Big Data

Jnaneshwar Bohara

Big Data Overview

- 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**.

Big Data

- Big data is high-**volume**, high-**velocity** and high-**variety** information assets that demand **cost-effective**, **innovative** forms of information processing for **enhanced insight and decision making**.

From Bits to GeopBytes

1024 Bytes	**1 Kilobyte**
1024 Kilobytes	**1 Megabyte**
1024 Megabytes	**1 Gigabyte**
1024 Gigabytes	**1 Terabyte**
1024 Terabytes	**1 Petabyte**
1024 Petabytes	**1 Exabyte**
1024 Exabytes	**1 Zettabyte**
1024 Zettabytes	**1 Yottabyte**
1024 Yottabytes	**1 Brontobyte**
1024 Brontobytes	**1 Geopbyte**

**One geopbyte is 1024^10 or 1267650600228229401496703205376 bytes.
Or simply a 1 followed by 30 digits. (Not zeroes)**

Big Data Statistics

- The big data analytics market is set to reach **$103 billion by 2023.**

- Poor data quality costs the US economy **up to $3.1 trillion yearly**.

- In 2020, **every person generated 1.7 megabytes** in just a second.

- Internet users generate about **2.5 quintillion bytes of data each day.**

Big Data Statistics

- **Google gets over 3.5 billion searches daily.**

- **WhatsApp users exchange up to 65 billion messages daily.**

- **Twitter users send over half a million tweets every minute.**

Big Data Statistics

- **45% of businesses worldwide are running at least one of their Big Data workloads in the cloud.**

- **80-90% of the data we generate today is unstructured.**

- **Data creation will grow to more than 180 zettabytes by 2025.**

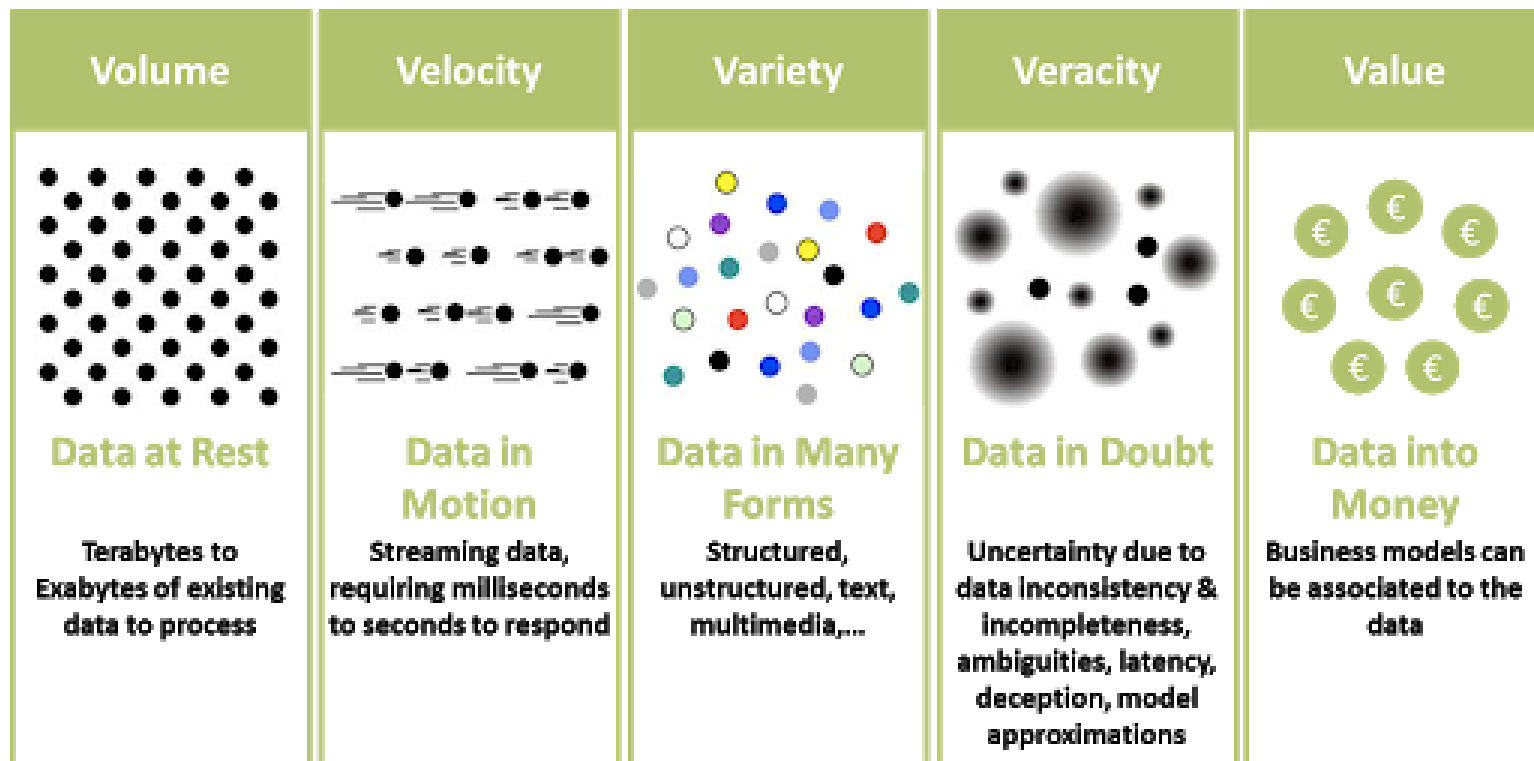- **Internet users generate about 2.5 quintillion bytes of data each day.**

Big Data Statistics

- **95% of businesses** cite the need to manage unstructured data as a problem for their business.

- **97.2% of organizations** are investing in big data and AI.

- Using big data, **Netflix saves $1 billion per year** on customer retention.

- In 2020, the big data market grew by 14%.

Characteristics of Big Data

- **Volume** - Data at rest (too big)
- **Variety -** Data in many forms (too complex)
- **Velocity -**Data in motion(too fast)
- **Veracity** - Data in doubt(uncertainty)
- **Value** - Data into money

Characteristics of Big Data

Volume	Velocity	Variety	Veracity	Value
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**	**Data into Money**
Terabytes to Exabytes of existing data to process	Streaming data, requiring milliseconds to seconds to respond	Structured, unstructured, text, multimedia,...	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations	Business models can be associated to the data

Adapted by a post of Michael Walker on 28 November 2012

Volume: Scale of Data

- Refers to the vast amounts of data generated every second.

- We are not talking Terabytes but Brontobytes or Geopbytes.

- If we take all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute.

Variety: Different Forms of Data

- This refers to the different types of data we can now use.

- In the past we focused on structured data that fits neatly into tables or relational databases, such as financial data.

- In fact, 80% of the world's data is unstructured (text, images, video, voice, etc.)

- Big data technology means we can now analyse and bring together data of different types such as messages, social media conversations, photos, sensor data, video or voice recordings.

Velocity: Analysis of Streaming Data

- Refers to the speed at which new data is generated and the speed at which data moves around.

- Just think of social media messages going viral in seconds.

- Technology allows us now to analyse the data while it is being generated (in-memory analytics), without ever putting it into databases.

Veracity: Uncertainty of Data

- Refers to the messiness or trustworthiness of the data.
- With many forms of big data, quality and accuracy are less controllable
- Big data and analytics technology now allows us to work with these type of data.

Value: Turning Big Data into Value

- Having access to big data is no good unless we can turn it into value.

- Companies are starting to generate amazing value from their big data.

Big Data Trends 2022

- As A Field, Predictive Analytics Will Grow

- The AI Market Will Reach A Record-Breaking High With Regards To Implementation And Usage

- Self-Service Analytics Will Become Even More Critical To Business Intelligence

Big Data Trends 2022

- Wholesale/Retail, Technology Organizations, And Financial Services Will Increase Their BI Budgets Over 50% By The Year 2022

- Big data will help climate change research

- Big data to search novel medical cures

Big Data Applications

- **Healthcare**
- **Manufacturing**
- **Media & Entertainment**
- **Internet of Things (IoT)**
- **Government**
- **Ecommerce**
- **Disaster Management**
- **Digital Marketing**
- **Telecommunication**
- **Retail Industry**
- **Finance**
- **Education**

Some Big Data Use Cases

- Credit Card Fraud Detection
- Sentiment Analysis
- Delivering personalized customer experience
- Preventing customer churn
- Genomic Research

How is Big Data Used

- Understanding and Targeting Customers
- Understanding and Optimizing Business Processes
- Improving Sports Performance
- Improving and Optimizing Cities and Countries
- Improving Healthcare and Public Health

https://www.bernardmarr.com/default.asp?contentID=1076

Big Data in Practice
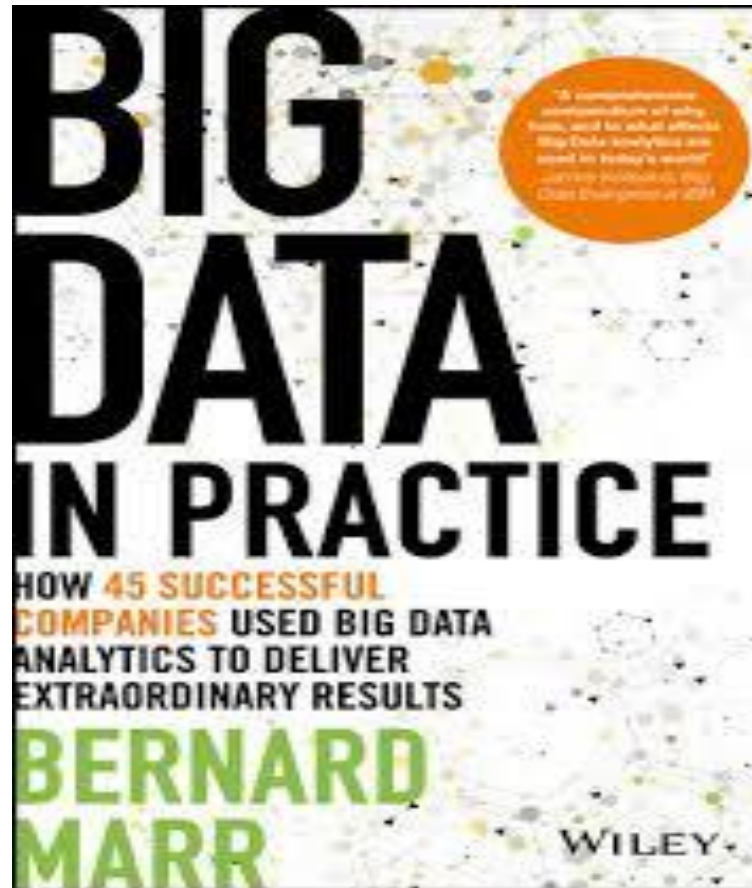
Big Data in COVID-19 Pandemic

- Identification of infected cases
  - It is capable of storing the complete medical history of all patients, due to its capability of storing a massive amount of data
  - By providing the captured data, this technology helps in identification of the infected cases and undertake further analysis of the level of risks

Big Data in COVID-19 Pandemic

- Travel history
  - Used to store the travel history of the people to analyze the risk
  - Helps to identify people who may be in contact with the infected patient of this virus

Big Data in COVID-19 Pandemic

- Fever symptoms
  - Big data can keep the record of fever and other symptoms of a patient and suggest if medical attention is required
  - Helps to identify the suspicious cases and other misinformation with the appropriate data

Big Data in COVID-19 Pandemic

- Identification of the virus at an early stage
  - Quickly helps to identify the infected patient at an early stage
  - Helps to analyze and identify persons who can be infected by this virus in future

Big Data in COVID-19 Pandemic

- Identification and analysis of fast-moving disease
  - Helps to effectively analyze the fast-moving disease as efficiently as possible
  - Potential to handle appropriate information regarding the disease

Big Data in COVID-19 Pandemic

- Information during lockdown
    - This technology collects information regarding this virus during the lockdown
    - Track and monitor the movement of people and entire health management

Big Data in COVID-19 Pandemic

- People entered or leaving the affected area
  - It helps to analyze the number of people entered or leaving from the affected city
  - With these vast amount of data, health specialist can quickly identify the chances of the virus in those peoples

Big Data in COVID-19 Pandemic

- Faster development of medical treatments

  - Assist in fast-tracking the development of new medicines and equipment needed for current and future medicinal needs

  - Provides previous data of virus inhabited or spread and, thus, helps in gaining a giving advantage over newer pandemic/epidemic with previously analyzed results

  https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7204193/

Role Of Big Data In The Fight Against COVID-19

- China's Surveillance Infrastructure Used to Track Exposed People

- Mobile App for Contact Tracing

- Official Dashboards Track the Virus and Outbreak Analytics

- Big Data Analytics and Successes in Taiwan

https://www.linkedin.com/pulse/vital-role-big-data-fight-against-covid-19-coronavirus-bernard-marr

Scope of Big Data

- Increasing demand for Data Analytics
- Increasing enterprise adoption of Big Data
- Big Data finds application across various parallels of the industry
- Huge Job Opportunities & Meeting the Skill Gap
- Promises exponential salary growth
- Key Decision-Making Power

Challenges of Big Data

- Dealing with data growth
- Generating insights in a timely manner
- Recruiting and retaining big data talent
- Integrating disparate data sources
- Securing big data
- Organizational resistance

Thank You !