Introduction to Contemporary Technologies

Avinash Maskey

Data Warehousing and Data Mining

Data Warehouse

- A ***Data Warehouse*** is a collection of software tools that facilitates analysis of a large set of business data used to help an organization make decisions.
- A large amount of data in data warehouses comes from numerous sources such that internal applications like marketing, sales, and finance; customer-facing apps; and external partner systems, among others. It is a centralized data repository for analysts that can be queried whenever required for business benefits.
- A data warehouse is mainly a data management system that's designed to enable and support business intelligence (BI) activities, particularly analytics. Data warehouses are alleged to perform queries, cleaning, manipulating, transforming and analyzing the data and they also contain large amounts of historical data.
- ***Data warehouses*** store historical data and handle requests faster, helping in online analytical processing, whereas a ***database*** is used to store current transactions in a business process that is called online transaction processing.

Features of Data Warehouses

- **Subject Oriented:**

  It provides you with important data about a specific subject like suppliers, products, promotion, customers, etc. Data warehousing usually handles the analysis and modeling of data that assist any organization to make data-driven decisions.

- **Integrated:**

  Different heterogeneous sources are put together to build a data warehouse, such as level documents or social databases.

- **Time-Variant:**

  The data collected in a data warehouse is identified with a specific period.

- **Nonvolatile:**

  This means the earlier data is not deleted when new data is added to the data warehouse. The operational database and data warehouse are kept separate and thus continuous changes in the operational database are not shown in the data warehouse.

Data Mining

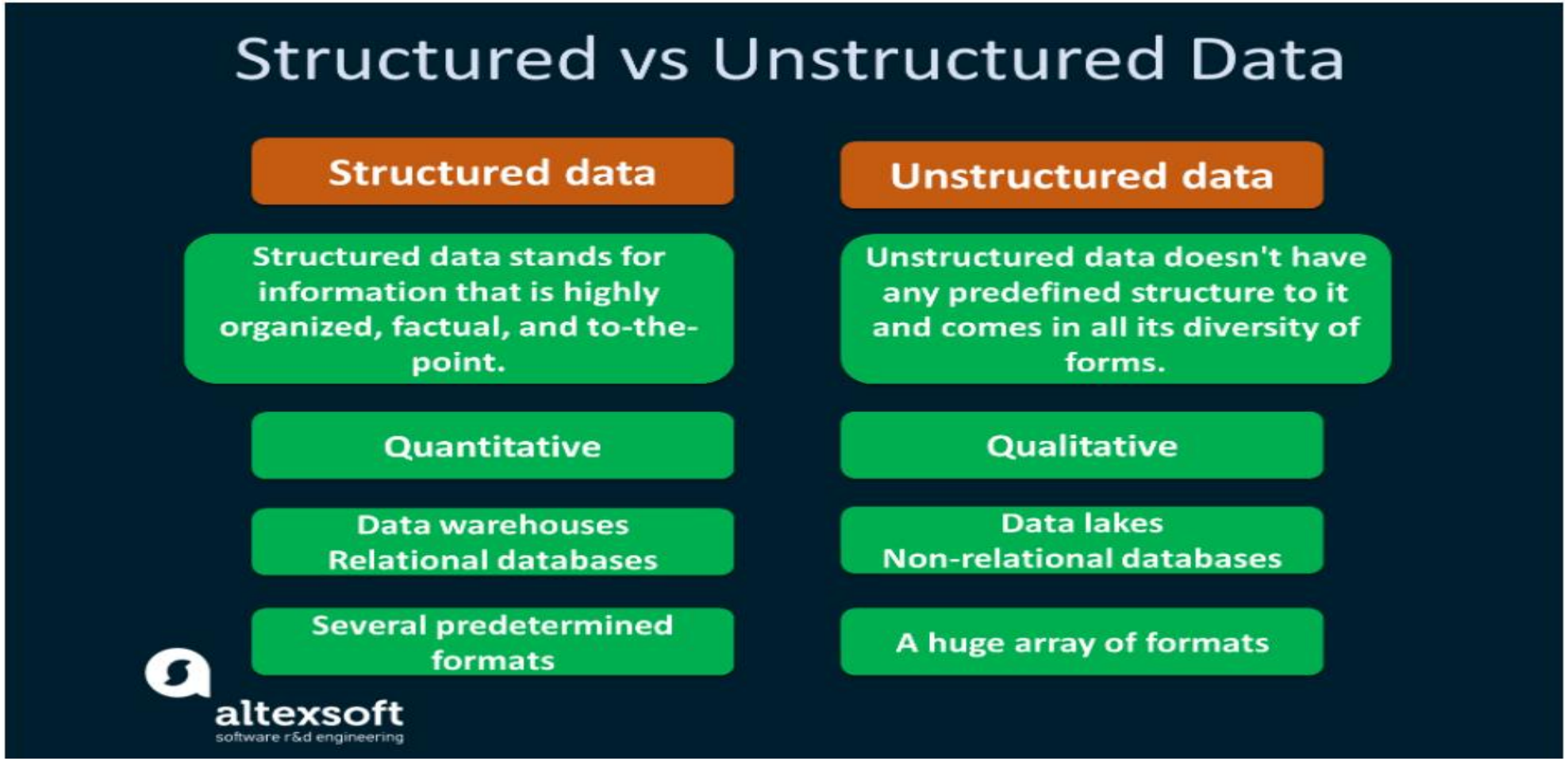
- ***Data mining*** aims to enable business organizations to view business behaviors, trends relationships that allow the business to make data-driven decisions.
- It is also known as **K**nowledge **D**iscovery in **D**atabase (**KDD**).
- ***Data mining*** uses statistics, artificial intelligence, machine learning systems, and some databases to find hidden patterns in the data. It supports business-related queries that are time-consuming to resolve.
- In this process, data is extracted and analyzed to fetch useful information. In data mining hidden patterns are researched from the dataset to predict future behavior. Data mining is used to indicate and discover relationships through the data.

Features of Data Mining

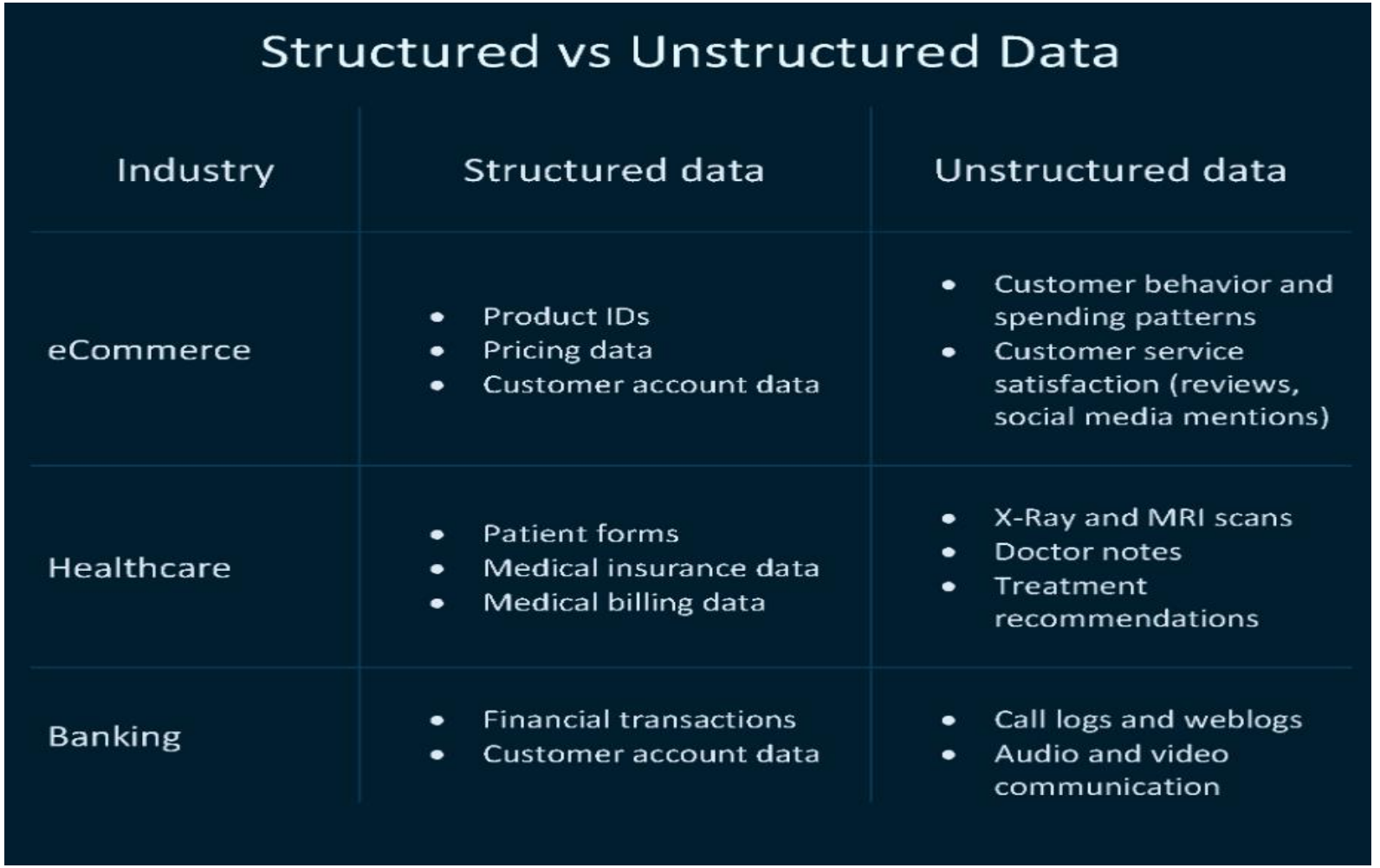
- It focuses on large data sets and databases.
- It predicts future results.
- It creates actionable insights.
- It utilizes the automated discovery of patterns.

Big Data and Data Science

Structured Data Vs Unstructured Data



Structured Data Vs Unstructured Data



Big Data

- ***Big data*** is exactly what the name suggests, a “big” amount of data. Big Data means a data set that is large in terms of volume and is more complex. Because of the large volume and higher complexity of Big Data, traditional data processing software cannot handle it.
- Big Data simply means datasets containing a large amount of diverse data, both structured as well as unstructured.
- Big Data allows companies to address issues they are facing in their business, and solve these problems effectively using Big Data Analytics. Companies try to identify patterns and draw insights from this sea of data so that it can be acted upon to solve the problem(s) at hand.

What are the 5 Vs of Big Data?

- Doug Laney introduced this concept of 3 Vs of Big Data, viz. Volume, Variety, and Velocity.
  - **Volume** refers to the amount of data that is being collected. The data could be structured or unstructured.
  - **Velocity** refers to the rate at which data is coming in.
  - **Variety** refers to the different kinds of data (data types, formats, etc.) that is coming in for analysis.
- Over the last few years, 2 additional Vs of data have also emerged – value and veracity.
  - **Value** refers to the usefulness of the collected data.
  - **Veracity** refers to the quality of data that is coming in from different sources.

Challenges of Big Data?

- **Data growth**
  - Managing datasets having terabytes of information can be a big challenge for companies. As datasets grow in size, storing them not only becomes a challenge but also becomes an expensive affair for companies.
- **Data security**
  - Data security is often prioritized quite low in the Big Data workflow, which can backfire at times. With such a large amount of data being collected, security challenges are bound to come up sooner or later.
  - Mining of ***sensitive information***, ***fake data generation***, and ***lack of cryptographic protection*** (encryption) are some of the challenges businesses face when trying to adopt Big Data techniques.
- **Data integration**
  - Data is coming in from a lot of different sources (social media applications, emails, customer verification documents, survey forms, etc.). It often becomes a very big operational challenge for companies to combine and reconcile all of this data.

The future of Big Data

- The volume of data being produced every day is continuously increasing, with increasing digitization. More and more businesses are starting to shift from traditional data storage and analysis methods to cloud solutions.
- Companies are starting to realize the importance of data. All of these imply one thing, the future of Big Data looks promising! It will change the way businesses operate, and decisions are made.

Data Science

- **Data Science** is a combination of multiple disciplines that uses statistics, data analysis, and machine learning to analyze data and to extract knowledge and insights from it.
- Data Science is about data gathering, analysis and decision-making.
- Data Science is about finding patterns in data, through analysis, and make future predictions.
- By using Data Science, companies are able to make:
  - Better decisions (should we choose A or B)
  - Predictive analysis (what will happen next?)
  - Pattern discoveries (find pattern, or maybe hidden information in the data)

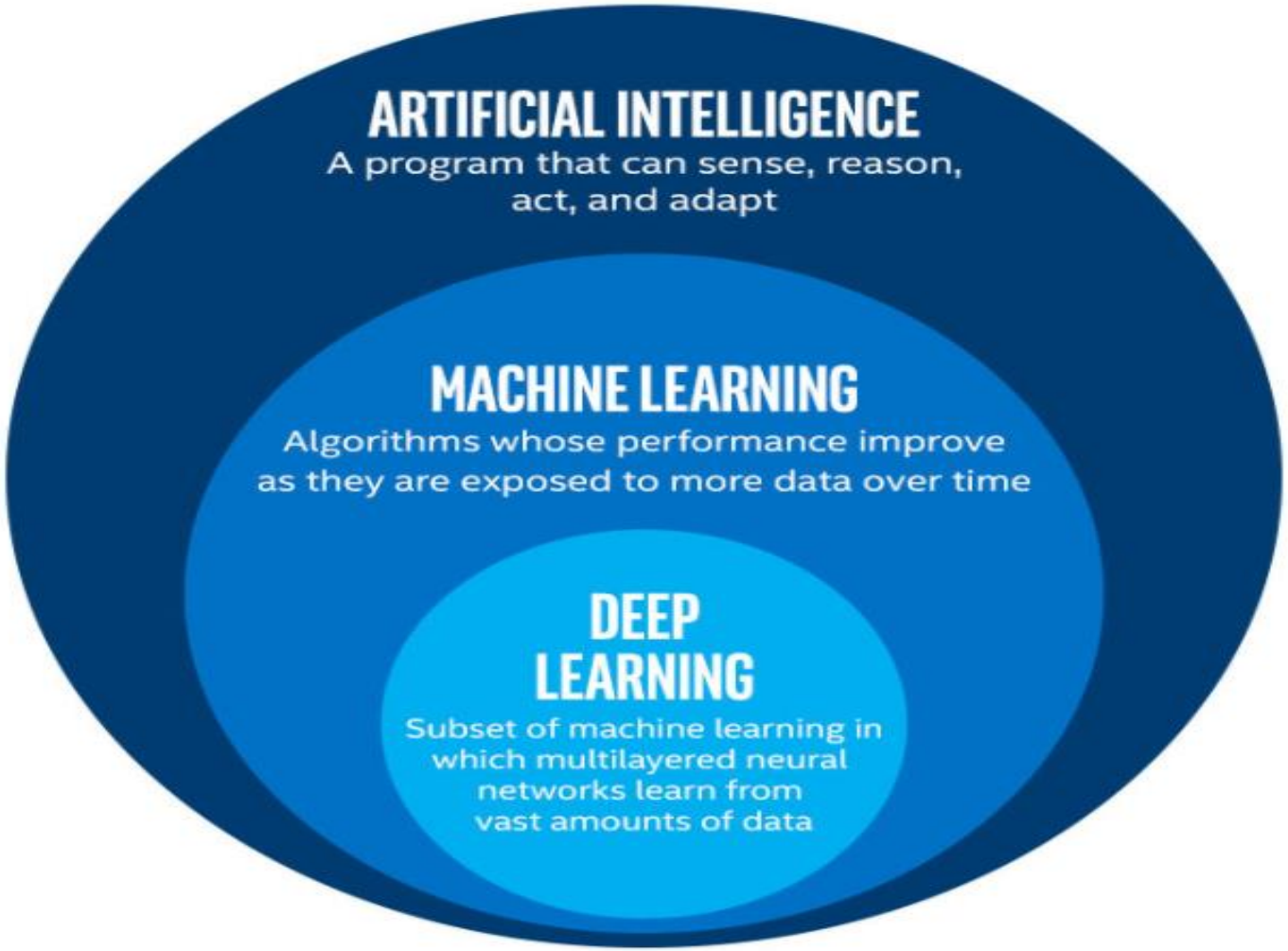
Where is Data Science Needed? (Application)

- Data Science is used in many industries in the world today, e.g. banking, consultancy, healthcare, and manufacturing.
- Examples of where Data Science is needed:
  - For route planning: To discover the best routes to ship
  - To foresee delays for flight/ship/train etc. (through predictive analysis)
  - To create promotional offers
  - To find the best suited time to deliver goods
  - To forecast the next years revenue for a company
  - To analyze health benefit of training
  - To predict who will win elections

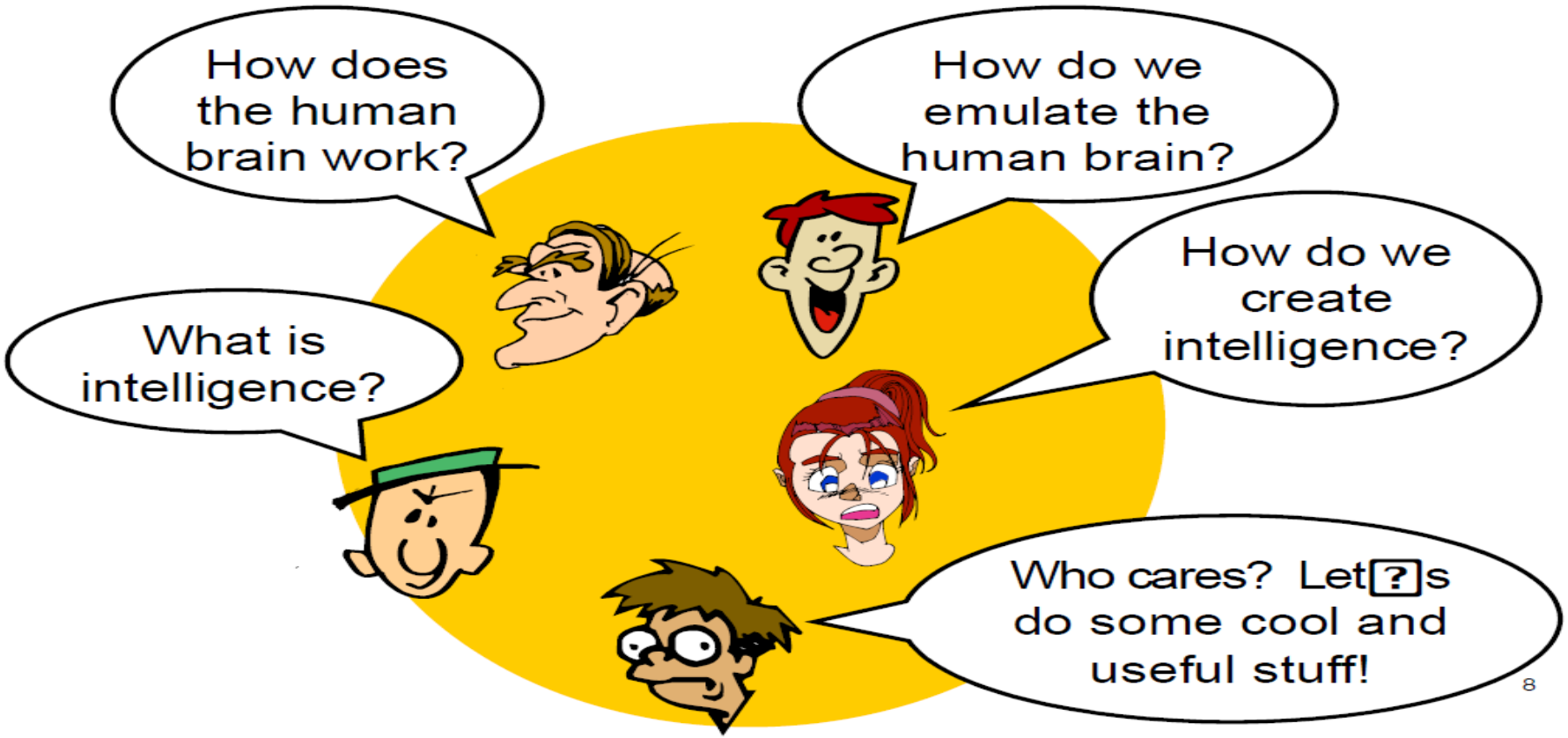
Where is Data Science Needed? (Application)

- Data Science can be applied in nearly every part of a business where data is available.
- Examples are:
  - Consumer goods
  - Stock markets
  - Industry
  - Politics
  - Logistic companies
  - E-commerce
- **A Data Scientist requires expertise in several backgrounds:**
  - Machine Learning
  - Statistics
  - Programming (Python or R)
  - Mathematics
  - Databases

Artificial Intelligence (AI)


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

What is AI?


How does the human brain work?
How do we emulate the human brain?
What is intelligence?
How do we create intelligence?
Who cares? Let[?]s do some cool and useful stuff!
8

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
- **Formal tasks:** prove a logic theorem, geometry, calculus, play chess, checkers etc.
- **Expert tasks:** engineering design, medical designers, financial analysis.

What is Artificial Intelligence?

- Based on the above, ***`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 is a branch of computer science that aims to create intelligent machines. It has become an essential part of the technology industry.
- Research associated with artificial intelligence is highly technical and specialized. The core problems of artificial intelligence include programming computers for certain traits such as: ***Knowledge Reasoning***, ***Problem solving***, ***Perception***, ***Learning***, ***Planning***, ***Ability to manipulate*** and ***move objects***.

Examples of AI Technology

- **Automation:** What makes a system or process function automatically.
- **Machine learning:** The science of getting a computer to act without programming.
- **Natural language processing (NLP):** The processing of human -- and not computer -- language by a computer program.
- **Robotics:** A field of engineering focused on the design and manufacturing of robots.
- **Self-driving cars:** These use a combination of computer vision, image recognition and deep learning to build automated skill at piloting a vehicle while staying in a given lane and avoiding unexpected obstructions, such as pedestrians.

Applications AI

- Artificial Intelligence has made its way into a number of areas.
  - **AI in healthcare:** The biggest bets are on improving patient outcomes and reducing costs. Companies are applying machine learning to make better and faster diagnoses than humans.
  - **AI in business:** Robotic process automation is being applied to highly repetitive tasks normally performed by humans. Machine learning algorithms are being integrated into analytics and CRM platforms to uncover information on how to better serve customers.
  - **AI in education:** AI can automate grading, giving educators more time. AI can assess students and adapt to their needs, helping them work at their own pace. AI tutors can provide additional support to students, ensuring they stay on track. AI could change where and how students learn, perhaps even replacing some teachers.

Applications AI (Contd.)

- **AI in finance:** AI in personal finance applications, such as Mint or Turbo Tax, is disrupting financial institutions. Applications such as these collect personal data and provide financial advice. Other programs, such as IBM Watson, have been applied to the process of buying a home. Today, software performs much of the trading on Wall Street.
- **AI in law:** The discovery process, sifting through of documents, in law is often overwhelming for humans. Automating this process is a more efficient use of time. Startups are also building question-and-answer computer assistants that can sift programmed-to-answer questions by examining the taxonomy and ontology associated with a database.
- **AI in manufacturing:** This is an area that has been at the forefront of incorporating robots into the workflow. Industrial robots used to perform single tasks and were separated from human workers, but as the technology advanced that changed.

Machine Learning (ML)

Introduction

- ***Machine learning*** is a specialized technology that falls under the umbrella of artificial intelligence (AI). This exciting field is the driving power behind many modern technologies, including ***image recognition***, ***self-driving cars***, and products like ***Amazon's Alexa***.
- This branch of AI focuses on using data and algorithms to mimic human learning, allowing machines to improve over time, becoming increasingly accurate when making predictions or classifications, or uncovering data-driven insights.
- It works in three basic ways, ***starting with using a combination of data*** and ***algorithms to predict patterns*** and ***classify data sets***, ***an error function that helps evaluate the accuracy***, and then ***an optimization process to fit the data points into the model best***.

What is ML?

"***Machine Learning*** is defined as the study of computer programs that leverage algorithms and statistical models to learn through inference and patterns without being explicitly programed. Machine Learning field has undergone significant developments in the last decade."

Applications of Machine Learning

- Machine learning is already used around us and you may not realize how it impacts your life. One example is the ***spam filter of your email provider***. That filter uses an algorithm to continually identify new types of junk messages and moves them to your spam folder instead of into your inbox. That's just one example of machine learning that you're probably already using.
- **Social media features:**

  Social media platforms integrate machine learning algorithms to help deliver personalized experiences to you. Facebook notes your activities, including your comments, likes, and the time you spend on different types of content. The algorithm learns from your activity and makes pages and friend suggestions tailored to you.
- **Virtual assistants:**

  Apple's Siri, Amazon's Alexa, and Google Now are all popular options if you're looking for a virtual personal assistant. These voice-activated devices can do everything from search for flights to check your schedule to set alarms and more. Machine learning is a key component of these smart devices and speakers. They collect information and refine it each time you interact with them. The machine can then use that data to give you results that are best matched to your preferences.

Applications of Machine Learning

- **Product recommendations:**

  Popular among e-commerce websites, product recommendations are a common machine learning application. It lets these sites track your behavior based on your searches, previous purchases, and your shopping cart history to make suggestions and recommendations about products you may be interested in.

- **Image recognition:**

  Image recognition is a complex technology that holds promise in a variety of fields. In your everyday life, you've probably come across this while uploading a photo to your social media platform. When you ***tag someone in an image***, the platform recognizes them. It can also be transformative for identifying potential threats or criminals, unlocking phones and mobile devices, and finding missing persons.

Artificial Neural Network (ANN)

Introduction

- ***Artificial Neural Network*** **(ANN)** is a deep learning algorithm that emerged and evolved from the idea of Biological Neural Networks of human brains. An attempt to simulate the workings of the human brain culminated in the emergence of ANN. ANN works very similar to the biological neural networks but doesn't exactly resemble its workings.
- The main objective is to develop a system to perform various computational tasks faster than the traditional systems. These tasks include pattern recognition and classification, approximation, optimization, and data clustering.
- ANN algorithm would accept only numeric and structured data as input. To accept unstructured and non-numeric data formats such as Image, Text, and Speech, ***Convolutional Neural Networks*** **(CNN)**, and ***Recursive Neural Networks*** **(RNN)** are used respectively.

What is Artificial Neural Network?

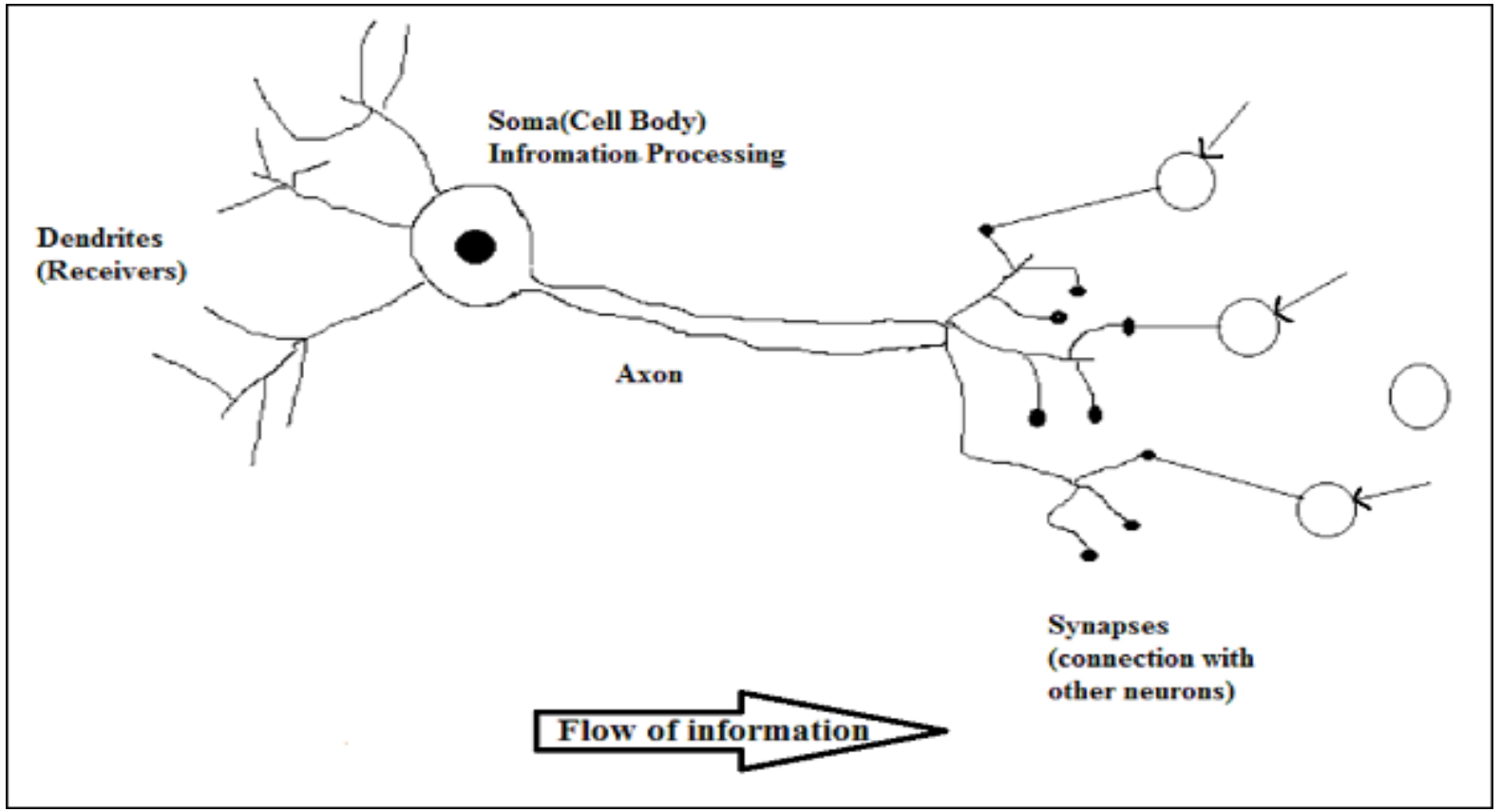
- Artificial Neural Network ANN is an efficient computing system whose central theme is borrowed from the analogy of biological neural networks. ANNs are also named as **“artificial neural systems,”** or **“parallel distributed processing systems,”** or **“connectionist systems”**. ANN acquires a large collection of units that are interconnected in some pattern to allow communication between the units. These units, also referred to as nodes or neurons, are simple processors which operate in parallel.
- Every neuron is connected with other neuron through a connection link. Each connection link is associated with a weight that has information about the input signal. This is the most useful information for neurons to solve a particular problem because the weight usually excites or inhibits the signal that is being communicated. Each neuron has an internal state, which is called an activation signal. Output signals, which are produced after combining the input signals and activation rule, may be sent to other units.

Characteristics of Artificial Neural Network

- It is neurally implemented mathematical model.
- It contains huge number of interconnected processing elements called neurons to do all operations.
- Information stored in the neurons are basically the weighted linkage of neurons.
- The input signals arrive at the processing elements through connections and connecting weights.
- It has the ability to learn , recall and generalize from the given data by suitable assignment and adjustment of weights.
- The collective behavior of the neurons describes its computational power, and no single neuron carries specific information.

Biological Neuron

- A nerve cell neuron is a special biological cell that processes information. According to an estimation, there are huge number of neurons, approximately 10^{11} with numerous interconnections, approximately 10^{15}.



Working of a Biological Neuron

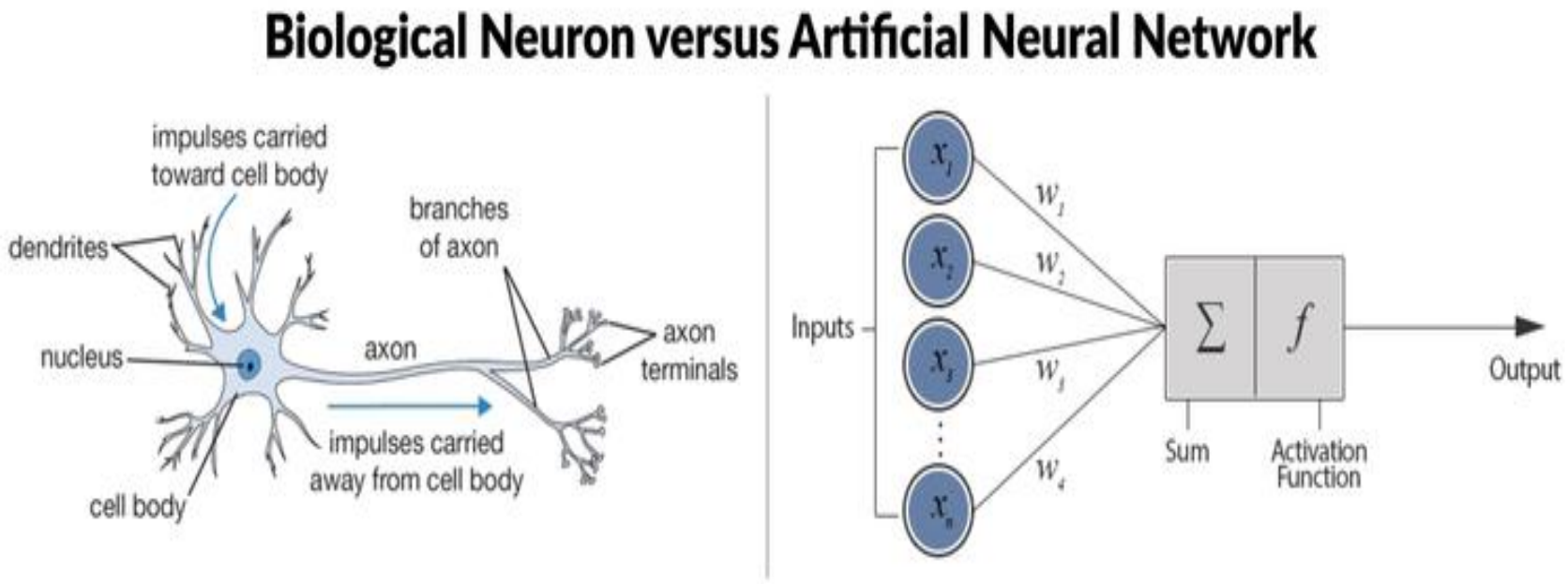
- As shown in the above diagram, a typical neuron consists of the following four parts with the help of which we can explain its working −
  - **Dendrites:** They are tree-like branches, responsible for receiving the information from other neurons it is connected to. In other sense, we can say that they are like the ears of neuron.
  - **Soma:** It is the cell body of the neuron and is responsible for processing of information, they have received from dendrites.
  - **Axon:** It is just like a cable through which neurons send the information.
  - **Synapses:** It is the connection between the axon and other neuron dendrites.

BNN Vs ANN

- Before taking a look at the differences between Artificial Neural Network (ANN) and Biological Neural Network (BNN), let us take a look at the similarities based on the terminology between these two:

Biological Neural Network BNN	Artificial Neural Network ANN
Soma	Node
Dendrites	Input
Synapse	Weights or Interconnections
Axon	Output

Fig: BNN Vs ANN



Applications of ANN

- Every new technology need assistance from the previous one i.e. data from previous ones and these data are analyzed so that every pros and cons should be studied correctly. All of these things are possible only through the help of neural network.
- Neural network is suitable for the research on ***Animal behavior, predator/prey relationships and population cycles.***
- It would be easier to do ***proper valuation*** of property, buildings, automobiles, machinery etc. with the help of neural network.
- Neural Network can be used in betting on horse races, sporting events, and most importantly in stock market.
- It can be used to predict the correct judgment for any crime by using a large data of crime details as input and the resulting sentences as output.
- By analyzing data and determining which of the data has any fault (files diverging from peers) called as ***Data mining, cleaning and validation*** can be achieved through neural network.

Applications of ANN

- Neural Network can be used to predict targets with the help of echo patterns we get from sonar, radar, seismic and magnetic instruments.
- It can be used efficiently in ***Employee hiring*** so that any company can hire the right employee depending upon the skills the employee has and what should be its productivity in future.
- It has a large application in ***Medical Research.***
- It can be used to for ***Fraud Detection*** regarding credit cards, insurance or taxes by analyzing the past records .

Business Intelligence (BI)

Business Intelligence (BI) Introduction

- **Business intelligence (BI)** is all about turning an organization's data into insights that can be used to inform business decisions. BI analysts will use BI tools, software or services to access and analyze datasets and translate their findings into reports, summaries, dashboards, graphs, charts or maps.
- In recent years, the advent of modern data visualization and reporting tools has transformed the discipline, empowering businesses to use big data insights to identify, develop and create new business opportunities.
- BI should not be confused with 'business analytics'. ***Business intelligence*** is descriptive and uses metrics to generate clear snapshots of business performance. Meanwhile, ***business analytics*** is predictive, and describes what organizations should do in future to generate better outcomes.

BI Examples

- There are two main strands within BI. The first is ***traditional BI***, where IT professionals use in-house transactional data to generate reports. The second is ***modern BI***, where business users interact with agile, intuitive systems to analyze data more quickly.
- Organizations generally use classic BI for types of reporting where accuracy is paramount and the questions and datasets used are standard or predicable. This might include regulatory or financial reports.
- Meanwhile, modern BI tools are best suited to situations where business users need insight into fast-changing dynamics so that they can respond to events as they unfold.
- For example, American restaurant Chipotle uses a self-service BI platform to create a centralized view of its operations and track their effectiveness.
- Similarly, Coca-Cola Bottling Company uses mobile BI dashboards to put timely, actionable CRM data in the hands of its sales teams.

Why is BI important?

- Creating KPI (Key Performance Indicator) based on historic data.
- Identify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problems that need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small Medium Enterprises)

BI Applications

- **Data mining:** Using databases, statistics and machine learning to uncover trends in large datasets.
- **Reporting:** Sharing data analysis to stakeholders so they can draw conclusions and make decisions.
- **Performance metrics and benchmarking:** Comparing current performance data to historical data to track performance against goals, typically using customized dashboards.
- **Querying:** Asking the data specific questions, BI pulling the answers from the datasets.
- **Statistical analysis:** Taking the results from descriptive analytics and further exploring the data using statistics such as how this trend happened and why.
- **Data visualization:** Turning data analysis into visual representations such as charts, graphs, and histograms to more easily consume data.
- **Data preparation:** Compiling multiple data sources, identifying the dimensions and measurements, preparing it for data analysis.

Remote Sensing and GIS

Geographical Information System (GIS)

- What is not GIS?
  - Is not a “software”
  - Is not “map making tool”
- GIS is a set of tools for collecting, storing, retrieving, transforming and displaying spatial data from the real world for a particular purpose. (Burrough, 1986)
- **Geographic**
  - Real world features or objects that can be referenced to a specific location in space.
- **Information**
  - Specific characteristics, descriptions or properties of real world features.
- **System**
  - Integrated and collaborative combination of hardware, software, analysis and presentation methods as well as user communities.

Components of GIS

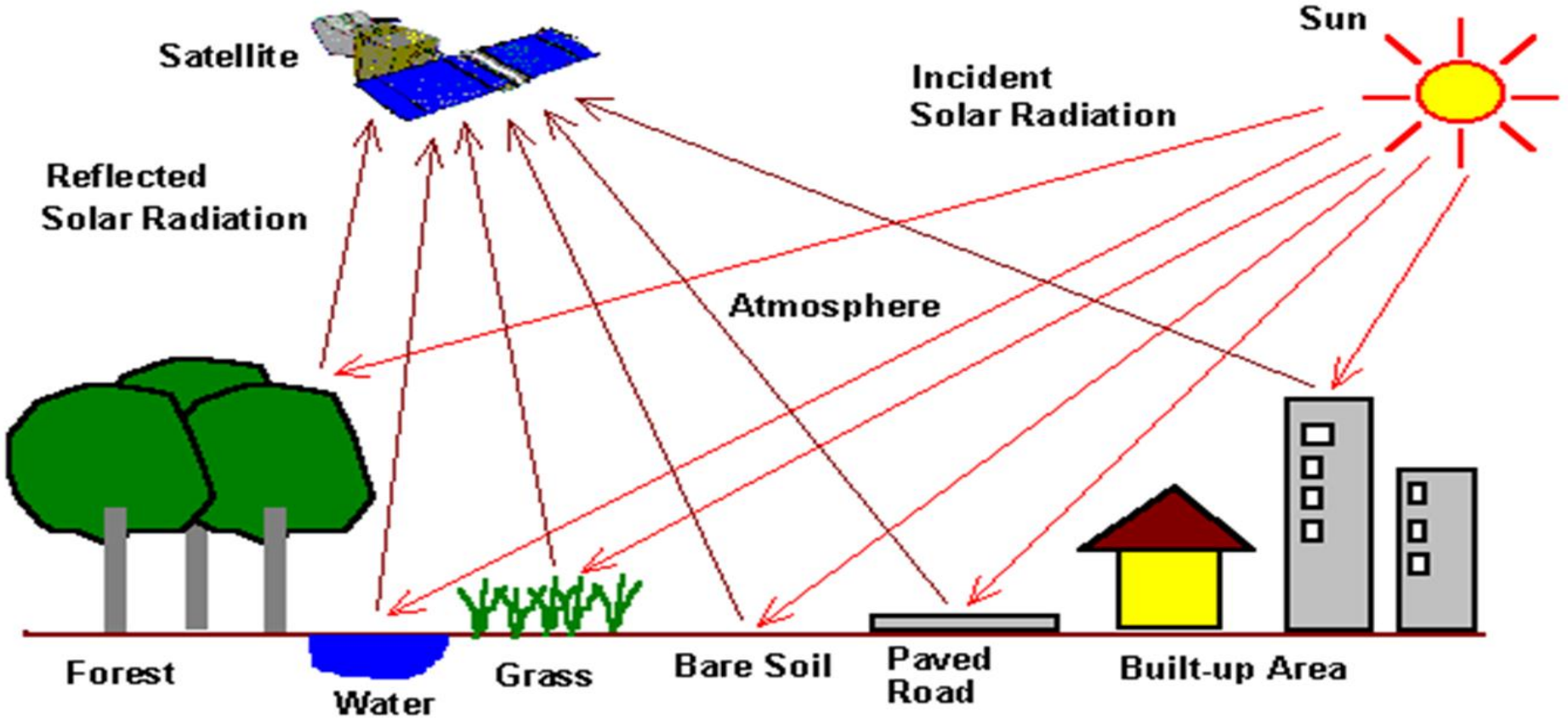
- Hardware
  - Computers, printers/plotters, scanners, network, digital tablets etc.
- Software
  - **ArcGIS / ArcINFO,** GRASS, MapINFO, ERDAS, IDRISI, ILWIS, etc.
- Data
  - Spatial and non-spatial data
- Methods
  - Methods for analysis, interpretation and presentation
- People/User Communities
  - Planners
  - Agriculture sector
  - Transportation
  - Telecommunication
  - Education etc.

Remote Sensing

- **Remote sensing** is the science and technology by which the properties of specified objects, area, or phenomenon can be identified, measured, and analyzed without direct contact with them in order to make useful decisions.
- Some examples are:
  - Cameras on satellites and airplanes take images of large areas on the Earth's surface, allowing us to see much more than we can see when standing on the ground.
  - Sonar systems on ships can be used to create images of the ocean floor without needing to travel to the bottom of the ocean.
  - Cameras on satellites can be used to make images of temperature changes in the oceans.

Remote Sensing Principle

- The source of remote sensing data is the electromagnetic radiations which are emitted or reflected by the object, which then helps in their identification and classification.



THANK YOU!