IT 231
Foundation of Information Technology

Unit 10

Introduction to Contemporary Technologies

Data Warehousing

Data warehouse is defined as a process of centralized data management and retrieval. It is a storage of historical data.

Data warehousing represents an ideal vison of maintaining a central repository of all organizational data.

Data warehouse is defined as a collection of huge amounts of data which are stored on the basis of some specific subject, data are time independent, non-volatile, and are integrated.

Data Warehousing

Subject oriented: data stored in the data warehouse are about some specific subject like customer, time etc.

Non volatile: data in the data warehouse are never updated or deleted.

Integrated: data from different sources are incorporated.

Time-independent: connection between data and the time when it was entered.

Applications of Data Warehouse

- Banking Industry
- Government
- Consumer Goods Industry
- Education
- Healthcare
- Manufacturing and Distribution Industry.

Data Mining

- Extracting or mining knowledge from large amounts of data. Remember that the mining of gold from rocks or sand is referred to as gold mining.
- Process used by organizations to extract specific data from huge databases to solve business problems. It primarily turns raw data into useful information.
- using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing.

How does data mining works?

Different analytical software are used for data mining: statistical, machine learning, neural network etc.

Classes: used to classify each item in a set of data into one of a predefined set of classes or groups.

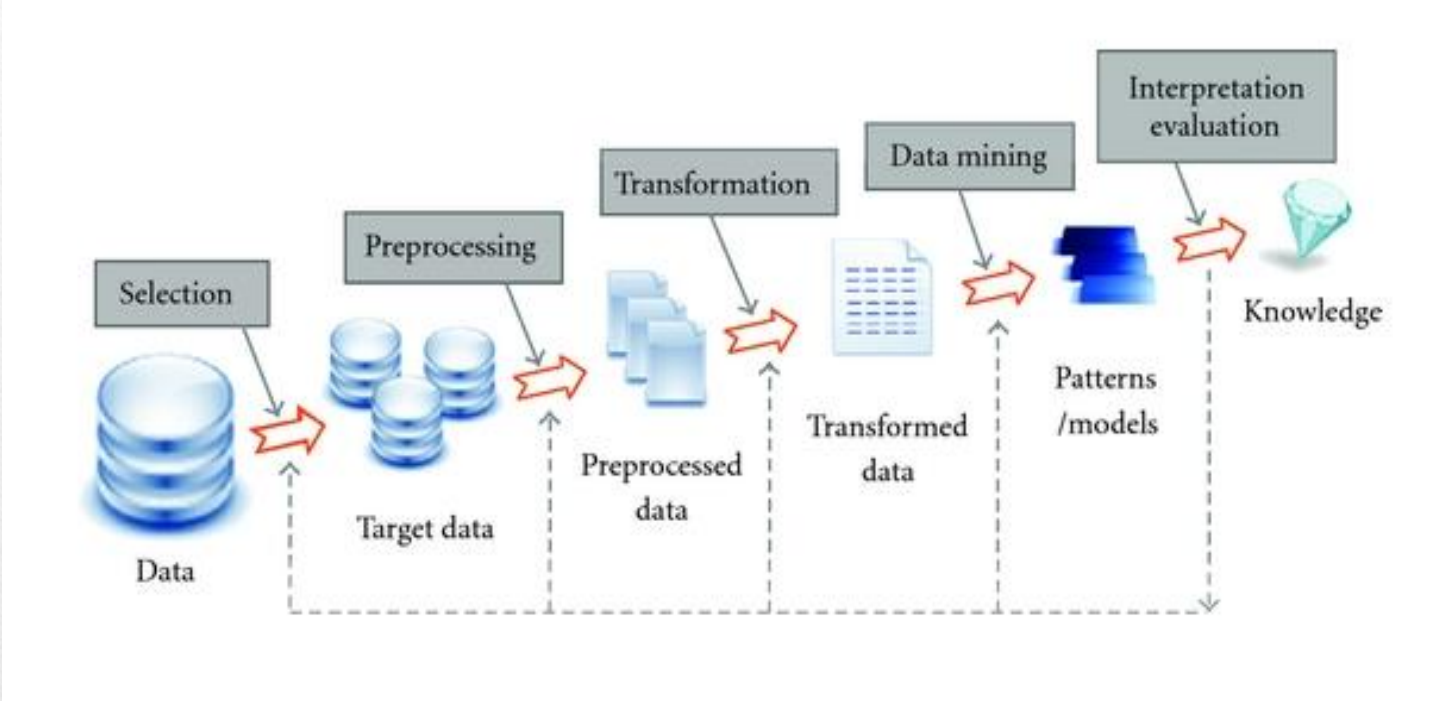
Clusters: It is a technique of partitioning a set of data into clusters or groups of objects. The clustering is done using algorithms.

Cont..

Association: In association, a pattern is discovered based on a relationship between items in the same transaction.

Sequential Patterns: A trend or some consistent patterns are recognized in this type of data mining. Understanding customer purchase behavior and sequential patterns are used by the stores to display their products on shelves.

The Knowledge Discovery in Databases (KDD) Process



The Knowledge Discovery in Databases (KDD) Process

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

Big Data

- Big data is data that contains greater variety, arriving in increasing volumes and with more velocity.
- Big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before.

The Three Vs of Big Data

- **Volume:** With big data, you'll have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a web page or a mobile app, etc.
- **Velocity :** The fast rate at which data is received and (perhaps) acted on.
- **Variety:** Variety refers to the many types of data that are available.

Big Data Fact

- We created 2.5 quintillion data bytes daily in 2020. (Source: Domo)
- 463 Exabyte's of data will be generated each day by people as of 2025. (*Source: Raconteur)*
- There were 4.66 billion active internet users around the world in January 2021.*(Source: Statista)*
- The end of 2021 could see two trillion Google searches. *(Source: ILS)*.
- Five hundred thousand new Tweets were posted every day in 2020.*(Source: David Syce)*

Big Data Fact

- Facebook generated four petabytes of data every day in 2020.*(Source: Kinsta)*
- In 2024, the number of emails will be about 361 billion every day.*(Source: Statista)*

Data growth statistics show that by 2022, over 70% of GDPs worldwide will have gone through some form of digitalization. The demand for collaboration software, cloud solutions, and contactless services will go up as Governments and organizations strive to move away from analog systems.

Big Data Technologies

- NoSQL: non-tabular databases and store data differently than relational tables.
- Hadoop : Hadoop is an open source software that stores and processes large volumes of data for analytical and batch operation purposes.
- Map Reduce: Map Reduce is a Hadoop framework used for writing applications that can process vast amounts of data on large clusters.

Artificial Intelligence

Meaning of the Word: ***"Intelligence"***

- Ability to reason.
- Ability to understand.
- Ability to create.
- Ability to learn from experience.
- Ability to plan and execute complex task.

What behaviors are Intelligent?

- **Everyday Tasks:** recognize a friend, recognize who is calling, translate from one language to another, interpret a photograph, talk, cook a dinner, etc.
- **Formal Tasks:** prove a logic Theorem, Geometry, calculus, play chess, checkers, etc.
- **Expert Tasks:** engineering design, medical designers, financial analysis, etc.

What is AI?

Artificial intelligence is about the science and engineering necessary to create artifacts that can

- acquire knowledge, i.e., can learn and extract knowledge; and
- reason with knowledge (leading to doing tasks such as planning, explaining, diagnosing, acting rationally, etc.),

Artificial Intelligence

- **Barr and Feigenbaum:**"Artificial Intelligence is the part of computer science concerned with designing intelligence computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior

- **Elaine Rich:** AI is the study of how to make computers do things at which, at the moment, people are better"

Different definitions of AI are given by different books/writers. These definitions can be divided into two dimensions.

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

Top dimension is concerned with thought processes and reasoning, where as bottom dimension addresses the behavior.

The definition on the left measures the success in terms of fidelity of human performance, whereas definitions on the right measure an ideal concept of intelligence, which is called rationality.

Human-centered approaches must be an empirical science, involving hypothesis and experimental confirmation. A rationalist approach involves a combination of mathematics and engineering.

Applications of AI

- Game Playing
- Speech Recognition
- Understanding Natural Language
- Computer Vision
- Expert System

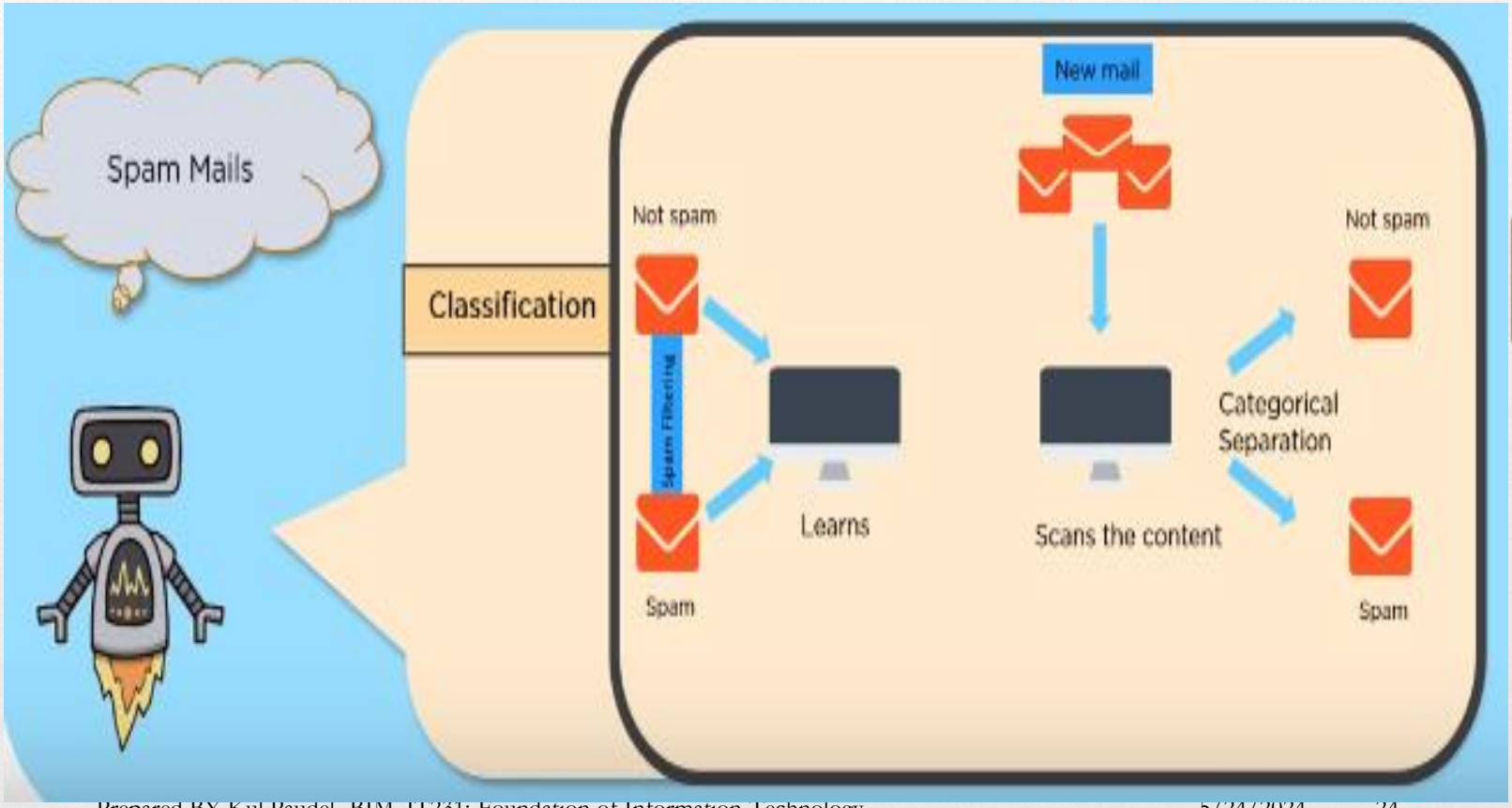
Machine Learning

- Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.
- Similar to how the human brain gains knowledge and understanding, machine learning relies on input, such as training data or knowledge graphs, to understand entities, domains and the connections between them. With entities defined, deep learning can begin.

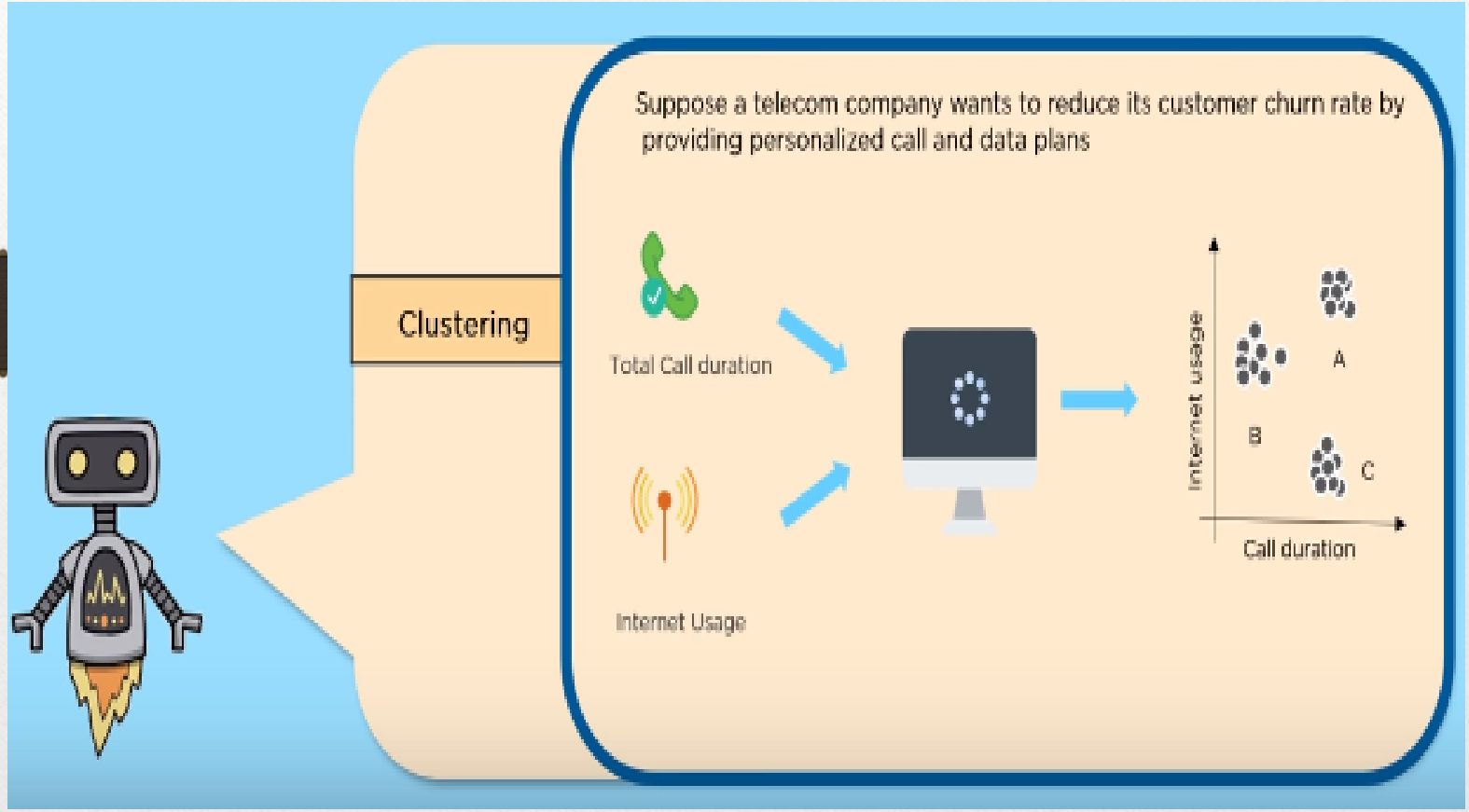
Methods of Machine Learning:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

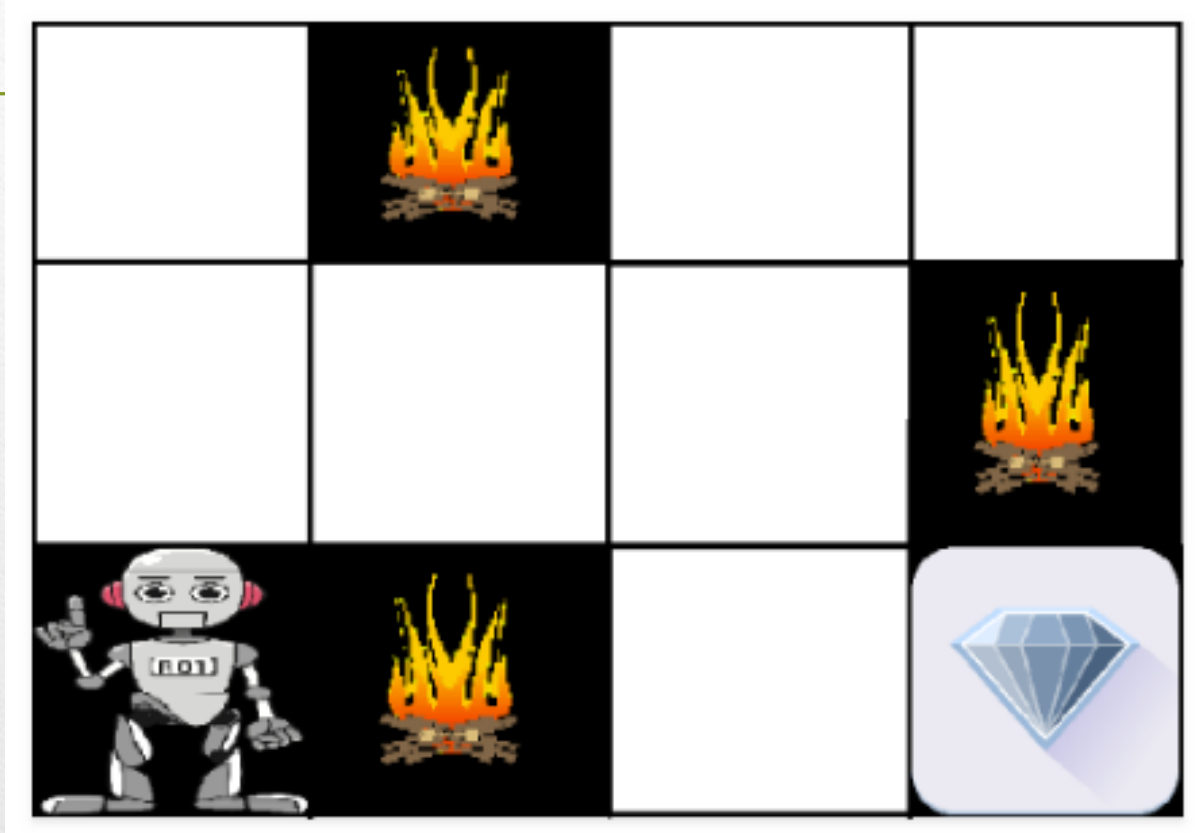
Supervised Learning



Unsupervised Learning



Reinforcement



Artificial Neural Network

Neural nets are basically mathematical models of information processing.

The discipline of ANN arose from the thought of mimicking the functioning of the same human brain that was trying to solve the problem.

AI cannot imitate human intelligence completely; computers can only be programmed to do some aspects of the human brain.

Cloud Computing

Cloud

The "Cloud" is the default symbol of the Internet in diagrams.

Computing

The broader term of "Computing" encompasses:
- Computation
- Coordination logic
- Storage

Cloud Computing is about moving computing from the single desktop pc/data centers to commercial service providers on the Internet.

Benefits of Cloud Computing

Self-Service Provisioning

- End users can spin up computing resources for almost any type of workload on-demand

Elasticity

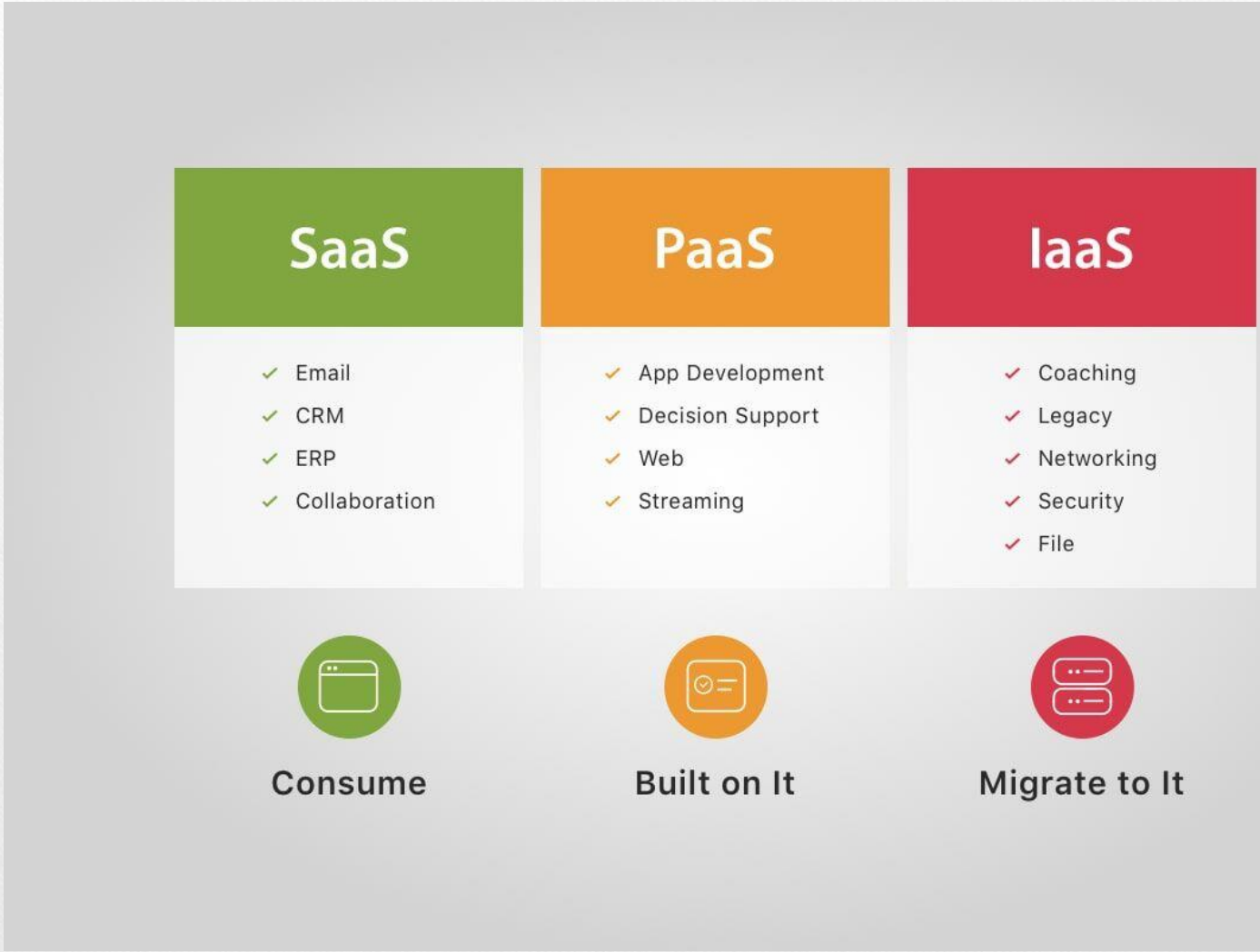
- Companies can scale up as computing needs increase and then scale down again as demand decreases.

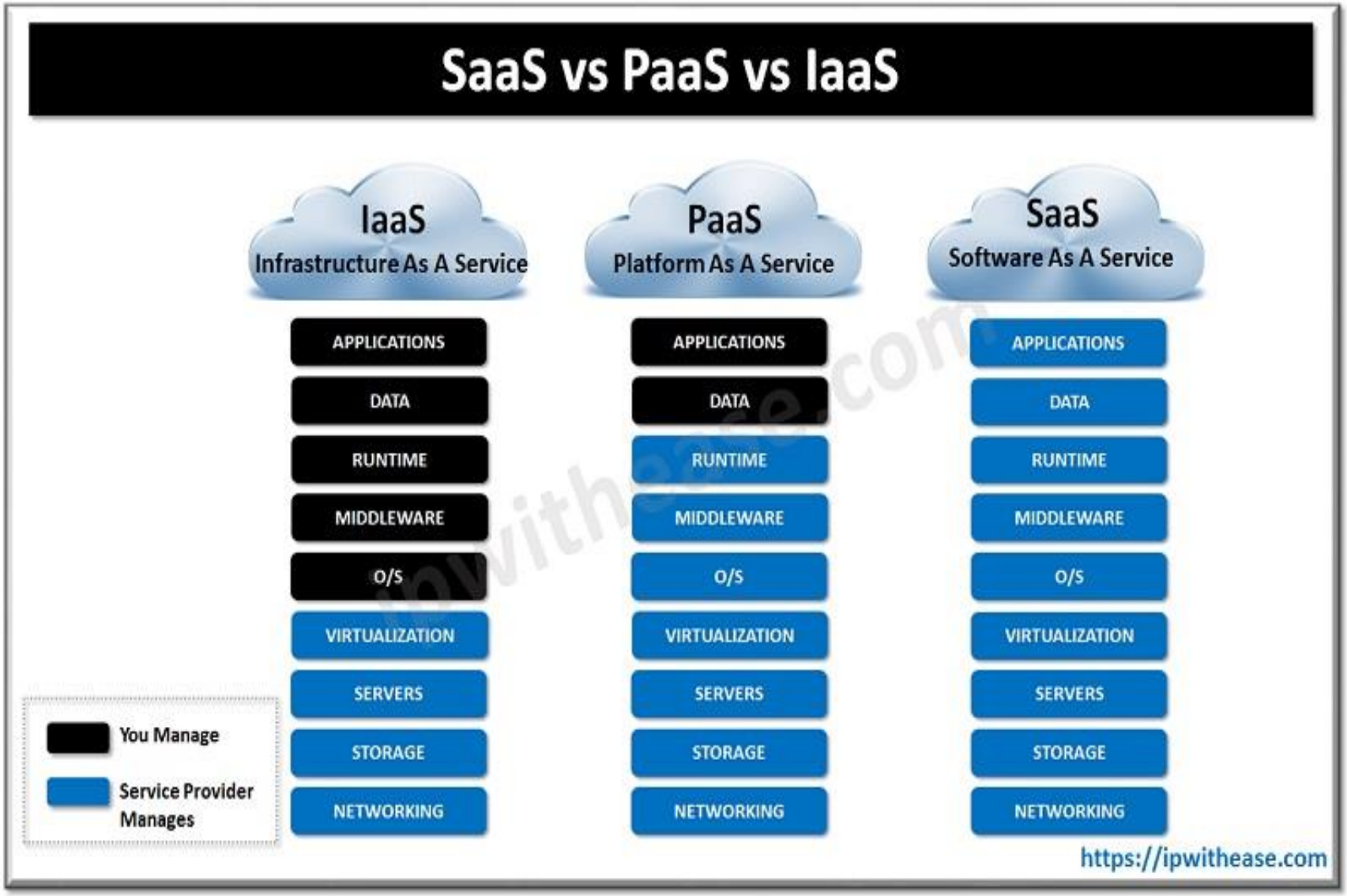
Pay per Use

- Computing resources are measured at a granular level, allowing users to par only for the resources and workloads they use.

Models of Cloud Computing

- **Infrastructure-as-a-service(IaaS)**
- **Platform-as-a-service(PaaS)**
- **Software-as-a-service(SaaS)**


SaaS
Email
CRM
ERP
Collaboration
PaaS
App Development
Decision Support
Web
Streaming
IaaS
Coaching
Legacy
Networking
Security
File
Consume
Built on It
Migrate to It


SaaS vs PaaS vs IaaS
IaaS
Infrastructure As A Service
PaaS
Platform As A Service
SaaS
Software As A Service
APPLICATIONS
DATA
RUNTIME
MIDDLEWARE
O/S
VIRTUALIZATION
SERVERS
STORAGE
NETWORKING
APPLICATIONS
DATA
RUNTIME
MIDDLEWARE
O/S
VIRTUALIZATION
SERVERS
STORAGE
NETWORKING
APPLICATIONS
DATA
RUNTIME
MIDDLEWARE
O/S
VIRTUALIZATION
SERVERS
STORAGE
NETWORKING
You Manage
Service Provider Manages
https://ipwithease.com

Green Computing

- Green computing, also known as green technology, is the use of computers and other computing devices and equipment in energy-efficient and eco-friendly ways. Organizations that use green computing methods often deploy energy-efficient central processing units (CPUs), servers, peripherals and power systems.

Green Computing Strategy

- Remote work.
- Smart technology.
- Upgrade and rearrange the data center.
- Power down.
- Strategic scheduling.
- Computer selection.
- Temperature check.
- E-waste.

Virtual Computing

Virtual computing refers to the use of a remote computer from a local computer where the actual computer user is located.

For example, a user at a home computer could log in to a remote office computer (via the Internet or a network) to perform job tasks. Once logged in via specialized software, the remote computer user can perform tasks via the keyboard, mouse, or other tools.

Digital Marketing

- Digital marketing is any marketing that uses the Internet and electronic devices that can be used by marketing specialists to convey promotional messaging and measure its impact through your customer journey.
- Digital marketing typically refers to marketing campaigns that appear on a computer, phone, tablet, or other device.
- Digital marketing, also called online marketing, is the promotion of brands to connect with potential customers using the internet and other forms of digital communication.
- Digital marketing includes not only email, social media, and web-based advertising, but also text and multimedia messages as a marketing channel.

Types of Digital Marketing

- SEO
- PPC
- Social Media Marketing
- Content Marketing
- Email Marketing
- Mobile Marketing
- Affiliate Marketing/Influencer Marketing

Internet of Things

- In general, Internet of Things, or IoT, refers to the physical devices that are connected to the internet.
- The IoT is a giant network of connected things and people – all of which collect and share data about the way they are used and about the environment around them.
- from smart microwaves, which automatically cook your food for the right length of time, to self-driving cars, whose complex sensors detect objects in their path,

Features of IoT

- Connectivity : connection of device to the network
- Scaling: connection cab be scaled up or down according to need.
- Artificial Intelligence: gain more advantages when IoT can combine with AI.
- Automated: automated system implementation in devices.
- Security: Store and save sensitive information regarding IoT operable devices.

Remote Sensing & GIS

Any location on the earth has massive amounts of data tied to it, which not only include physical features, but political, economic and social data, as well. GIS facilitates the process by which we can visualize, analyze and understand this data.

A geographic information system (GIS) is a computer-based tool for mapping and analyzing feature events on earth.

Remote sensing is one of the methods commonly used for collecting physical data to be integrated into GIS. Remote sensors collect data from objects on the earth without any direct contact.

Business Intelligence

- Business Intelligence represents the technical infrastructure that collects, stores, and analyzes company data that drives to profitable business actions.
- BI parses data and produces reports and information that help managers to make better decisions.
- BI tools and software come in a wide variety of forms such as spreadsheets, reporting/query software, data visualization software, data mining tools, and online analytical processing (OLAP).

Types of BI Tools and Software

- **Spreadsheets:** Spreadsheets like Microsoft Excel and Google Docs are some of the most widely used BI tools.
- **Reporting software:** Reporting software is used to report, organize, filter, and display data.
- **Data visualization software:** Data visualization software translates datasets into easy-to-read, visually appealing graphical representations to quickly gain insights.

Types of BI Tools and Software

- **Data mining tools:** Data mining tools "mine" large amounts of data for patterns using things like artificial intelligence, machine learning, and statistics.
- **Online analytical processing (OLAP):** OLAP tools allow users to analyze datasets from a wide variety of angles based on different business perspectives.

Benefits of BI

- faster, more accurate reporting and analysis
- improved data quality
- better employee satisfaction
- reduced costs and increased revenues
- the ability to make better business decisions.