Big Data Analytics and Visualization (CDSC 715)

Addisu G. Semie, PhD
Asst. Prof.,Computational Science Program,
Addis Ababa University
Email: addisu.semie@aau.edu.et

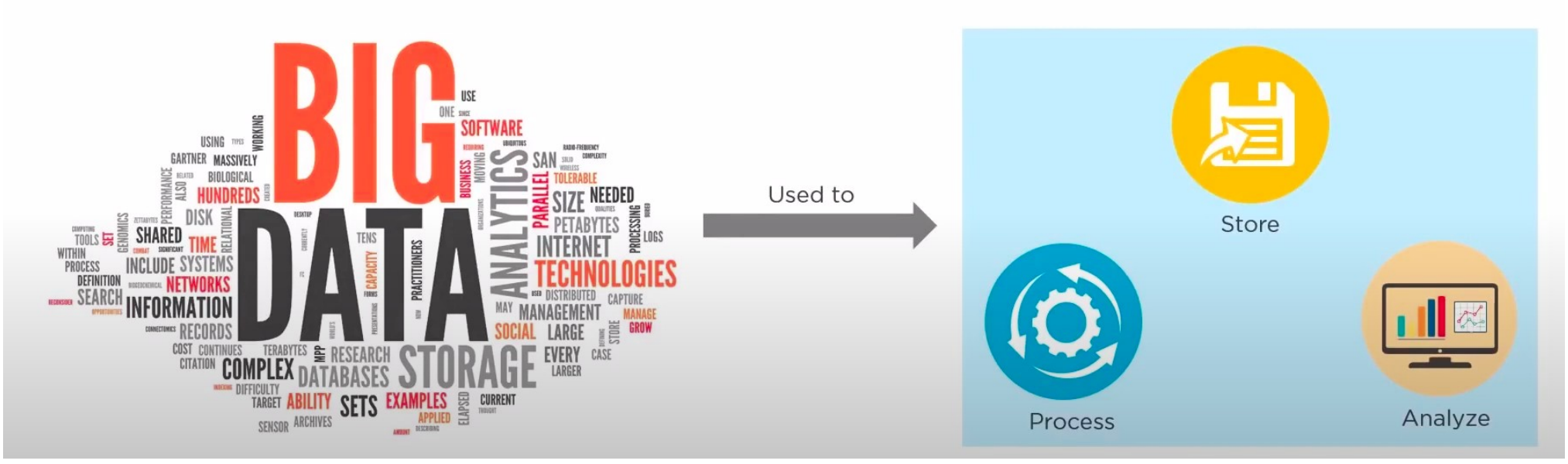
Basics of Big Data Analytics

- What is Big Data?
- Evolution of Big Data
- Why Big data?
- Characteristics of Big Data
- Applications of Big Data


What is Big Data?

What is Big Data?

- Big Data is a term used to describe data that is too large and complex to store, manage, process and analyze the data using traditional databases and data processing tools.



Evolution of Big Data

- Data has evolved in the last 5 years like never before. Lots of data is being generated each day in every business sector

Evolution of Big Data

- Here are some facts to convince you that data is exploding and needs your attention

Every minute, users send 31.25 million messages and watch 2.77 million videos on Facebook

Walmart handles more than 1 million customer transactions every hour

40,000 search queries are performed on Google per second, i.e. 3.46 million searches a day

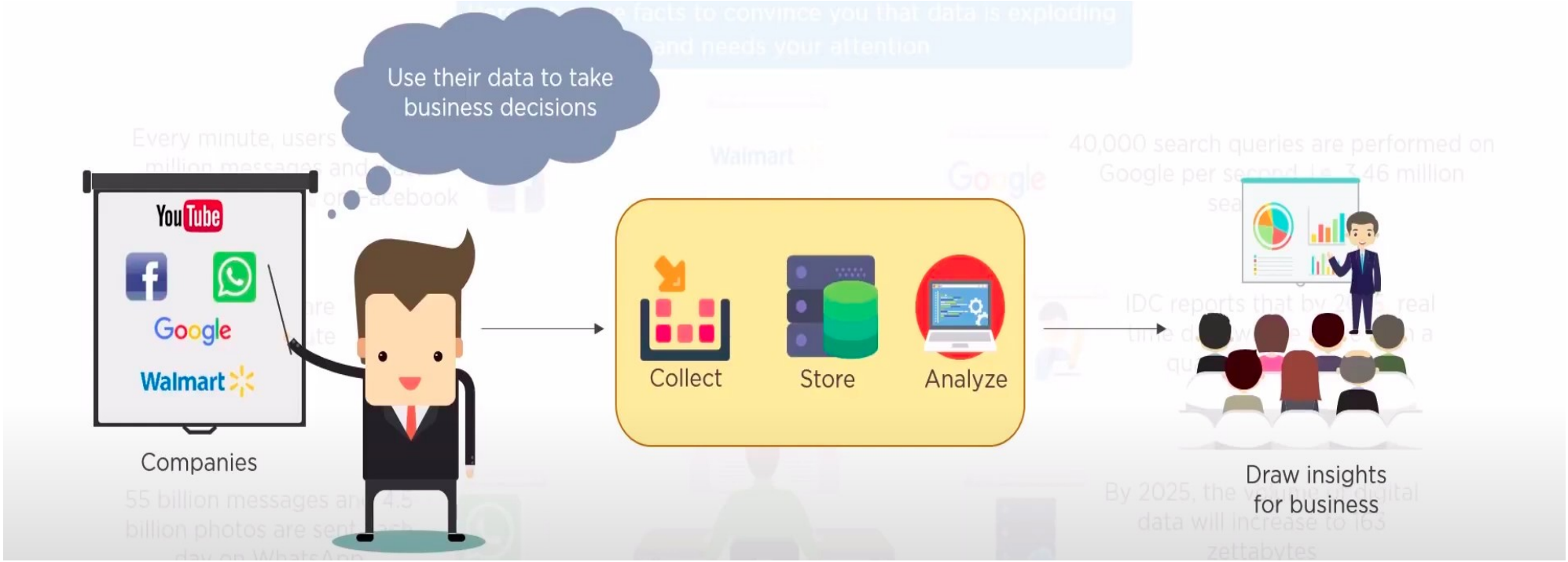
300 hours of video are uploaded every minute on YouTube

IDC reports that by 2025, real time data will be more than a quarter of all the data

55 billion messages and 4.5 billion photos are sent each day on WhatsApp

By 2025, the volume of digital data will increase to 163 zettabytes

Why Big Data?

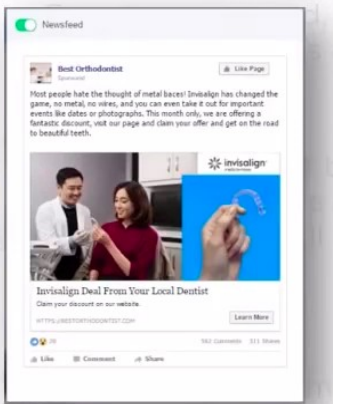

Use their data to take business decisions
YouTube
Google
Walmart
Companies
Collect
Store
Analyze
Draw insights for business

Why Big Data?

Here is what Facebook does with its big data

Facebook collects huge volumes of user data

Analyses the data

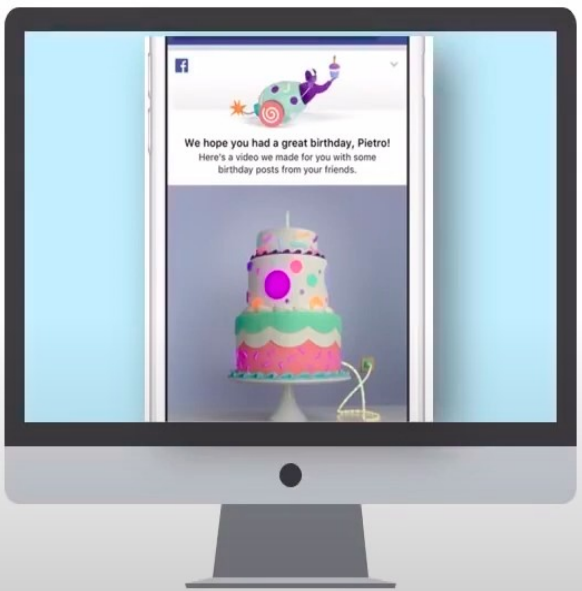


Personalized News Feed

Why Big Data?

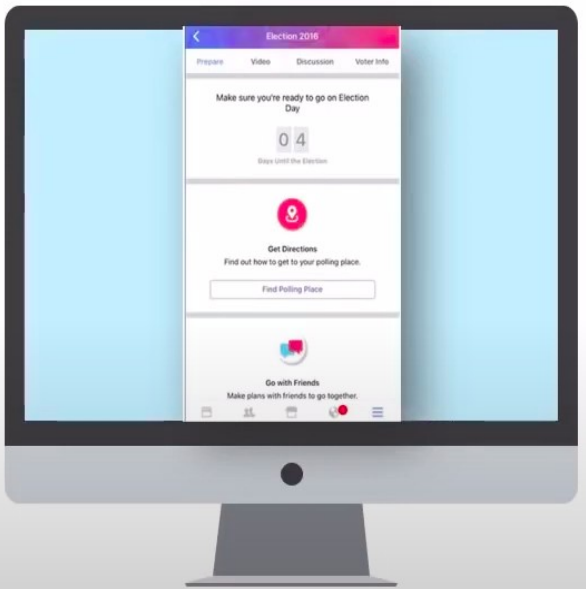
- Some more examples that show how Facebook uses its big data

The Flashback



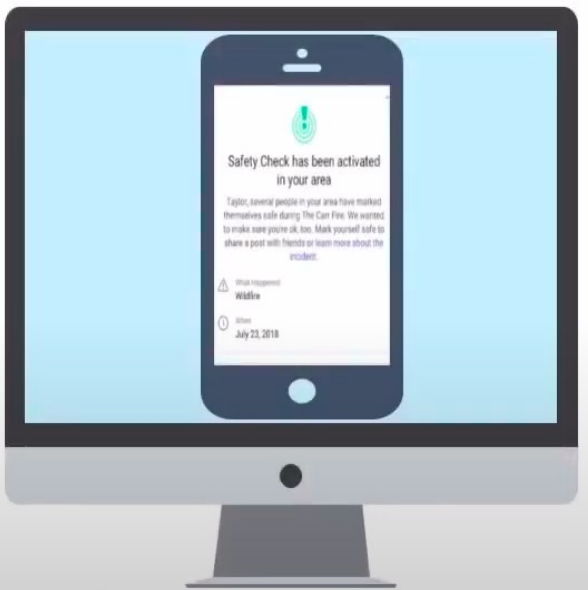
Collection of photos and posts that received the most comments and likes

I Voted



Used for 2016 elections, with reminders and directions to tell users their time and place of polling

Safety Checks



In incidents such as earthquake, hurricane or mass shooting, Facebook gives you safety checks

Characteristics of Big Data

Characteristics of Big Data

Size of the data

Volume

Incredible amount of data

Data generated each second

From cell phones, social media, online transactions, etc.

Characteristics of Big Data

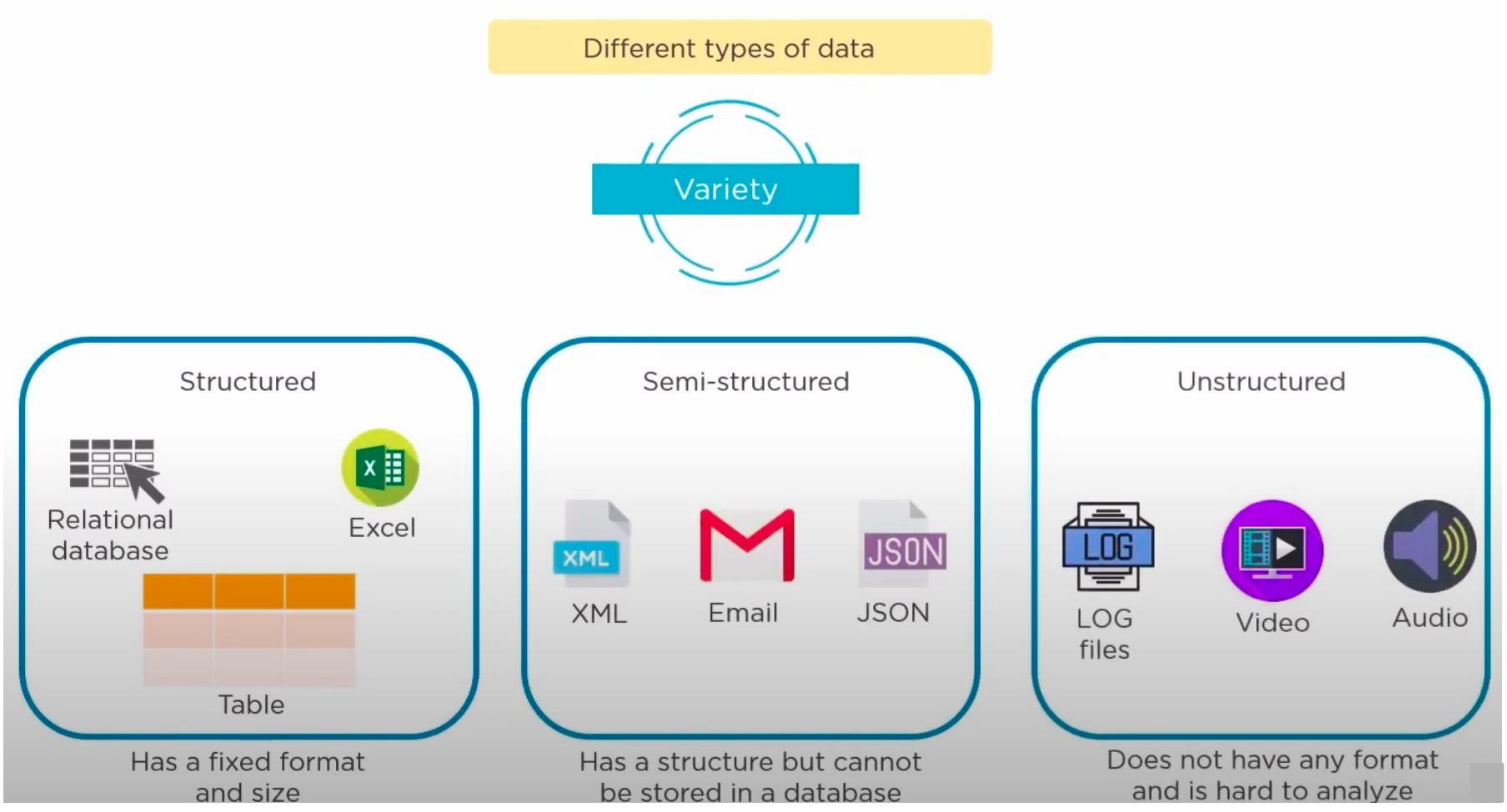
Speed at which data is generated

Velocity

Speed at which data is:

- Generated
- Collected
- Analyzed

Characteristics of Big Data



Characteristics of Big Data

- Has a defined organizational structure (excel,csv)
- Represented using a schema
- Databases such as **Teradata, Greenplum** (commercial) and **Redis, Cassandra, and Hive** (open source) are examples of technologies that provide the ability to manage and query structured data.

Characteristics of Big Data

Semi-structured Data

- Has elements of organizational schema as well as aspects that are arbitrary (ex. A personal phone diary, emails)
- Represented by JSON formats
- Mongo and CouchDB are generally used to store and query semi-structured datasets

Characteristics of Big Data

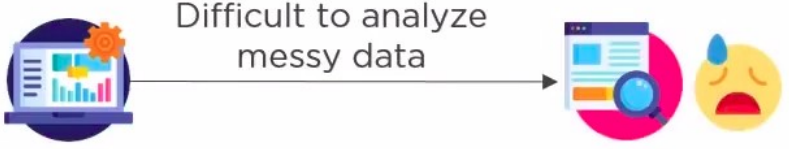
- Doesn’t have predefined organizational schema (Ex. Spoken words, music, videos, and books)
- It can be stored in various formats such as Blobs or textual data.
- Technologies such as Lucene/Solr, Elasticsearch, and others are generally used to query, index, and other operations on textual data.

Characteristics of Big Data

- Trustworthiness of data in terms of quality and accuracy

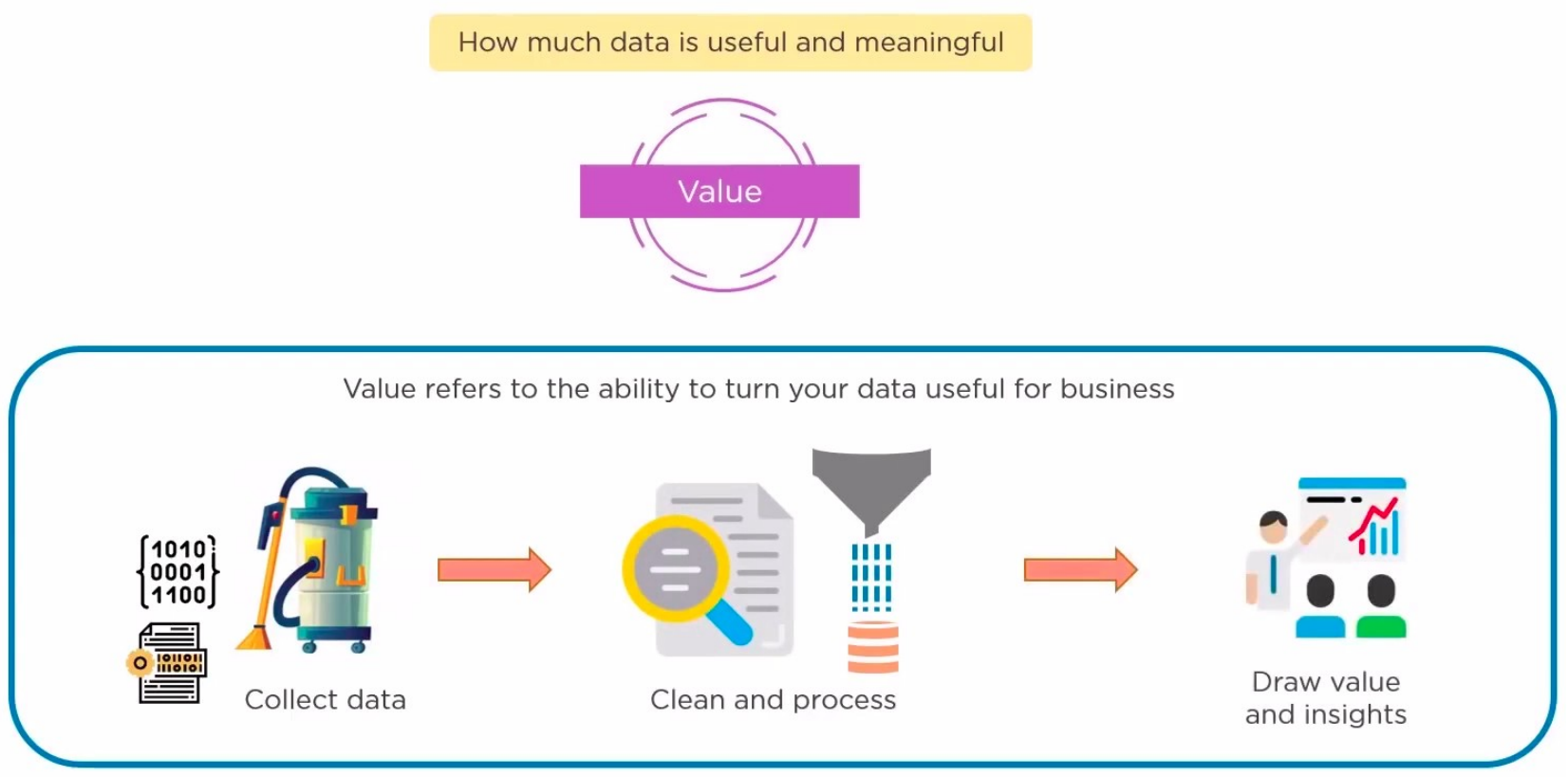
Extracting loads of data is not useful if the data is messy and poor in quality

Twitter posts with abbreviations, spelling mistakes, etc.



Characteristics of Big Data

- How much data is useful and meaningful



Application of Big Data

- Big Data has found a wide range of applications across various industries and fields due to its ability to process and analyze vast amounts of data quickly and efficiently.

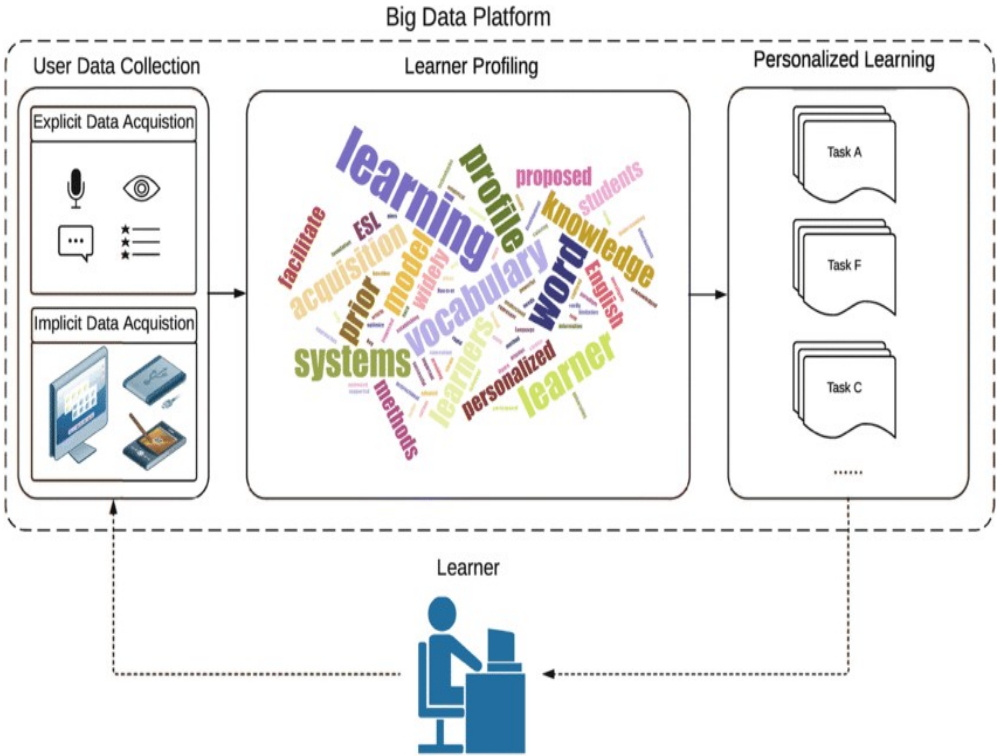
Healthcare and Medicine

- Disease Surveillance: Analyzing large healthcare datasets to track and predict disease outbreaks.
- Personalized Medicine: Tailoring medical treatments and therapies based on individual patient data.
- Drug Discovery: Identifying potential drug candidates by analyzing genetic and chemical data.

Agriculture

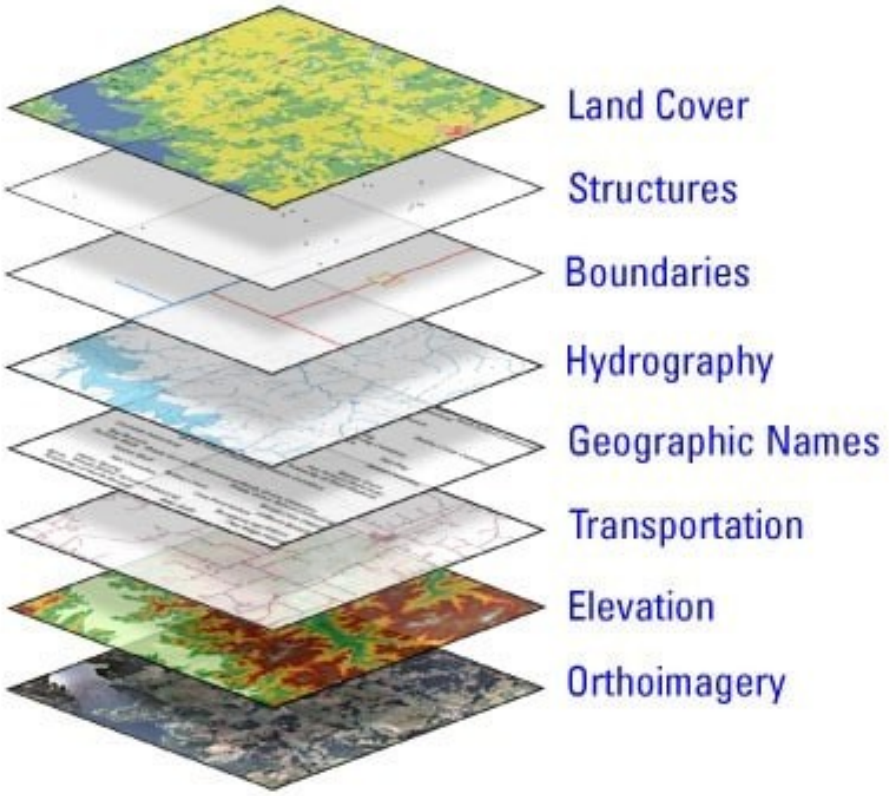
- Precision Agriculture: Using sensors and data analysis to optimize crop yields and reduce resource use.
- Weather and Climate Forecasting: Utilizing large-scale environmental data for more accurate weather predictions and climate modeling.

Education



- Personalized Learning: Tailoring educational content to individual student needs and learning patterns.
- Student Performance Tracking: Analyzing student data to identify areas of improvement and evaluate educational programs.

Geospatial Analysis



- Big data analytics has revolutionized the field of geo-spatial analysis by enabling the processing, analysis, and visualization of massive and complex spatial datasets.
  - For example: Satellite imagery analysis, Urban planning and smart cities, Environment monitoring and management and others

Thank you!

Questions?