients. It is because of Web 2.0 that we have interactive and dynamic web pages. It also increases flexibility among web pages. Popular examples of web 2.0 include Google Maps, Facebook, Twitter, etc. Needless to say, social media is possible because of this technology only. In gained major popularity in 2004.

- **Service orientation:** It acts as a reference model for cloud computing. It supports low-cost, flexible, and evolvable applications. Two important concepts were introduced in this computing model. These were Quality of Service (QoS) which also includes the SLA (Service Level Agreement) and Software as a Service (SaaS).

- **Utility computing:** It is a computing model that defines service provisioning techniques for services such as compute services along with other major services such as storage, infrastructure, etc which are provisioned on a pay-per-use basis.

Applications of Cloud Computing

Cloud Computing has its applications in almost all the fields such as business, entertainment, data storage, social networking, management, entertainment, education, art and global positioning system, etc. Some of the widely famous cloud computing applications are as below:

	Application Type	Application Name	Description
1	Business Applications	MailChimp	It offers an e-mail publishing platform. It is widely employed by the businesses to design and send their e-mail campaigns.
2		Chatter	Chatter app helps the employee to share important information about organization in real time. One can get the instant feed regarding any issue.
3		Google Apps for Business	Google offers creating text documents, spreadsheets, presentations, etc., on Google Docs which allows the business users to share them in collaborating manner.
4		Quickbooks	It offers online accounting solutions for a business. It helps in monitoring cash flow, creating VAT returns and creating business reports.
5	Data Storage and Backup	Box.com	Box.com offers drag and drop service for files. The users need to drop the files into Box and access from anywhere.
6		Mozy	Mozy offers online backup service for files to prevent data loss.
7		Joukuu	Joukuu is a web-based interface. It allows to display a single list of contents for files stored in Google Docs, Box.net and Dropbox.
8	Management Applications	Toggl	It helps in tracking time period assigned to a particular project.
9		Evernote	It organizes the sticky notes and even can read the text from images which helps the user to locate the notes easily.
10		Outright	It is an accounting app. It helps to track income, expenses, profits and losses in real time.
11	Social Applications	Facebook	It offers social networking service. One can share photos, videos, files, status and much more.
12		Twitter	It helps to interact with the public directly. One can follow any celebrity, organization and any person, who is on twitter and can have latest updates regarding the same.

13	**Entertainment Applications**	Audio box.fm	It offers streaming service. The music files are stored online and can be played from cloud using the own media player of the service.
14	**Art Applications**	Moo	It offers art services such as designing and printing business cards, postcards and mini cards.

Benefits –Limitations of Cloud Computing

Advantage of Cloud Computing

	Advantage	Description
1	**Back-up and restore data**	Once the data is stored in the cloud, it is easier to get back-up and restore that data using the cloud.
2	**Improved collaboration**	Cloud applications improve collaboration by allowing groups of people to quickly and easily share information in the cloud via shared storage.
3	**Excellent accessibility**	Cloud allows us to quickly and easily access store information anywhere, anytime in the whole world, using an internet connection. An internet cloud infrastructure increases organization productivity and efficiency by ensuring that our data is always accessible.
4	**Low maintenance cost**	Cloud computing reduces both hardware and software maintenance costs for organizations.
5	**Mobility**	Cloud computing allows us to easily access all cloud data via mobile.
6	**Services in the pay-per-use model**	Cloud computing offers Application Programming Interfaces (APIs) to the users for access services on the cloud and pays the charges as per the usage of service.
7	**Unlimited storage capacity**	Cloud offers us a huge amount of storing capacity for storing our important data such as documents, images, audio, video, etc. in one place.
8	**Data security**	Data security is one of the biggest advantages of cloud computing. Cloud offers many advanced features related to security and ensures that data is securely stored and handled.

Disadvantages of Cloud Computing

	Disadvantage	Description
1	**Internet Connectivity**	As you know, in cloud computing, every data (image, audio, video, etc.) is stored on the cloud, and we access these data through the cloud by using the internet connection. If you do not have good internet connectivity, you cannot access these data. However, we have no any other way to access data from the cloud.

2	**Vendor lock-in**	Vendor lock-in is the biggest disadvantage of cloud computing. Organizations may face problems when transferring their services from one vendor to another. As different vendors provide different platforms, that can cause difficulty moving from one cloud to another.
3	**Limited Control**	As we know, cloud infrastructure is completely owned, managed, and monitored by the service provider, so the cloud users have less control over the function and execution of services within a cloud infrastructure.
4	**Security**	Although cloud service providers implement the best security standards to store important information. But, before adopting cloud technology, you should be aware that you will be sending all your organization's sensitive information to a third party, i.e., a cloud computing service provider. While sending the data on the cloud, there may be a chance that your organization's information is hacked by Hackers.

Cloud Services

There are the following three types of cloud service models

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

Infrastructure as a Service (IaaS)

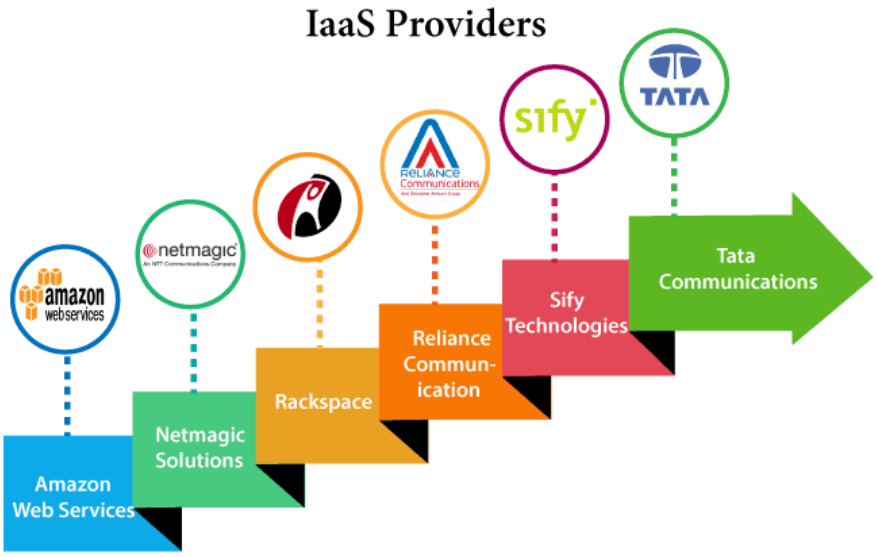
Iaas is also known as Hardware as a Service (HaaS). It is one of the layers of the cloud computing platform. It allows customers to outsource their IT infrastructures such as servers, networking, processing, storage, virtual machines, and other resources. Customers access these resources on the Internet using a pay-as-per use model.

In traditional hosting services, IT infrastructure was rented out for a specific period of time, with pre-determined hardware configuration. The client paid for the configuration and time, regardless of the actual use. With the help of the IaaS cloud computing platform layer, clients can dynamically scale the configuration to meet changing requirements and are billed only for the services actually used.

IaaS cloud computing platform layer eliminates the need for every organization to maintain the IT infrastructure. IaaS provider provides the following services -

- **Compute:** Computing as a Service includes virtual central processing units and virtual main memory for the Vms that is provisioned to the end- users.
- **Storage:** IaaS provider provides back-end storage for storing files.
- **Network:** Network as a Service (NaaS) provides networking components such as routers, switches, and bridges for the Vms.
- **Load balancers:** It provides load balancing capability at the infrastructure layer.

Top Iaas Providers



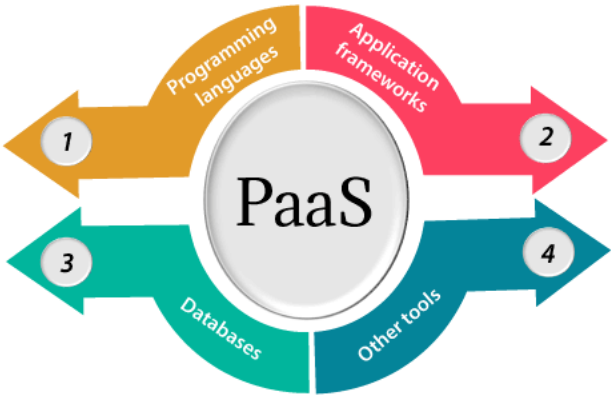
Platform as a Service (PaaS)

Platform as a Service (PaaS) provides a runtime environment. It allows programmers to easily create, test, run, and deploy web applications. You can purchase these applications from a cloud service provider on a pay-as-per use basis and access them using the Internet connection. In PaaS, back end scalability is managed by the cloud service provider, so end- users do not need to worry about managing the infrastructure.

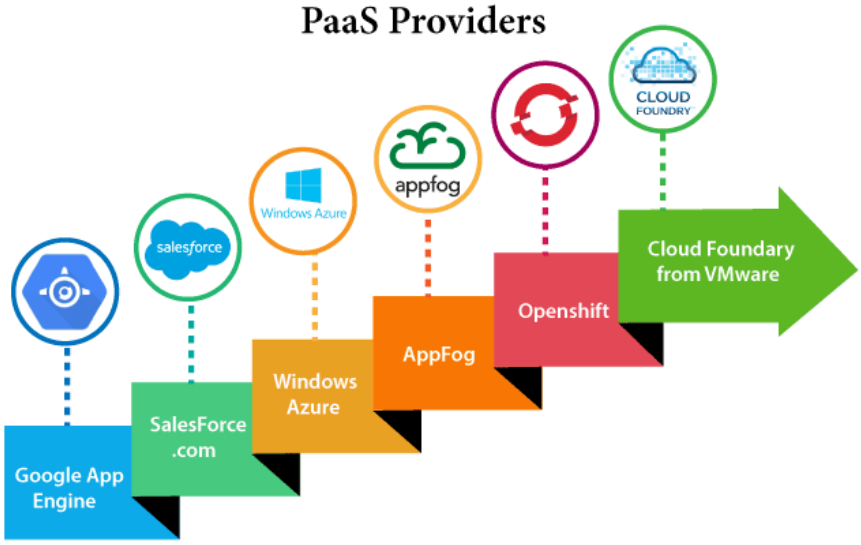
PaaS includes infrastructure (servers, storage, and networking) and platform (middleware, development tools, database management systems, business intelligence, and more) to support the web application life cycle. PaaS providers provide the Programming languages, Application frameworks, Databases, and Other tools:

- **Programming languages:** PaaS providers provide various programming languages for the developers to develop the applications. Some popular programming languages provided by PaaS providers are Java, PHP, Ruby, Perl, and Go.
- **Application frameworks:** PaaS providers provide application frameworks to easily understand the application development. Some popular application frameworks provided by PaaS providers are Node.js, Drupal, Joomla, WordPress, Spring, Play, Rack, and Zend.
- **Databases:** PaaS providers provide various databases such as ClearDB, PostgreSQL, MongoDB, and Redis to communicate with the applications.

- **Other tools:** PaaS providers provide various other tools that are required to develop, test, and deploy the applications.



Popular PaaS Providers



Software as a Service (SaaS)

SaaS is also known as "On-Demand Software". It is a software distribution model in which services are hosted by a cloud service provider. These services are available to end-users over the internet so, the end-users do not need to install any software on their devices to access these services. There are the following services provided by SaaS providers -

- **Business Services** - SaaS Provider provides various business services to start-up the business. The SaaS business services include ERP (Enterprise Resource Planning), CRM (Customer Relationship Management), billing, and sales.
- **Document Management** - SaaS document management is a software application offered by a third party (SaaS providers) to create, manage, and track electronic documents.
- **Social Networks** - As we all know, social networking sites are used by the general public, so social networking service providers use SaaS for their convenience and handle the general public's information.
- **Mail Services** - To handle the unpredictable number of users and load on e-mail services, many e-mail providers offering their services using SaaS.

Popular SaaS Providers

Difference between IaaS, PaaS, and SaaS

IaaS	Paas	SaaS
It provides a virtual data center to store information and create platforms for app development, testing, and deployment.	It provides virtual platforms and tools to create, test, and deploy apps.	It provides web software and apps to complete business tasks.

It provides access to resources such as virtual machines, virtual storage, etc.	It provides runtime environments and deployment tools for applications.	It provides software as a service to the end-users.
It is used by network architects.	It is used by developers.	It is used by end users.
IaaS provides only Infrastructure.	PaaS provides Infrastructure+Platform.	SaaS provides Infrastructure+Platform +Software.

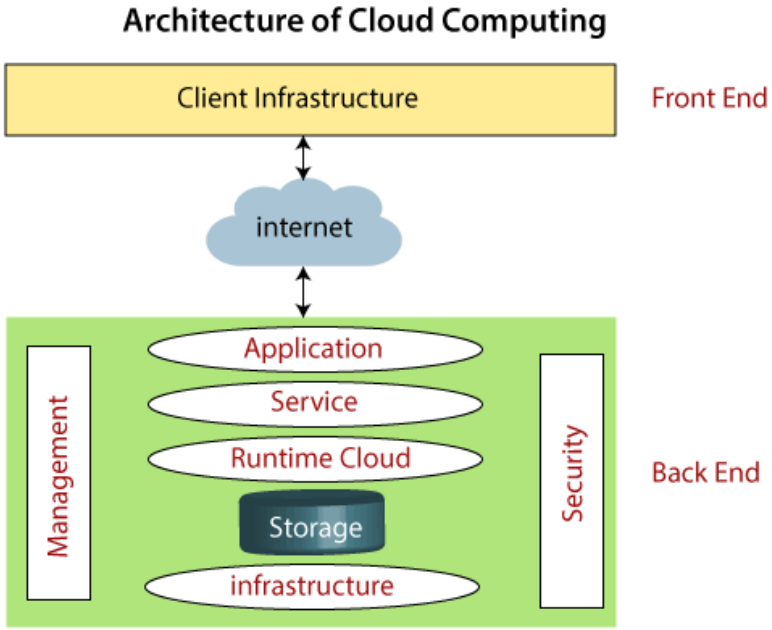
Cloud Computing hardware and infrastructure

Cloud Computing Architecture

Cloud computing architecture is a combination of **service-oriented architecture** and **event-driven architecture**. Cloud computing architecture is divided into the following two parts –

- **Front End**
- **Back End**

The below diagram shows the architecture of cloud computing -



Front End: The front end is used by the client. It contains client-side interfaces and applications that are required to access the cloud computing platforms. The front end includes web servers (including Chrome, Firefox, internet explorer, etc.), thin & fat clients, tablets, and mobile devices.

Back End: The back end is used by the service provider. It manages all the resources that are required to provide cloud computing services. It includes a huge amount of data storage, security mechanism, virtual machines, deploying models, servers, traffic control mechanisms, etc.

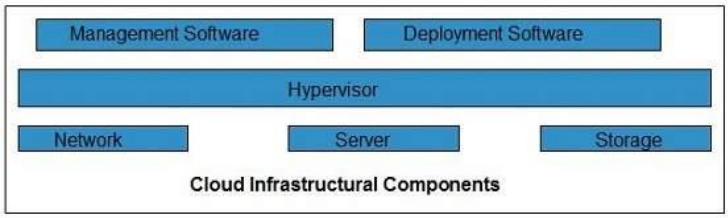
Components of Cloud Computing Architecture

There are the following components of cloud computing architecture –

1. **Client Infrastructure:** Client Infrastructure is a Front-end component. It provides GUI (Graphical User Interface) to interact with the cloud.
2. **Application:** The application may be any software or platform that a client wants to access.
3. **Service:** A Cloud Services manages that which type of service you access according to the client's requirement. Cloud computing offers the following three type of services:
    a. Software as a Service (SaaS)
    b. Platform as a Service (PaaS)
    c. Infrastructure as a Service (IaaS)
4. **Runtime Cloud:** Runtime Cloud provides the execution and runtime environment to the virtual machines.
5. **Storage:** Storage is one of the most important components of cloud computing. It provides a huge amount of storage capacity in the cloud to store and manage data.
6. **Infrastructure:** It provides services on the host level, application level, and network level. Cloud infrastructure includes hardware and software components such as servers, storage, network devices, virtualization software, and other storage resources that are needed to support the cloud computing model.
7. **Management:** Management is used to manage components such as application, service, runtime cloud, storage, infrastructure, and other security issues in the backend and establish coordination between them.
8. **Security:** Security is an in-built back end component of cloud computing. It implements a security mechanism in the back end.
9. **Internet:** The Internet is medium through which front end and back end can interact and communicate with each other.

Cloud Computing Infrastructure

Cloud infrastructure consists of servers, storage devices, network, cloud management software, deployment software, and platform virtualization.

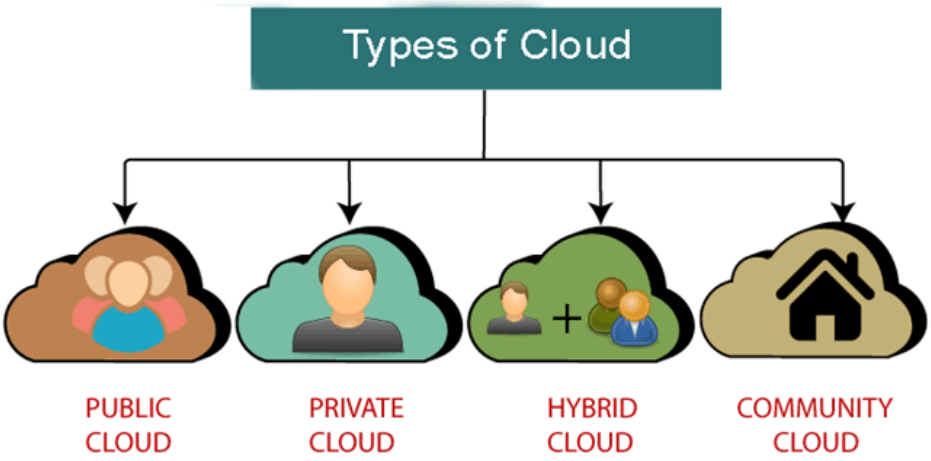


- **Hypervisor:** Hypervisor is a firmware or low-level program that acts as a Virtual Machine Manager. It allows to share the single physical instance of cloud resources between several tenants.
- **Management Software:** It helps to maintain and configure the infrastructure.
- **Deployment Software:** It helps to deploy and integrate the application on the cloud.
- **Network:** It is the key component of cloud infrastructure. It allows to connect cloud services over the Internet. It is also possible to deliver network as a utility over the Internet, which means, the customer can customize the network route and protocol.

- **Server:** The server helps to compute the resource sharing and offers other services such as resource allocation and de-allocation, monitoring the resources, providing security etc.
- **Storage:** Cloud keeps multiple replicas of storage. If one of the storage resources fails, then it can be extracted from another one, which makes cloud computing more reliable.

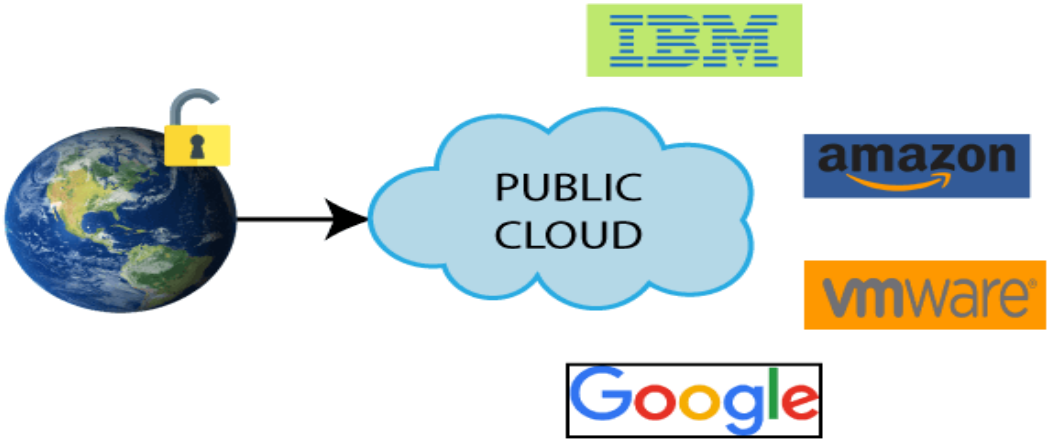
Types of Clouds

There are the following 4 types of cloud that you can deploy according to the organization's needs-



Public Cloud

- Public Cloud provides a shared platform that is accessible to the general public through an Internet connection.
- Public cloud operated on the pay-as-per-use model and administrated by the third party, i.e., Cloud service provider.
- In the Public cloud, the same storage is being used by multiple users at the same time.
- Public cloud is owned, managed, and operated by businesses, universities, government organizations, or a combination of them.
- Amazon Elastic Compute Cloud (EC2), Microsoft Azure, IBM's Blue Cloud, Sun Cloud, and Google Cloud are examples of the public cloud.



Advantages of Public Cloud

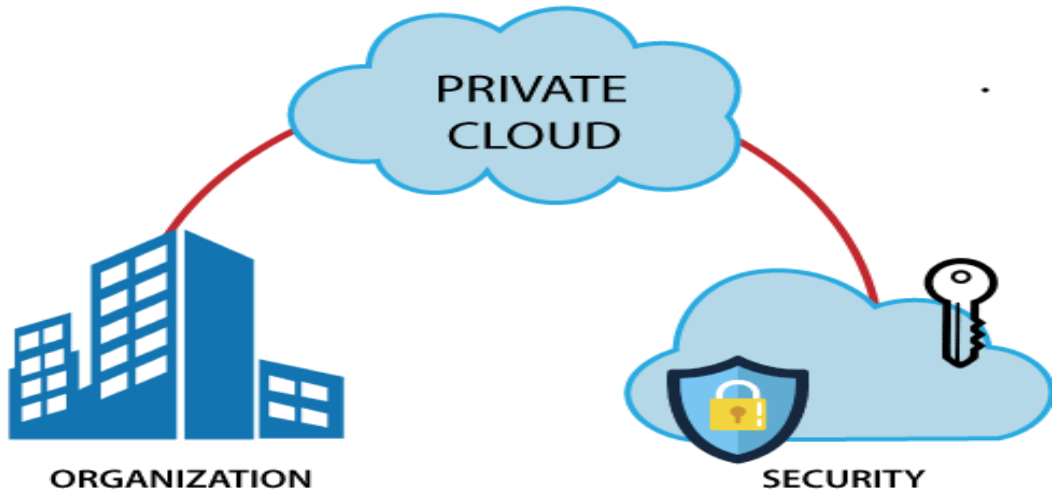
- **Low Cost:** Public cloud has a lower cost than private, or hybrid cloud, as it shares the same resources with a large number of consumers.
- **Location Independent:** Public cloud is location independent because its services are offered through the internet.
- **Save Time:** In Public cloud, the cloud service provider is responsible for the manage and maintain data centers in which data is stored, so the cloud user can save their time to establish connectivity, deploying new products, release product updates, configure, and assemble servers.
- **Quickly and easily set up:** Organizations can easily buy public cloud on the internet and deployed and configured it remotely through the cloud service provider within a few hours.
- **Business Agility:** Public cloud provides an ability to elastically re-size computer resources based on the organization's requirements.
- **Scalability and reliability:** Public cloud offers scalable (easy to add and remove) and reliable (24*7 available) services to the users at an affordable cost.

Disadvantages of Public Cloud

- **Low Security:** Public Cloud is less secure because resources are shared publicly.
- **Performance:** In the public cloud, performance depends upon the speed of internet connectivity.
- **Less customizable:** Public cloud is less customizable than the private cloud.

Private Cloud

- Private cloud is also known as an internal cloud or corporate cloud.
- Private cloud provides computing services to a private internal network (within the organization) and selected users instead of the general public.
- Private cloud provides a high level of security and privacy to data through firewalls and internal hosting. It also ensures that operational and sensitive data are not accessible to third-party providers.
- HP Data Centers, Microsoft, Elastra-private cloud, and Ubuntu are the example of a private cloud.



Advantages of Private cloud

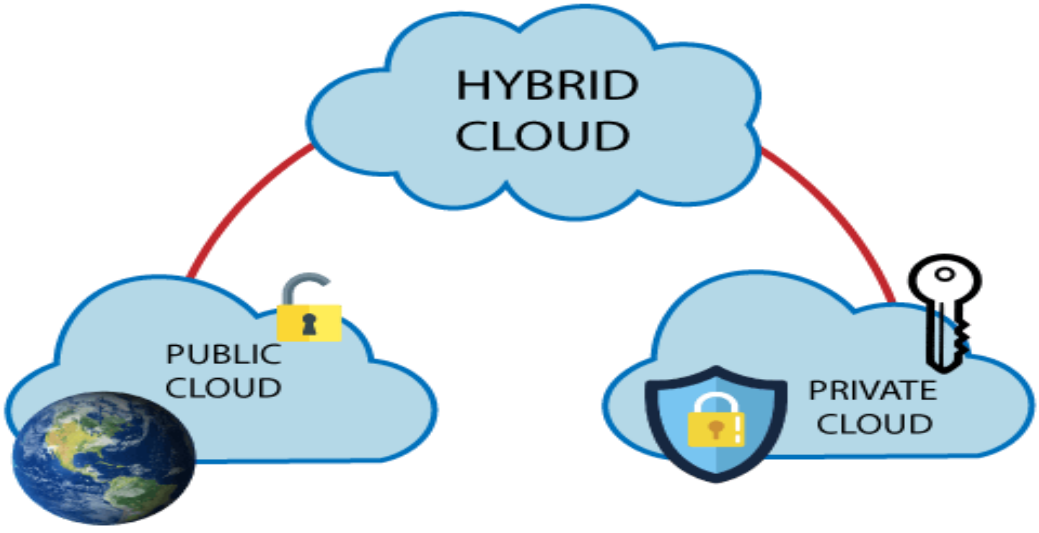
- **More Control:** Private clouds have more control over their resources and hardware than public clouds because it is only accessed by selected users.
- **Security & privacy:** Security & privacy are one of the big advantages of cloud computing. Private cloud improved the security level as compared to the public cloud.
- **Improved performance:** Private cloud offers better performance with improved speed and space capacity.

Disadvantages of Private cloud

- **High cost:** The cost is higher than a public cloud because set up and maintain hardware resources are costly.
- **Restricted area of operations:** As we know, private cloud is accessible within the organization, so the area of operations is limited.
- **Limited scalability:** Private clouds are scaled only within the capacity of internal hosted resources.
- **Skilled people:** Skilled people are required to manage and operate cloud services.

Hybrid Cloud

- Hybrid cloud is a combination of public and private clouds.
  Hybrid cloud = public cloud + private cloud
- The main aim to combine these cloud (Public and Private) is to create a unified, automated, and well-managed computing environment.
- In the Hybrid cloud, non-critical activities are performed by the public cloud and critical activities are performed by the private cloud.
- Mainly, a hybrid cloud is used in finance, healthcare, and Universities.
- The best hybrid cloud provider companies are Amazon, Microsoft, Google, Cisco, and NetApp.



Advantages of Hybrid Cloud

- **Flexible and secure:** It provides flexible resources because of the public cloud and secure resources because of the private cloud.

- **Cost effective:** Hybrid cloud costs less than the private cloud. It helps organizations to save costs for both infrastructure and application support.
- **Cost effective:** It offers the features of both the public as well as the private cloud. A hybrid cloud is capable of adapting to the demands that each company needs for space, memory, and system.
- **Security:** Hybrid cloud is secure because critical activities are performed by the private cloud.
- **Risk Management:** Hybrid cloud provides an excellent way for companies to manage the risk.

Disadvantages of Hybrid Cloud

- **Networking issues:** In the Hybrid Cloud, networking becomes complex because of the private and the public cloud.
- **Infrastructure Compatibility:** Infrastructure compatibility is the major issue in a hybrid cloud. With dual-levels of infrastructure, a private cloud controls the company, and a public cloud does not, so there is a possibility that they are running in separate stacks.
- **Reliability:** The reliability of the services depends on cloud service providers.

Community Cloud

- Community cloud is a cloud infrastructure that allows systems and services to be accessible by a group of several organizations to share the information. It is owned, managed, and operated by one or more organizations in the community, a third party, or a combination of them.
- Example: Our government organization within India may share computing infrastructure in the cloud to manage data.

Advantages of Community Cloud

- **Cost effective:** Community cloud is cost effective because the whole cloud is shared between several organizations or a community.
- **Flexible and Scalable:** The community cloud is flexible and scalable because it is compatible with every user. It allows the users to modify the documents as per their needs and requirement.

- **Security:** Community cloud is more secure than the public cloud but less secure than the private cloud.
- **Sharing infrastructure:** Community cloud allows us to share cloud resources, infrastructure, and other capabilities among various organizations.

Disadvantages of Community Cloud

- Community cloud is not a good choice for every organization.
- Slow adoption to data
- The fixed amount of data storage and bandwidth is shared among all community members.
- Community Cloud is costly than the public cloud.
- Sharing responsibilities among organizations is difficult.

Difference between public cloud, private cloud, hybrid cloud, and community cloud

Parameter	Public Cloud	Private Cloud	Hybrid Cloud	Community Cloud
Host	Service provider	Enterprise (Third party)	Enterprise (Third party)	Community (Third party)
Users	General public	Selected users	Selected users	Community members
Access	Internet	Internet, VPN	Internet, VPN	Internet, VPN
Owner	Service provider	Enterprise	Enterprise	Community