



WE'VE DECIDED  
TO TAKE BIG  
DATA TO THE  
NEXT LEVEL...

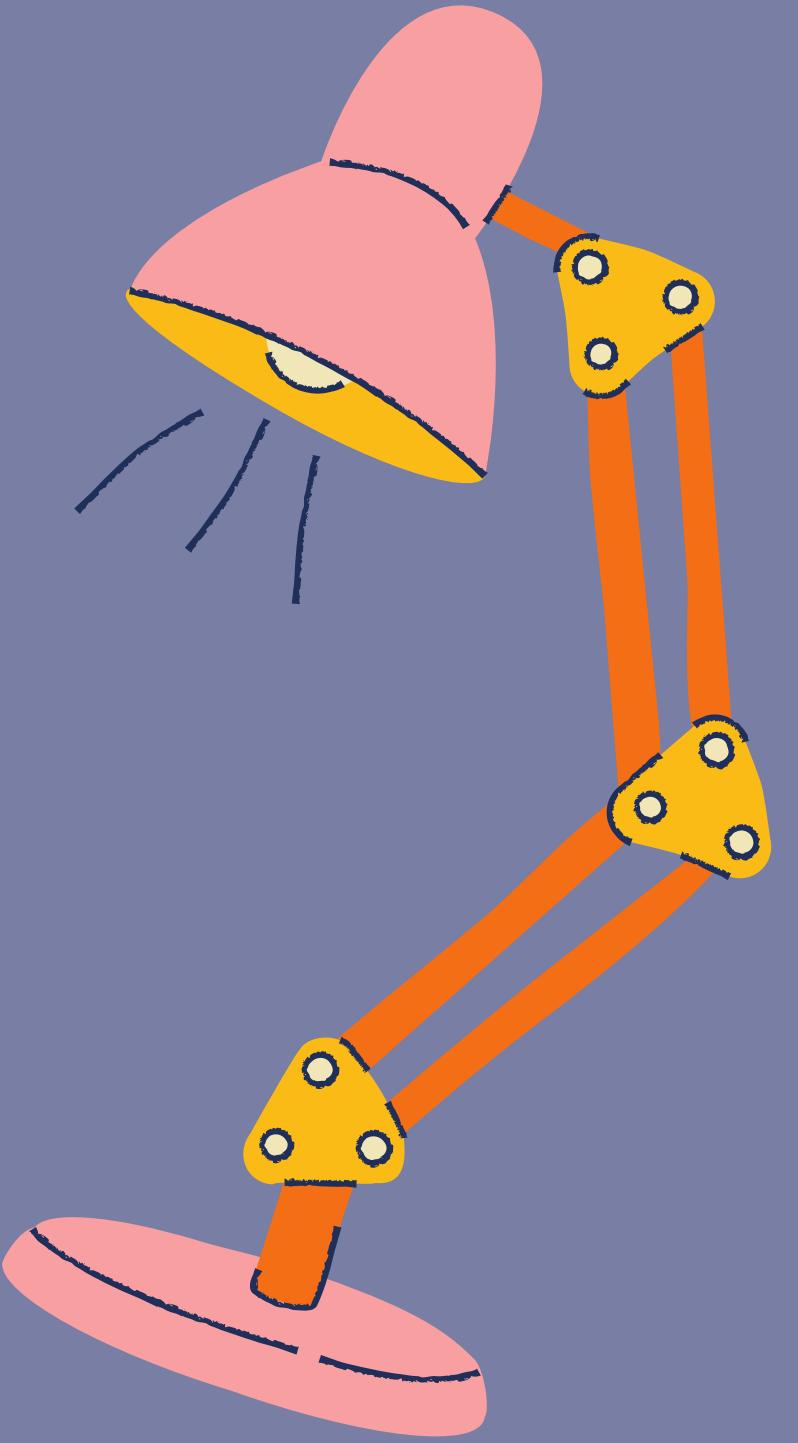
HUMONGOUS  
DATA

Big  
Data

Elective 2

# WHAT IS A BIG DATA?

-Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.



# 4 V's of Big Data

## Volume

Terabytes, Records, Transactions,  
Tables and Files



## Velocity

numbers of likes, tweets, google  
searches, emails sent, etc.



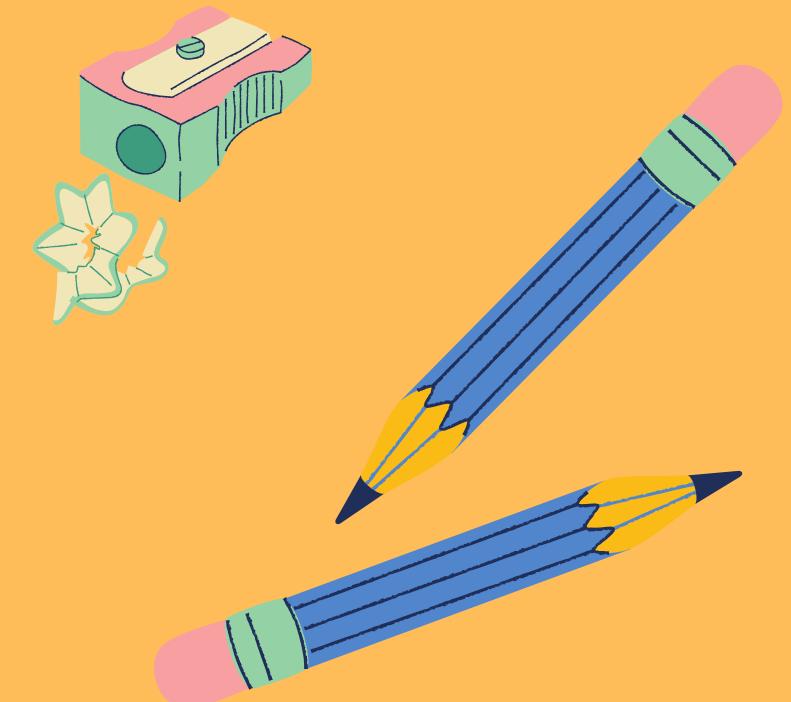
## Variety

data can be gathered using emails,  
PDFs, photos, videos, audios, SM posts,  
etc.



## Veracity

medical experiment or trial



# Example of Big Data Careers



## Big Data Engineer

An IT professional who is responsible for designing, building, testing and maintaining complex data processing systems that work with large data sets.

## Data Architect

Builds and maintain a company's database by identifying structural and installation solutions.

## Data Security Analyst

A professional in charge of designing and implementing protection for organization networks

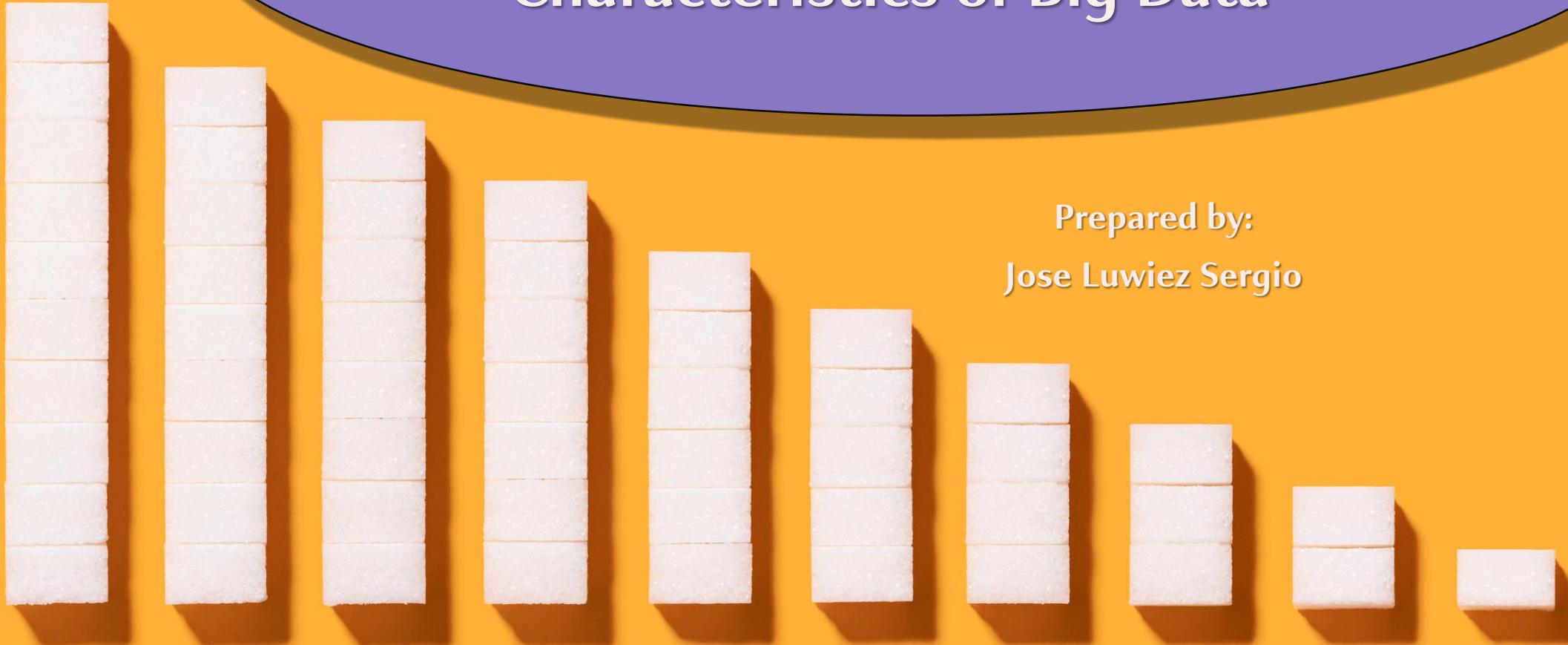
## Data Scientist

a professional responsible for collecting, analyzing and interpreting extremely large amounts of data. The data scientist role is an offshoot of several traditional technical roles, including mathematician, scientist, statistician and computer professional.

# 5 V's of Big Data

## Characteristics of Big Data

Prepared by:  
Jose Luwiez Sergio



## 5 V's of Big Data

Volume

Variety

Value

Veracity

Velocity

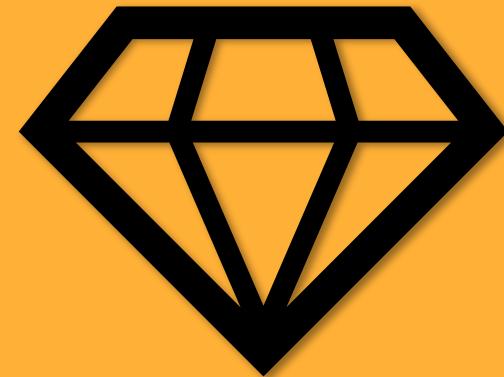
# Volume

- Initial size and amount of data that is being collected. It can only be considered a big data if the volume of data is abundant.
- On this day and age, we are currently using distributed systems to store data in several locations and brought together by a software Framework like Hadoop.
- Loads of data are managed by big data companies such as Meta, Twitter, or any applications that exchange information.



- Valuable, reliable, and trustworthy data needs to be stored, processed, and analyzed to get a better understanding of big data a company or organization can receive in the future.
- It is said to be the most essential "V" from the business perspective. The value of usually comes from insight discovery, pattern recognition, marketing strategies that can lead to more effective operations, stronger business-to-business & customer relationships, and other business benefits.

# Value



- Unstructured data is data that is unorganized or messy because it comes in different files or formats. It is not a good fit for a relational database because it does not fit into conventional data models. It can be referred as raw data.
- Semi-structured data is data that has not been organized into a specialized category or repository but has associated information, such as metadata, which makes the data easier to process.
- Structured data is data that has been allocated or organized into a formatted repository which means the data is more addressable for more effective data processing and analysis.

# Variety



**It refers to the range and diversity of different data types. The collected data may come in unstructured, semi-structured, or structured form.**

- It refers to the speed and how quickly a data is generated.
- An important business aspect especially companies that need fast data flow for more effective operations.
- A good example is search queries that are received within a time period and specific social media posts.

# Velocity



# Veracity

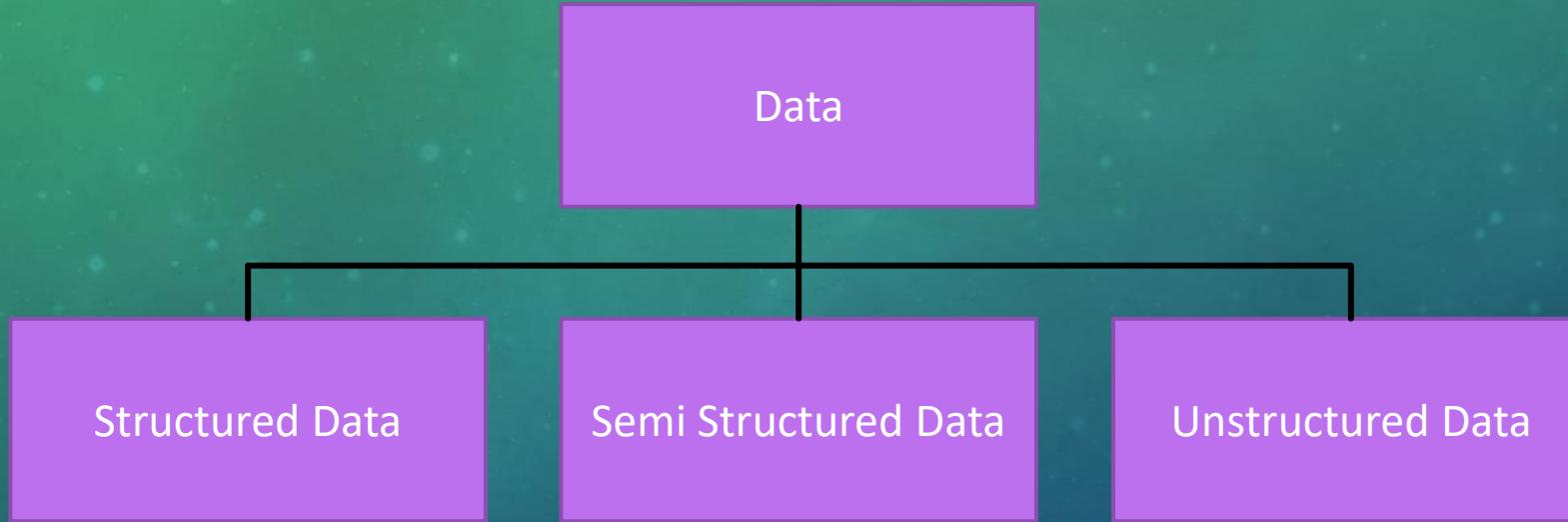
- It refers to the accuracy or the degree of reliability that the data has to offer.
- Big data needs to filter or translate them out as the data is crucial in business developments.



# TYPES OF DATA UNDER BIG DATA

BY JATSEN GESTA

# TYPE OF DATA UNDER BIG DATA



What are the different types of data?

- The big data are categorized into three different types:
  1. Structured Data
  2. Semi Structured Data
  3. Unstructured Data

# STRUCTURED DATA

ID	Company	Last Name	First Name
1	Company A	Bedecs	Anna
2	Company B	Gratacos Solsona	Antonio
3	Company C	Axen	Thomas
4	Company D	Lee	Christina
5	Company E	O'Donnell	Martin
6	Company F	Pérez-Olaeta	Francisco
7	Company G	Xie	Ming-Yang
8	Company H	Andersen	Elizabeth

## Structured Data

Structured Data refers to a data type that is highly organized in a fixed format, both as numerical and alphabetic values. It is displayed in sets of rows and columns, similar to Excel spreadsheets or tables. It is easily stored in a database or warehouse and is simple to query using Structured Query Language (SQL) or ETL tools (Extract, Transform, Load). SQL is the standard language employed to communicate with databases and is effective when managing Structured Data — such as, querying, inserting, updating, and deleting data records.

# UNSTRUCTURED DATA



Image



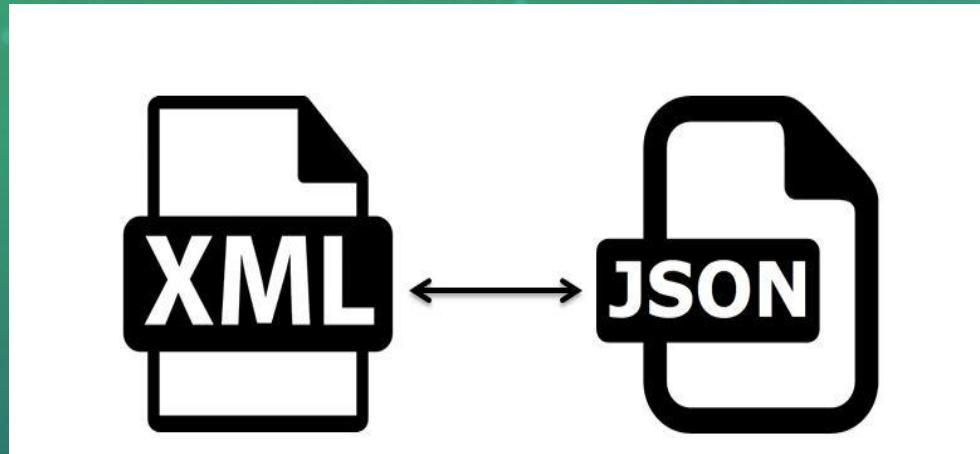
Video

## Unstructured Data

Unstructured Data or also called Qualitative Data refers to **data that does not have any predefined structure**.

Typically, Unstructured Data is text-heavy, such as Social Media conversations and open-ended survey responses, but it also includes audio, videos, and images.

# SEMI STRUCTURED DATA



## Semi-Structured Data

Semi-structured data is a type of data that combines features of both structured data and unstructured data and it also possess qualitative properties that require either translation into a machine language or interpretation by humans. semi-structured data does not conform to relational databases such as Excel or SQL, but nonetheless contains some level of organization through semantic elements like tags. For instance, consider HTML, which does not restrict the amount of information you can collect in a document, but enforces a certain hierarchy

## REFERENCES:

- <https://www.linkedin.com/pulse/structured-data-definition-examples-comparison-cubeware-gmbh/>
- <https://www.indeed.com/career-advice/career-development/semi-structured-data>
- [https://hevodata.com/learn/unstructured-data-definition/#:~:text=Unstructured%20Data%20\(also%20called%20Qualitative,audio%2C%20videos%2C%20and%20images.](https://hevodata.com/learn/unstructured-data-definition/#:~:text=Unstructured%20Data%20(also%20called%20Qualitative,audio%2C%20videos%2C%20and%20images.)

THANK YOU!  
GOD BLESS!



# THE EVOLUTION OF DATA



presentation by  
**MARIA LEILA MAE L.BOCAR**

# What is Data?

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical.



# What is Big Data?

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.



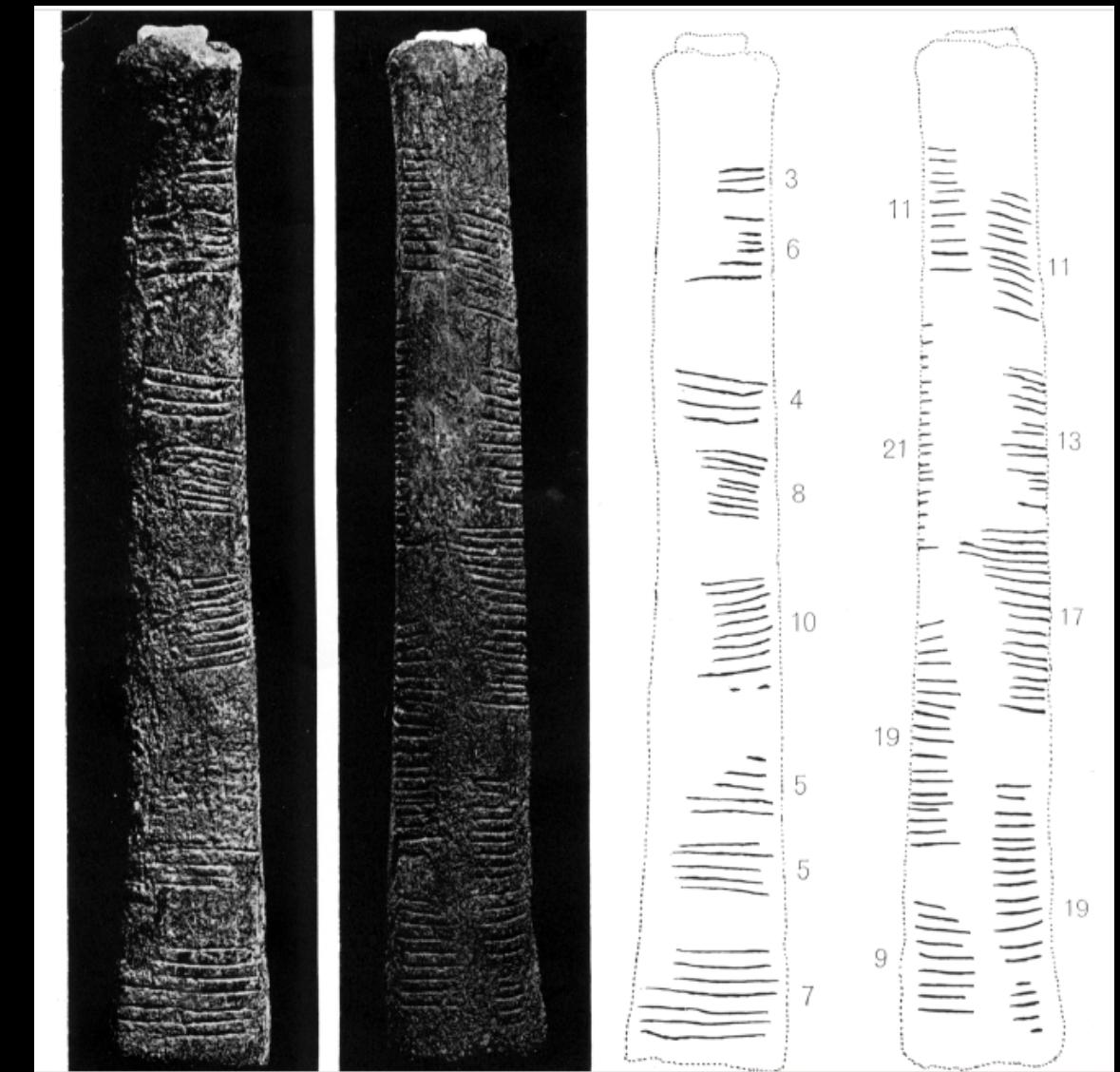
# Evolution of Data



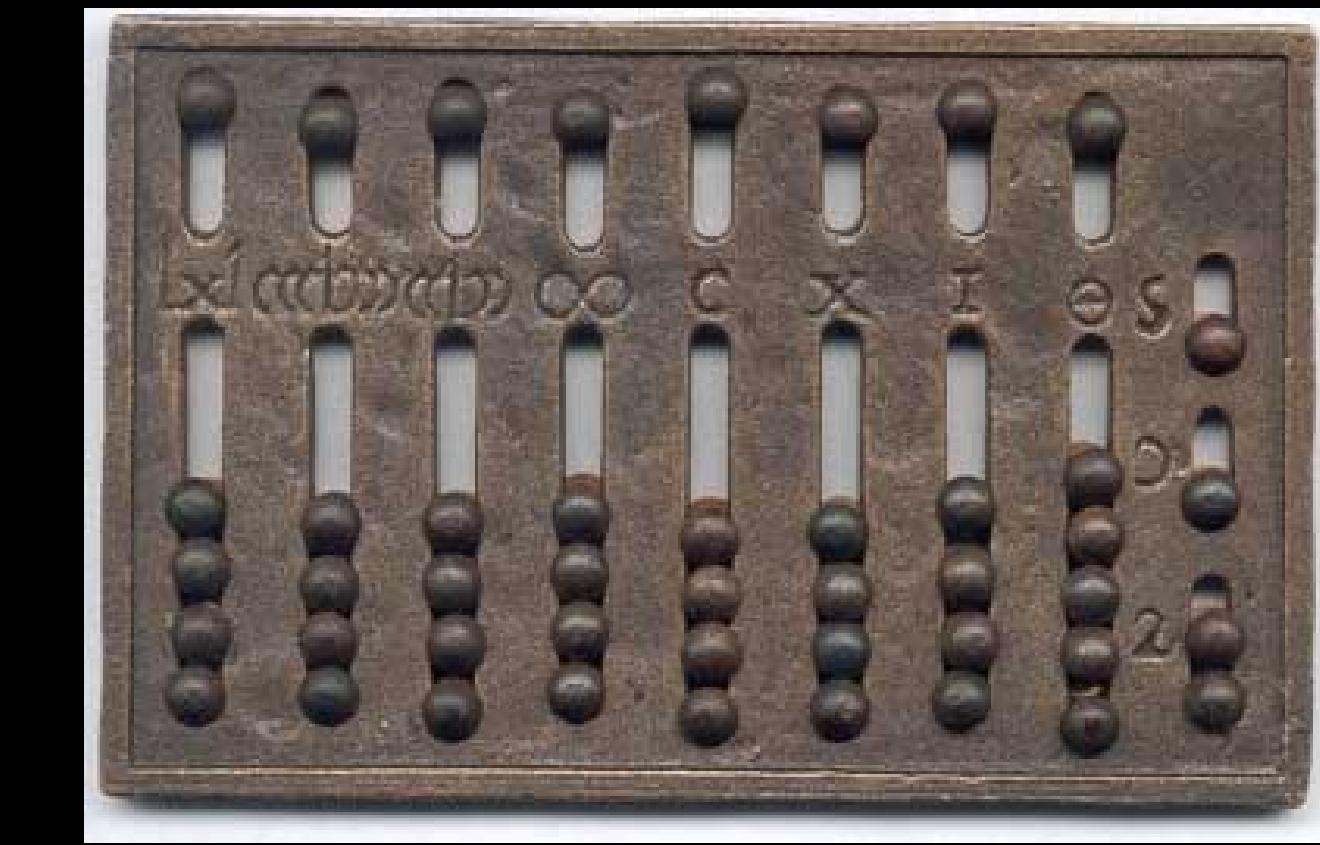
# The ancient history of Big Data

The earliest examples we have of humans storing and analyzing data are the tally sticks, which date back to 18,000 BCE! The Ishango Bone was discovered in 1960 in what is now known as Uganda and is thought to be one of the earliest pieces of evidence of prehistoric data storage.

in 2400 BCE came, the abacus.



# The ancient history of Big Data

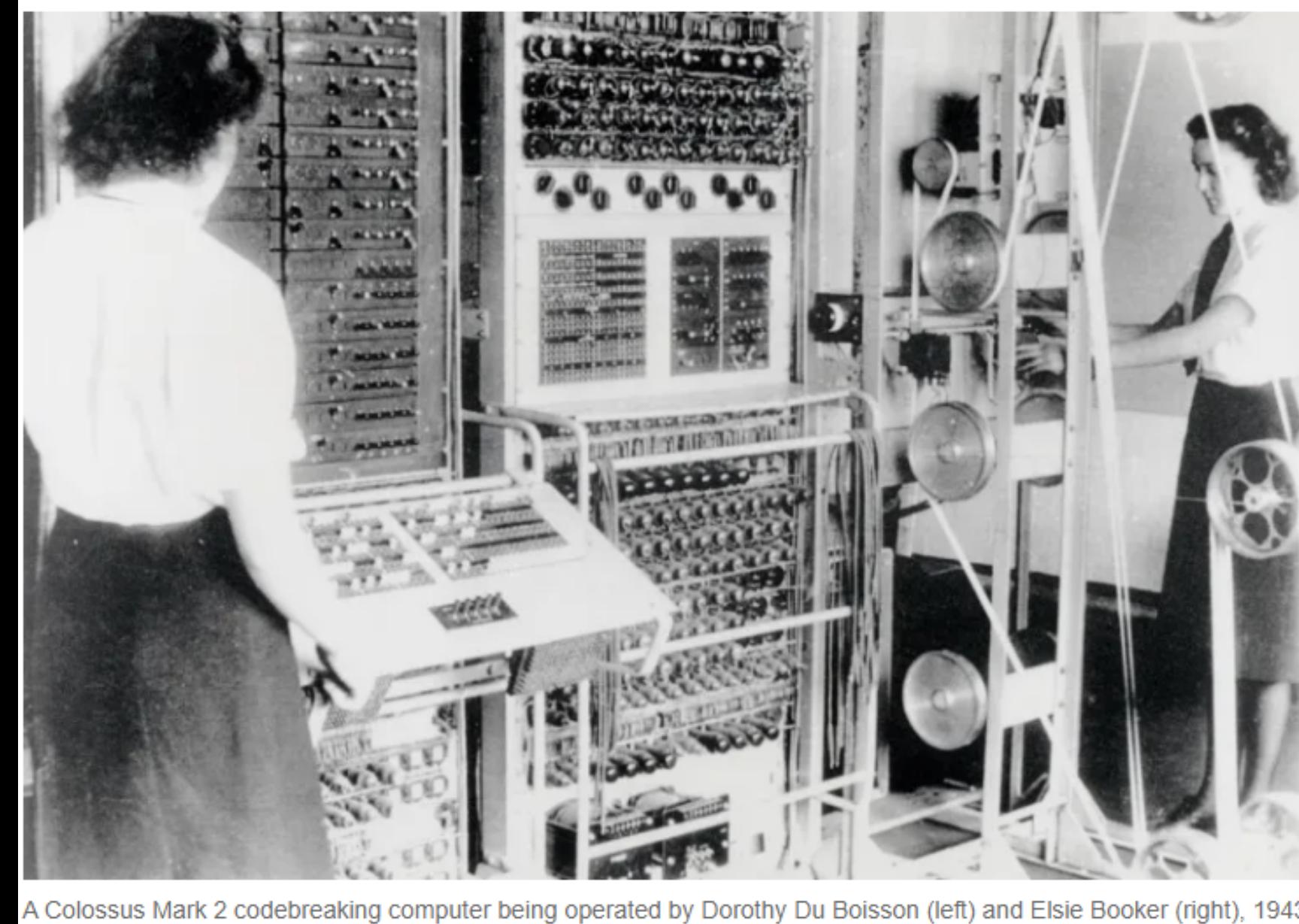


# Big Data in 20th Century

The first major data project was created in 1937 and was ordered by the Franklin D. Roosevelt administration after the Social Security Act became law. The government had to keep track of contributions from 26 million Americans and more than 3 million employers. IBM got the contract to develop punch card-reading machine for this massive bookkeeping project.

The first data-processing machine appeared in 1943 and was developed by the British to decipher Nazi codes during World War II.

# Big Data in 20th Century

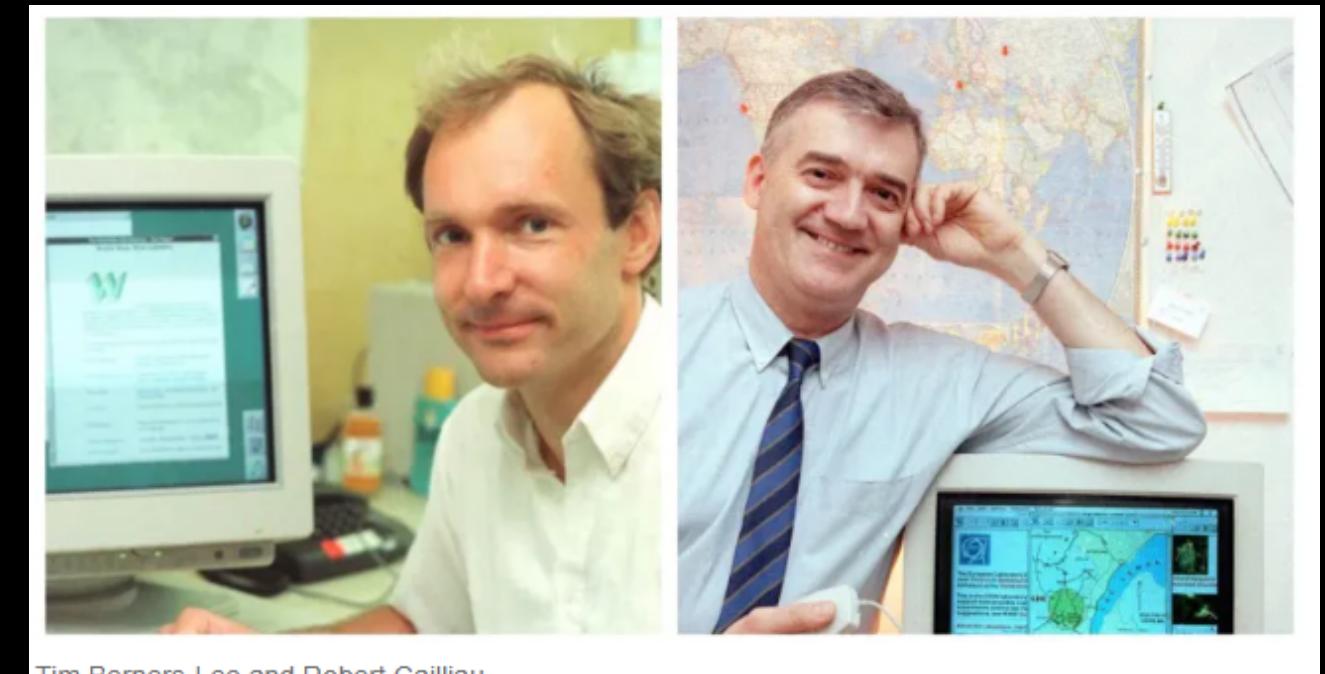


# **Big Data in 20th Century**

Then, in 1965, the United States Government decided to build the first ever data centre to store over 742million tax returns and 175 million sets of fingerprints.

# The Internet age and the dawn of big data

Between 1989 and 1990 Tim Berners-Lee and Robert Cailliau created the World Wide Web and developed HTML, URLs and HTTP, all while working for CERN. The internet age with widespread and easy access to data had begun and by 1996 digital data storage had become more cost-effective than storing information on paper.



# The Internet age and the dawn of big data

In 1998, Carlo Strozzi developed NoSQL, an open-source relational database that provided a way to store and retrieve data modelled differently from the traditional tabular methods found in relational databases.

In 1999, the first edition of *How Much Information* by Hal R. Varian and Peter Lyman attempted to quantify the amount of digital information available in the world at that point.

# The Information ages

Since the early 2000s, the Internet and the Web has offered unique data collections and data analysis opportunities. With the expansion of web traffic and online stores, companies such as Yahoo, Amazon and eBay started to analyze customer behavior by looking at click-rates, IP-specific location data and search logs.

In 2005, Big Data was labelled by Roger Mousalas as he referred to a large set of data that, at the time, was almost impossible to manage and process using the traditional business intelligence tools available. In the same year, Hadoop, which could handle Big Data, was created.

# The Future of Big Data

Since Big Data first entered the scene, its definition, its use cases, technology and strategy of harnessing its value evolved significantly across different industries. Innovations in cloud computing, quantum computing, Internet of Things (IoT), artificial intelligence, and so on will allow for Big Data to evolve further as we'll find new ways of harnessing its potential.



# **COMPANIES WITH BIG DATA**



# 1. Amazon

- Amazon is a well-known e-commerce platform. They store every single piece of information related to their customer as a means of figuring out how customers are spending their money on an individual product.
- All this information is being collected to use in social media advertising algorithms that can be further used to expand customer relations, recommending products, improving customer experience and services, etc.



- For example, you might have noticed that if you wishlist something or add a product to your cart then it recommends some items related to your product or they show items bought together with that product.

This way, Amazon uses big data in leveraging recommendations to facilitate immediate purchases from a customer and also increase the entire shopping experience.



## 2. Apple

- We all know that Apple is an expert in using advanced technology. So, they are using big data technologies and now they become involved in big data analytics, with the technology driving their plenty of decisions.
- The data collected by them is used by the company to consider the best approach towards consumers with its new products and services.
- By using big data, Apple can find how people are using apps in real life and change future designs to fit with customer preferences.



- The two major areas of use of big data by Apple are:
    1. Application design
    2. Keep a track of people's health and improving lifestyle
  - For instance, the Apple watch is not just wearable but also has the ability of data gathering. Now, Apple watches can track what users do during the day.
- Therefore, wearable devices by Apple have great potential with big data and they would want to capitalize on it.



### 3. Google

- Google uses big data to understand what we want from it based on several parameters such as search history, locations, trends, and many more.
- After that, it goes through an algorithm where complex estimations are done and afterward Google easily shows the arranged or positioned indexed lists as far as significance and authority intended to coordinate the users prerequisite.



- Google easily shows the ranked search results in terms of relevance and authority formulated to match the user's requirement.
- Google has acquired some techniques to understand user's requirements like Indexed pages, real-time feeds, sorting tools, knowledge graph pages, literal and semantic search, google translate, etc.



## 4. Spotify

- This company has lots of users and it is fully a data-driven company. They use intel about components like songs' playtime, where they are being streamed, what sort of device is being used for streaming, and when they are being played.
- Thus, all that data is providing the music-tech industry with wonderful insights to affect listeners' experience.



- Spotify has also launched Spotify for Artists, which gives artists and their administration access to data so they can improve their promotion and entity.
- Simply, Spotify uses big data for digitizing the taste of users, developing personalized content, for enhanced marketing through targeted ads, Spotify wrapped, and so on.



## 5. Facebook

- You have noticed how Facebook reminds us of birthdays, friendship anniversaries, but do you ever realize how Facebook does that? This work is done by big data and Facebook uses big data to enhance the experience of their users.
- With the help of big data, Facebook creates a short video that contains our old pictures as a memory. Data plays a very vital role for Facebook as they create flashbacks based on data collected.



- Image recognition is one of the big data technologies adopted by Facebook. It is a technology that advises devices on how to observe the details in a particular picture or video just by directing it through several other images.
- Thus, Facebook assesses every single piece of data and gives you better services each time you log in.



Thank you!