Parul® University

BIG DATA ANALYTICS
CHAPTER 1

Prepared By: Prof. Himadri Vegad

Parul®
University

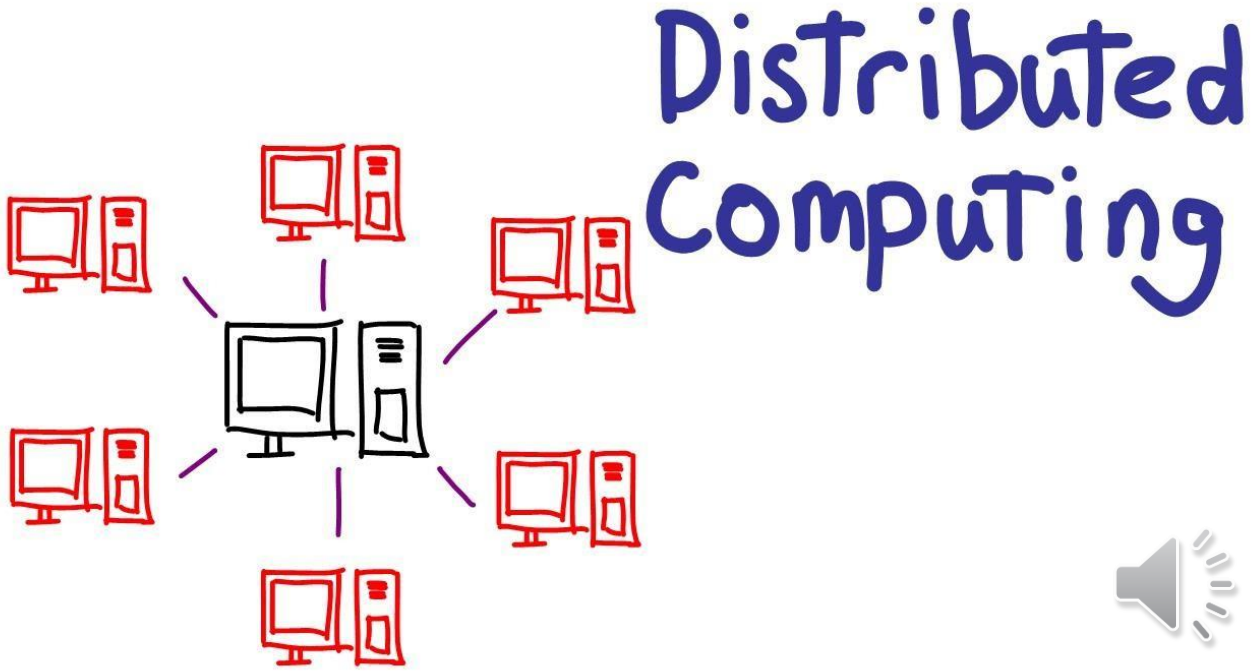
DIGITAL LEARNING CONTENT

CHAPTER-1

Distributed Learning

What is Distributed System?

A distributed system, also known as distributed computing, is a system with multiple components located on different machines that communicate and coordinate actions in order to appear as a single coherent system to the end-user.



Parul®
University

Advantage of Distributed System

- Can easily share data among themselves.
- Easily can added nodes.
- Failure of one or more node do not lead to the failure of distributed system.
- Resources can be shared with every nodes.

Disadvantage of Distributed System

- Difficult to provide security in distributed system because to make the system secure the nodes as well as the connection should be secured.
- Some message and data can be lost in the network while moving from one node to another.
- when database is connected with distributed it is become complicated.
- Overloading may occur if all the node send data at the same time.

Parul® University

DIGITAL LEARNING CONTENT

What is Data?

- Data are quantities, characters, or symbols
- Computer performed operation on quantifiers, characters or symbol.
- It may be stored and transmitted in the form of electrical signal.
- It can be also recorded on magnetic recording media.

Big Data

- Big Data is also data but big in size.
- Big data is a data which is huge in volume and still growing exponentially.
- Ex:
  - New York Stock Exchange
  - Social Media
  - Jet engine

Type of Data

- Structured
- Unstructured
- Semi Structured

Structured data	Semi-structured data	Unstructured data
Databases	XML / JSON data Email Web pages	Audio Video Image data Natural language Documents

Structured Data

- Any kind of data that are stored in a predefined format are known as structured data.
- Predefined format : DBMS
- It is well organized.
- Easily used by machine learning algorithm, easily used by business user.
- Because of limited storage options.

Structured Data at a Glance

Characteristics of Structured Data

- High organized
- Clearly defined
- Easy to access
- Easy to analyze

Examples of Structured Data

- Name
- Age
- Gender
- Address
- Phone number
- Currency
- Dates
- Billing info

Sources of Structured Data

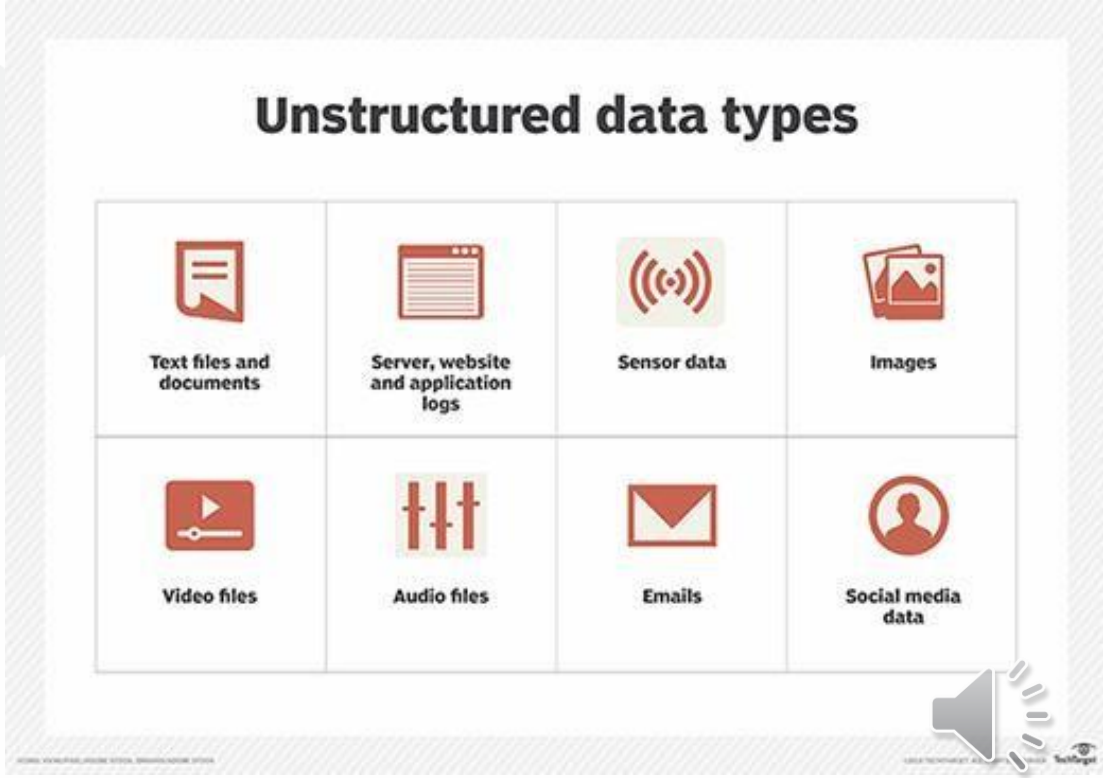
- SQL databases
- Spreadsheets
- Sensors
- Medical Devices
- Online Forms
- Point of Sales Systems
- Web and Server Logs

Semi Structured Data

- Semi-structured data is like unstructured data but it contains metadata.
- Mata data contains information of the data.
- With the help of metadata the data is easily accessable,analized like structured data.

Unstructured Data

- The data which are stored in a random format, it is not processed until it is used.
- Freedom to store in any format, faster accumulate rate.
- Require expert data scientist, specialized tools.




BIG
DATA
Evolution

Three Characteristic of Big Data

- Volume : Volume refers to the amount of data generated through websites, portal and online application.

- Velocity: Velocity refer the speed by which data is generated. Like facebook,google,twitter many other social media create data every day.

- Variety: Variety refers to all structured and unstructured data that are generated by human or machines.

Challenges of Big Data

- Lack of proper understanding of Big Data: As it is new technology ,there is insuffient understanding of the Big Data.
- Data growth issues : As it growing rapidly , to store such amount of data is huge task.
- Confusion while Big Data tool selection: As it is new professional
- Lack of data professionals
- Securing data

Why Big Data?

1. Data driven decisions provide advantage : Studies have shown that data driven decision are more effective than human generated decision. Big Data help the organization to know the trends and pattern that can be used for future benefit.

2. Big Data provides a spring board for AI: Artificial intelligence (AI) is the most desire areas of expertise in business today. Big Data acts like food for the AI projects. For running AI algorithm it require a large amount of data which it get from Big Data.

3. Big Data Demand: As every day new technology is coming in the market. All this technology is creating or required data. In both the situation high demand of big data expert is in the market.

Why Big Data Analytics?

- Risk Management :Big Data analytics to identify fraudulent activities and discrepancies.
- Product Development and Innovations: Big Data analytics to analyze how efficient the product design and whether there is any scope of improvement.
- Quicker and Better Decision Making Within Organizations: Big Data analytics uses features of the data and give a quicker and better decision than human.
- Improve Customer Experience: Big Data analytics analyse the feedback of the customer so that they can improve the experience of the customer by improving their product and service.