

Cloud Computing

It is the online delivery of computing services including server, storage, Databases, networking, softwares, analytic and intelligence over the internet (cloud) to offer faster innovative, flexible resource and economies of scale

Cloud computing is a pay per use on demand model or pay as you go model

Example

Amazon web service (AWS), Microsoft Azure, IBM cloud, VM ware, google cloud platform

Why we go for cloud

Huge amount needed for old model

Experts is needed for old model

Scalability is essential

Security

Instant deployment (huge time for deployment for old model)

Characteristics or advantages of Cloud Computing

Scalability

High availability and reliability

Multi-sharing

Device and location independence

Maintenance is easy

Low cost, security

Types of Cloud Computing

Three types of cloud computing . They are Public Cloud, Private Cloud and Hybrid Cloud

- **Public Cloud:** The cloud resources that are owned and operated by a third-party cloud service provider are termed as public clouds. It delivers computing resources such as servers, software, and storage over the internet
- **Private Cloud:** The cloud computing resources that are exclusively used inside a single business or organization are termed as a private cloud. A private cloud may physically be located on the company's on-site datacentre or hosted by a third-party service provider.
- **Hybrid Cloud:** It is the combination of public and private clouds, which is bounded together by technology that allows data applications to be shared between them. Hybrid cloud provides flexibility and more deployment options to the business.

Types of Cloud Services

Infrastructure as a Service (IaaS): In IaaS, we can rent IT infrastructures like servers and virtual machines (VMs), storage, networks, operating systems from a cloud service vendor

Platform as a Service (PaaS): This service provides an on-demand environment for developing, testing, delivering, and managing software applications. The developer is responsible for the application, and the PaaS vendor provides the ability to deploy and run it.

Software as a Service (SaaS): It provides a centrally hosted and managed software services to the end-users. It delivers software over the internet, on-demand, and typically on a subscription basis. E.g., Microsoft One Drive, Dropbox, WordPress, Office 365

What is Big Data

Data which are very large in size is called Big Data. Normally we work on data of size MB(WordDoc ,Excel) or maximum GB(Movies, Codes) but data

in Peta bytes i.e. 10^{15} byte size is called Big Data. It is stated that almost 90% of today's data has been generated in the past 3 years.

Big data refers to collection of data that are huge and complex that none of that none of the traditional data management tools are able to store and process efficiently.

Characteristics of big data

- Mostly unstructured
- Store in PB, Eb
- Increases exponentially
- Globally present, distributed

Sources of Big Data

These data come from many sources like

- **Social networking sites:** Facebook, Google, LinkedIn all these sites generate huge amount of data on a day to day basis as they have billions of users worldwide.
- **E-commerce site:** Sites like Amazon, Flipkart, Alibaba generate huge amount of logs from which users buying trends can be traced.
- **Weather Station:** All the weather station and satellite give very huge data which are stored and manipulated to forecast weather.
- **Telecom company:** Telecom giants like Airtel, Vodafone study the user trends and accordingly publish their plans and for this they store the data of its million users.
- **Share Market:** Stock exchange across the world generates huge amount of data through its daily transaction.
- 3V's of Big Data
 1. **Velocity:** The data is increasing at a very fast rate. It is estimated that the volume of data will double in every 2 years.
 2. **Variety:** Now a days data are not stored in rows and column. Data is structured as well as unstructured. Log file, CCTV footage is unstructured data. Data which can be saved in tables are structured data like the transaction data of the bank.

3. **Volume:** The amount of data which we deal with is of very large size of Peta bytes.

Data mining

The process of extracting information to identify patterns, trends, and useful data that would allow the business to take the data-driven decision from huge sets of data is called Data Mining.

In other words, we can say that Data Mining is the process of investigating hidden patterns of information to various perspectives for categorization into useful data, which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cost-cutting and generating revenue.

Data mining is also called ***Knowledge Discovery in Database (KDD)***.

Advantages of Data Mining

- The Data Mining technique enables organizations to obtain knowledge-based data.
- Data mining enables organizations to make lucrative modifications in operation and production.
- Compared with other statistical data applications, data mining is a cost-efficient.
- 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.

Disadvantages of Data Mining

- There is a probability that the organizations may sell useful data of customers to other organizations for money. As per the report, American Express has sold credit card purchases of their customers to other organizations.
- 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.
- The data mining techniques are not precise, so that it may lead to severe consequences in certain conditions.

These are the following areas where data mining is widely used

- Data Mining in Healthcare
- Data Mining in Market Basket Analysis
- Data mining in Education
- Data Mining in CRM (Customer Relationship Management)
- Data Mining in Fraud detection

Data Mining Techniques

Clustering

Clustering is a technique used to represent data visually — such as in graphs that show buying trends or sales demographics for a particular product.

Association

Association rules are used to find correlations, or associations, between points in a data set.

Data Cleaning

Data cleaning involves organizing data, eliminating duplicate or corrupted data, and filling in any null values. When this process is complete, the most useful information can be harvested for analysis.

Data Visualization

Data visualization is the translation of data into graphic form to illustrate its meaning to business stakeholders.

Prediction:

Prediction used a combination of other data mining techniques such as trends, clustering, classification, etc. It analyzes past events or instances in the right sequence to predict a future event.

Mobile computing

Mobile Computing refers a technology that allows transmission of data, voice and video via a computer or any other wireless enabled device. It is free from having a connection with a fixed physical link. It facilitates the users to move from one physical location to another during communication.

Applications of Mobile Computing

Following is a list of some significant fields in which mobile computing is generally applied:

- Web or Internet access.
- Global Position System (GPS).
- Emergency services.
- Entertainment services.
- Educational services.

There are some wireless/mobile computing technologies given below:

Global System for Mobile Communications (GSM) :

GSM is a Current circuit-switched wireless data communication technology.

Code-Division Multiple Access (CDMA) :

CDMA is a type of wireless computing technology. It is developed during World War II. This technology is mostly used as it provides better network quality, more storage capacity for voice and data communications

General Packet Radio Service (GPRS) :

GPRS is a type of Packet-based Wireless communication technology.

Embedded Systems.

An **Embedded System** is a system that has software embedded into computer-hardware, which makes a system dedicated for a variety of application or specific part of an application or product or part of a larger system.

An embedded system can be a small independent system or a large combinational system. It is a microcontroller-based control system used to perform a specific task of operation.

Eg. industrial machines, consumer electronics, agricultural and processing industry devices, automobiles, medical equipment, cameras, digital watches, household appliances, airplanes, vending machines and toys, as well as mobile devices, are possible locations for an embedded system.

An embedded system is a combination of three major components:

- **Hardware:** Hardware is physically used component that is physically connected with an embedded system. It comprises of microcontroller based integrated circuit, power supply, LCD display etc.
- **Application software:** Application software allows the user to perform varieties of application to be run on an embedded system by changing the code installed in an embedded system.
- **Real Time Operating system (RTOS):** RTOS supervises the way an embedded system work. It act as an interface between hardware and application software which supervises the application software and provide mechanism to let the processor run on the basis of scheduling for controlling the effect of latencies.

Characteristics of embedded system

- **Single-functioned**
- **Microprocessors based**
- Requires real time performance
- It should have high availability and reliability.
- Developed around a real-time operating system

- Usually, have easy and a diskless operation, ROM boot
- Designed for one specific task
- It must be connected with peripherals to connect input and output devices.
- Offers high reliability and stability
- Needed minimal user interface

Use of Computers in Education and Research:

E-Library

An e-library or [Digital library](#) is a physical site and/ or website that provide around the clock online access to [digitized](#) audio, video, and [written material](#). It provides free copies of books, [journals](#), etc. available to the users.

An e-library is electronic information, library organizers usage of the catalog, tag and search books, and journals. It maintains a database as the collection of e-materials and provides services in digital form.

Advantages of E-Library

1. E-library is the easiest to use the available online research tool.
2. Standards searching help the educators to integrate the technology into the curriculum, by increasing the technology literacy.
3. Students can use the search by topic feature to retrieve a manageable amount of quality content, quickly and easily.
4. Public libraries need to offer an easy-to-use research solution to patrons.
5. Point-and-click functionality ensures all the users finding the information they need.
6. The reference desk gives integrated access to a dictionary, encyclopedia, almanacs, and much more.
7. Visually impaired people are no longer disabled in searching and surfing information on the digital library.

Google scholar

Google Scholar is a freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines. Released in beta in November 2004, the Google Scholar index includes peer-reviewed online academic journals and books,

conference papers, theses and dissertations, preprints, abstracts, technical reports, and other scholarly literature, including court opinions and patents.

Google Scholar allows users to search for digital or physical copies of articles, whether online or in libraries. It indexes "full-text journal articles, technical reports, preprints, theses, books, and other documents, including selected Web pages that are deemed to be 'scholarly.' Because many of Google Scholar's search results link to commercial journal articles, most people will be able to access only an abstract and the citation details of an article, and have to pay a fee to access the entire article.^[20] The most relevant results for the searched keywords will be listed first, in order of the author's ranking, the number of references that are linked to it and their relevance to other scholarly literature, and the ranking of the publication that the journal appears in.

What is MATLAB?

MATLAB is a software package for high-performance mathematical computation, visualization, and programming environment. It provides an interactive environment with hundreds of built-in functions for technical computing, graphics, and animations.

MATLAB stands for **Matrix Laboratory**. MATLAB was written initially to implement a simple approach to matrix software developed by the **LINPACK** (Linear system package) and **EISPACK** (Eigen system package) projects.

MATLAB is a modern programming language environment, and it has refined data structures, includes built-in editing and debugging tools, and supports object-oriented programming.

MATLAB is **Multi-paradigm**. So, it can work with multiple types of programming approaches, such as Functional, Object-Oriented, and Visual.

MATLAB

MATLAB
Programming
Language

User-written Functions

Built-in Functions

Graphics

- 2-D Graphics
- 3-D Graphics
- Color and Lighting
- Animation
- Audio and Video

E
X
T
R
A
\$
F
U
N
C
T
I
O

Computations

- Linear Algebra
- Data Analysis
- Signal Processing
- Polynomials & Interpolation
- Quadrature
- Solution of ODEs

MATLAB Features

As there are numerous features to describe, but here, we will focus on some of the key features:

- It is designed for numerical as well as symbolic computing.
- It's a high-level language used mainly for engineering and scientific computing.
- It works within a Desktop environment providing full features for iterative exploration, design, and problem-solving.
- Creation of custom plots for visualizing data and tools, with the help of built-in Graphics.
- Specific applications are designed to work with any particular type of problems, such as data classification, control system design and tuning, signal analysis.
- Provides several add-on toolboxes to build a wide range of engineering, scientific, and custom user interface applications.
- Provide interfaces to work with other programming languages such as C, C++, Java, .NET, Python, SQL, Hadoop.

Mathematica

Mathematica is a mathematical computation program used in many scientific, engineering, mathematical, and computing fields. Unlike other systems, Mathematica applies intelligent automation in every part of the system, from algorithm selection to plot layout and user interface design.

Heterogeneous Storage

Heterogeneous data are any data with high variability of data types and formats. They are possibly ambiguous and low quality due to missing values, high data redundancy, and untruthfulness. It is difficult to integrate heterogeneous data to meet the business information demands. For example, heterogeneous data are often generated from Internet of Things (IoT). Because of the variety of data acquisition devices, the acquired data are also different in types with heterogeneity.

- Syntactic heterogeneity occurs when two data sources are not expressed in the same language.
- Conceptual heterogeneity, also known as semantic heterogeneity or logical mismatch, denotes the differences in modelling the same domain of interest.
- Terminological heterogeneity stands for variations in names when referring to the same entities from different data sources.
- Semiotic heterogeneity, also known as pragmatic heterogeneity, stands for different interpretation of entities by people.

Domain specific languages

A domain-specific language (DSL) is a computer language specialized to a particular application domain. This is in contrast to a general-purpose language (GPL), which is broadly applicable across domains. A domain-specific language is created specifically to solve problems in a particular domain and is not intended to be able to solve problems outside of it (although that may be technically possible). In contrast, general-purpose languages are created to solve problems in many domains. Examples of domain-specific languages include HTML, Logo for pencil-like drawing, Verilog and VHDL hardware description languages, MATLAB and GNU Octave for matrix programming, Mathematica, Maple and Maxima for symbolic mathematics, Specification and Description Language for reactive and distributed systems,