6CS030 :Big Data

Lecture 1 : Big Data Overview

by Prakriti Regmi

Module Overview

ASSESSMENT	DESCRIPTION	% COVERAGE
Coursework	Implementation based research work involving handling and analysing of big data in your interest specific subject domain.	Report (80%) Code (20%) out of total 70%
Timed Constrained Assessment.	MCQ Exam based on Big Data subjects covered over the weeks.	30%

Module Team

MODULE LEADER
ASHISH ACHARYA

LECTURER
PRAKRITI REGMI

LECTURER
PRAKRITI REGMI

LECTURER
ANIL POUDEL

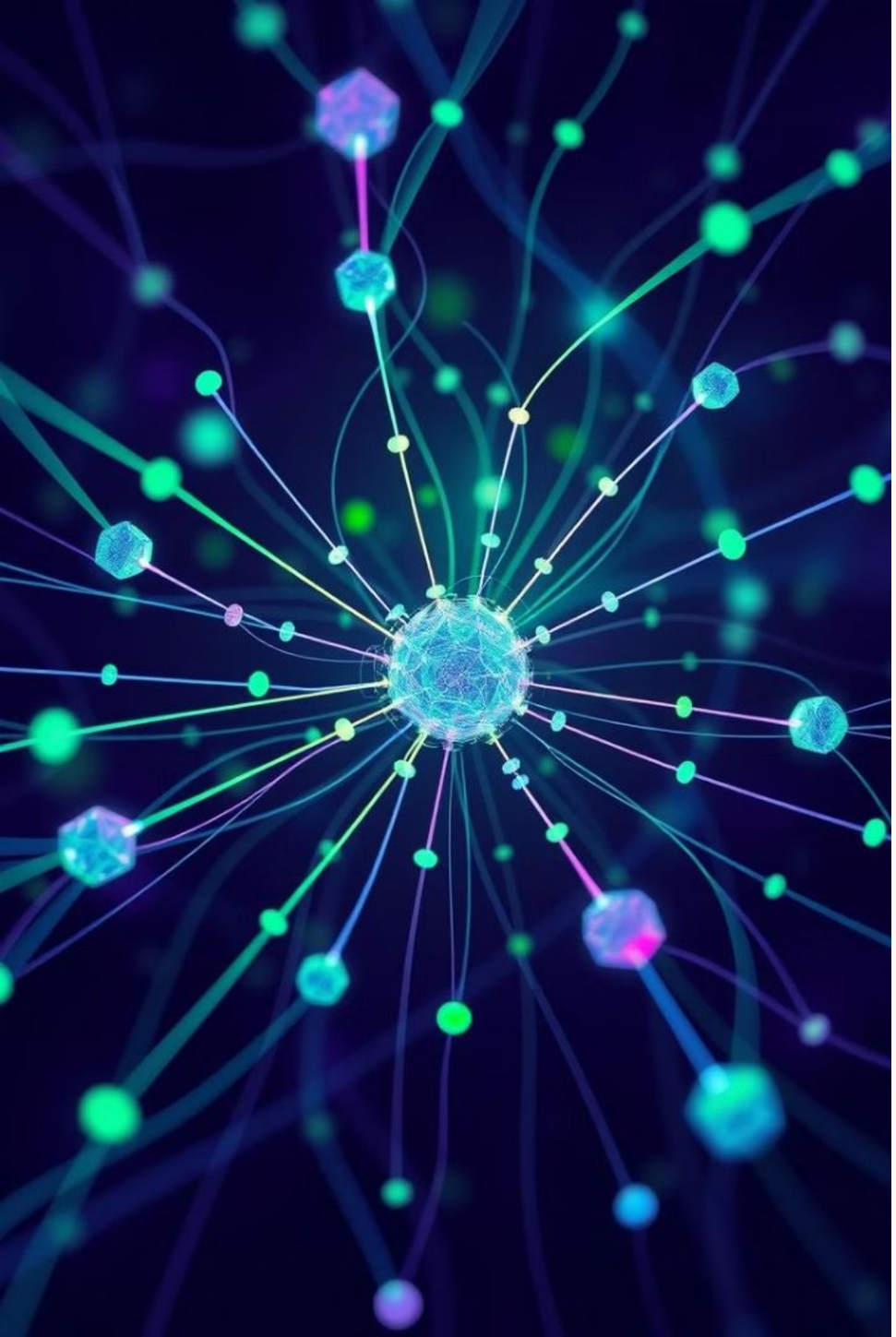
TUTOR
GUNJAN KUMAR MISHRA

TUTOR
SANDESH LAMSAL

GTA
SIDDHARTHA SHAKYA

What is Big Data?

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 Characteristics : **The Vs**

1

Volume

Data at rest (too big).

2

Variety

Data in many forms (too complex).

3

Velocity

Data in motion (too fast).

4

Veracity

Data in doubt (uncertainty).

Big Data Scale Breakdown

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 by the Numbers

$862.31B

Market Size

Big data analytics market by 2030.

$3.1T

Poor Quality Costs

Yearly cost to the US economy.

2.5Q

Daily Bytes

Data generated by internet users.

Big Data Trends 2022

1

Predictive Analytics

The field will continue to grow.

2

AI Market

Will reach a record high in implementation.

3

Self-Service Analytics

Will become even more critical.

Big Data Statistics

3.5B+ Daily Searches

on Google

65B Daily Messages

exchanged on WhatsApp

500K+ Tweets/Minute

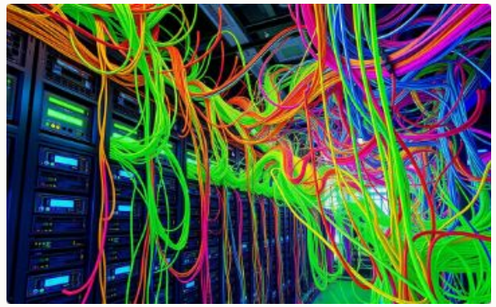
sent by Twitter users

45% Businesses in Cloud

running Big Data workloads

80-90% Unstructured Data

generated today



180+ Zettabytes by 2025

data creation forecast

2.5 Quintillion Bytes/Day

generated by users

More on Big Data Statistics

95%

Unstructured Data Problem

Businesses cite this as a problem

97.2%

Investment in Big Data & AI

Organizations are investing

$1B

Netflix Savings

Per year on customer retention

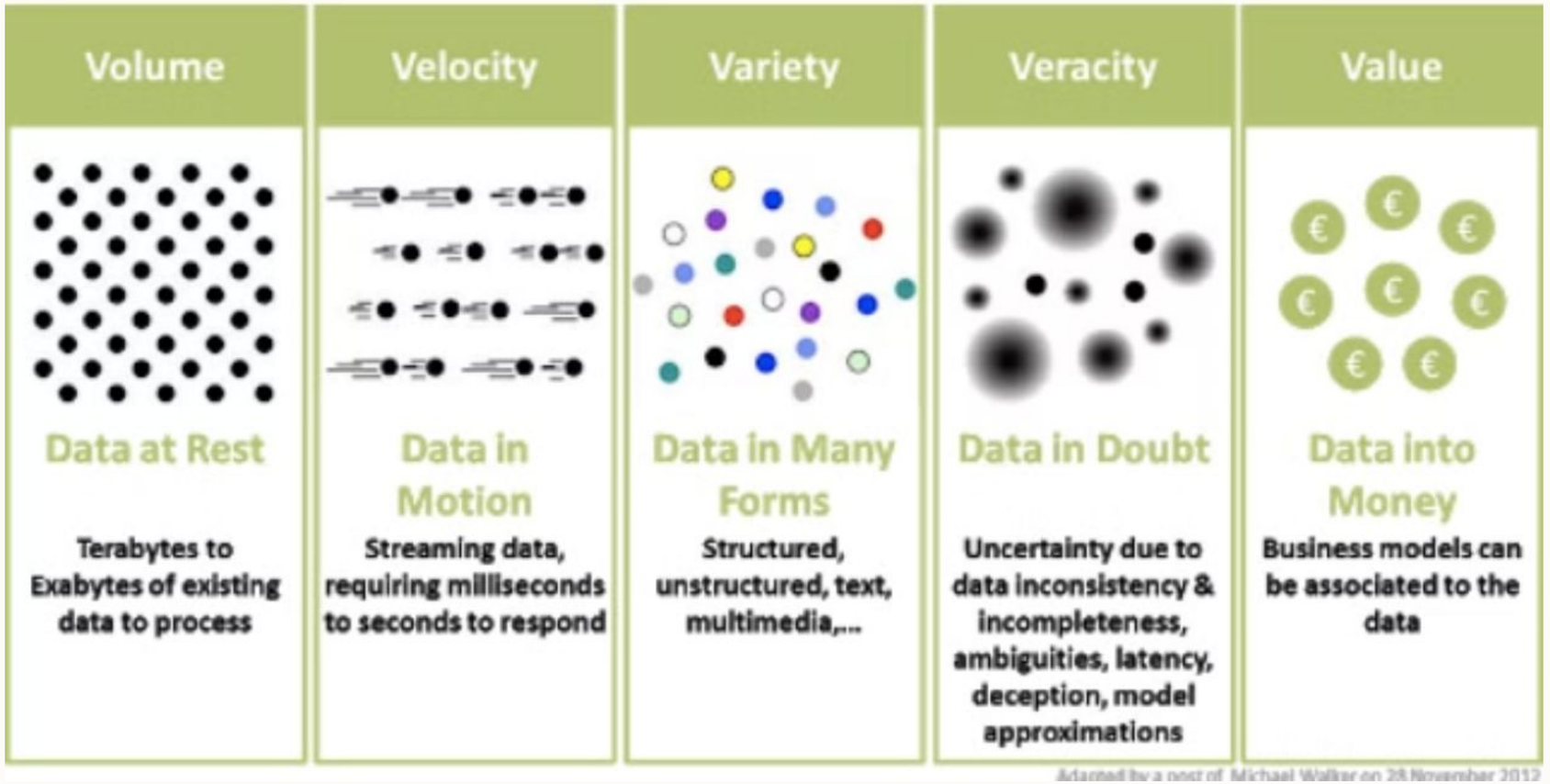
14%

Big Data Market Growth

In 2020

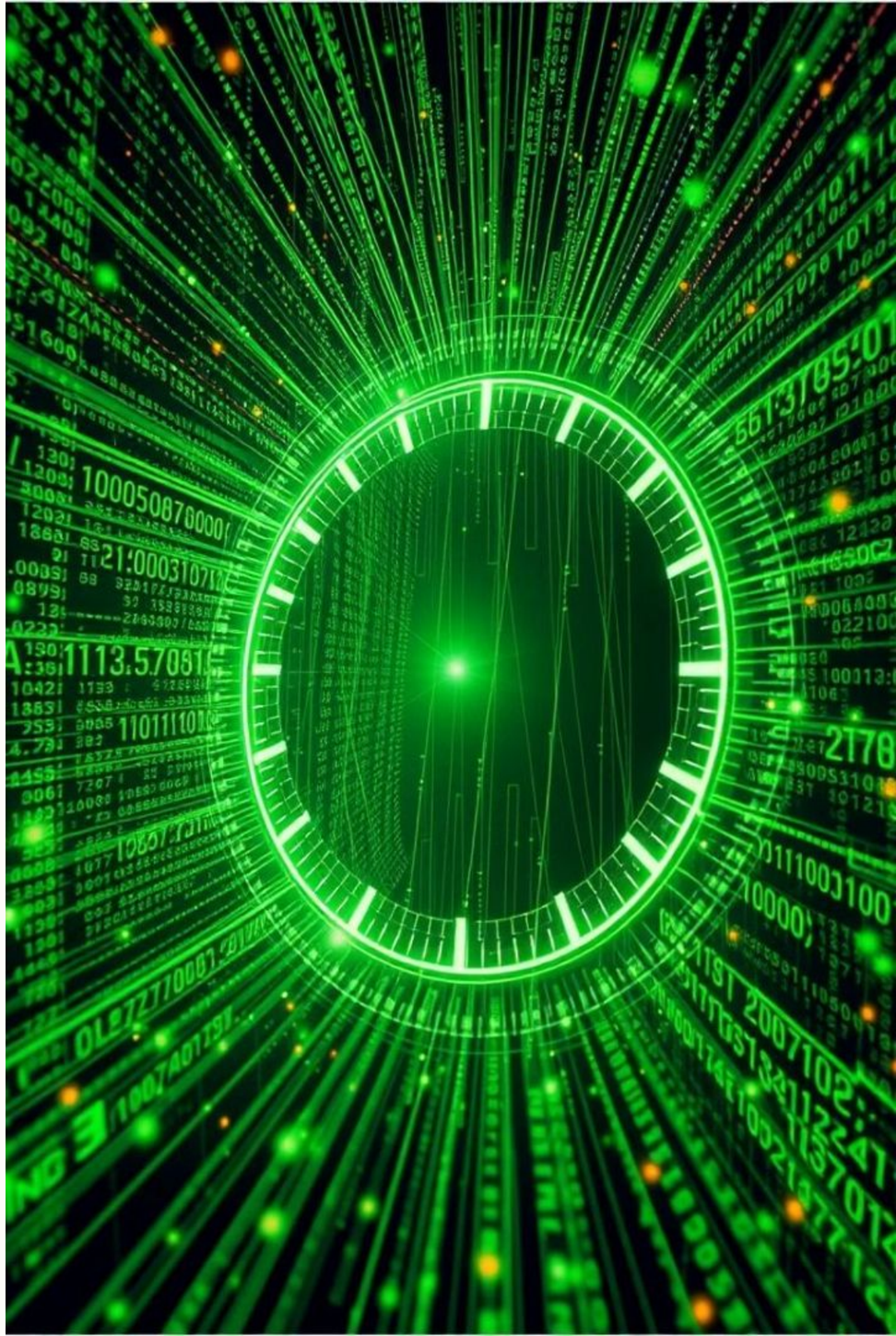
What makes a Data

Big?



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

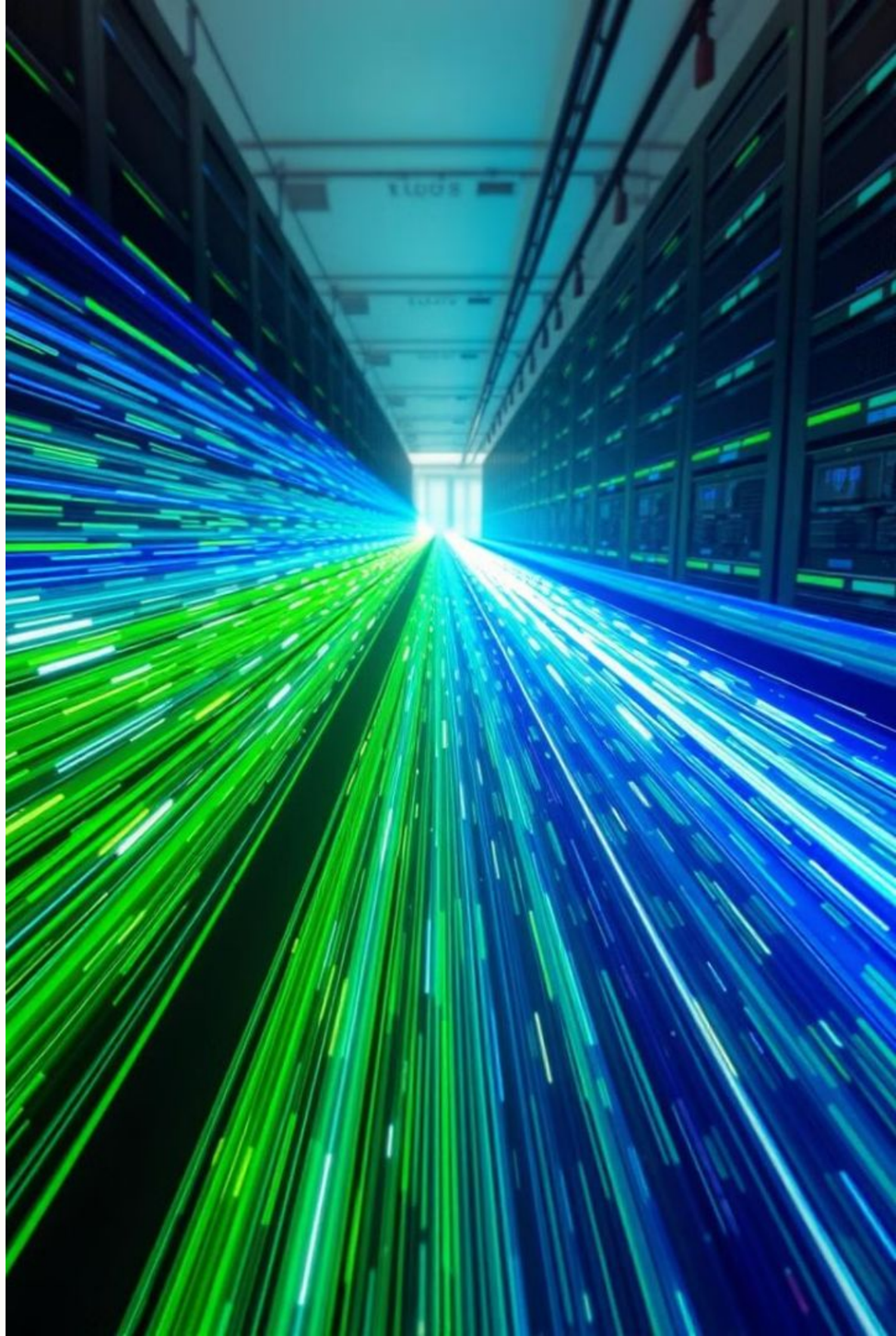
Refers to the **different types of data** we can now use.

- 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.

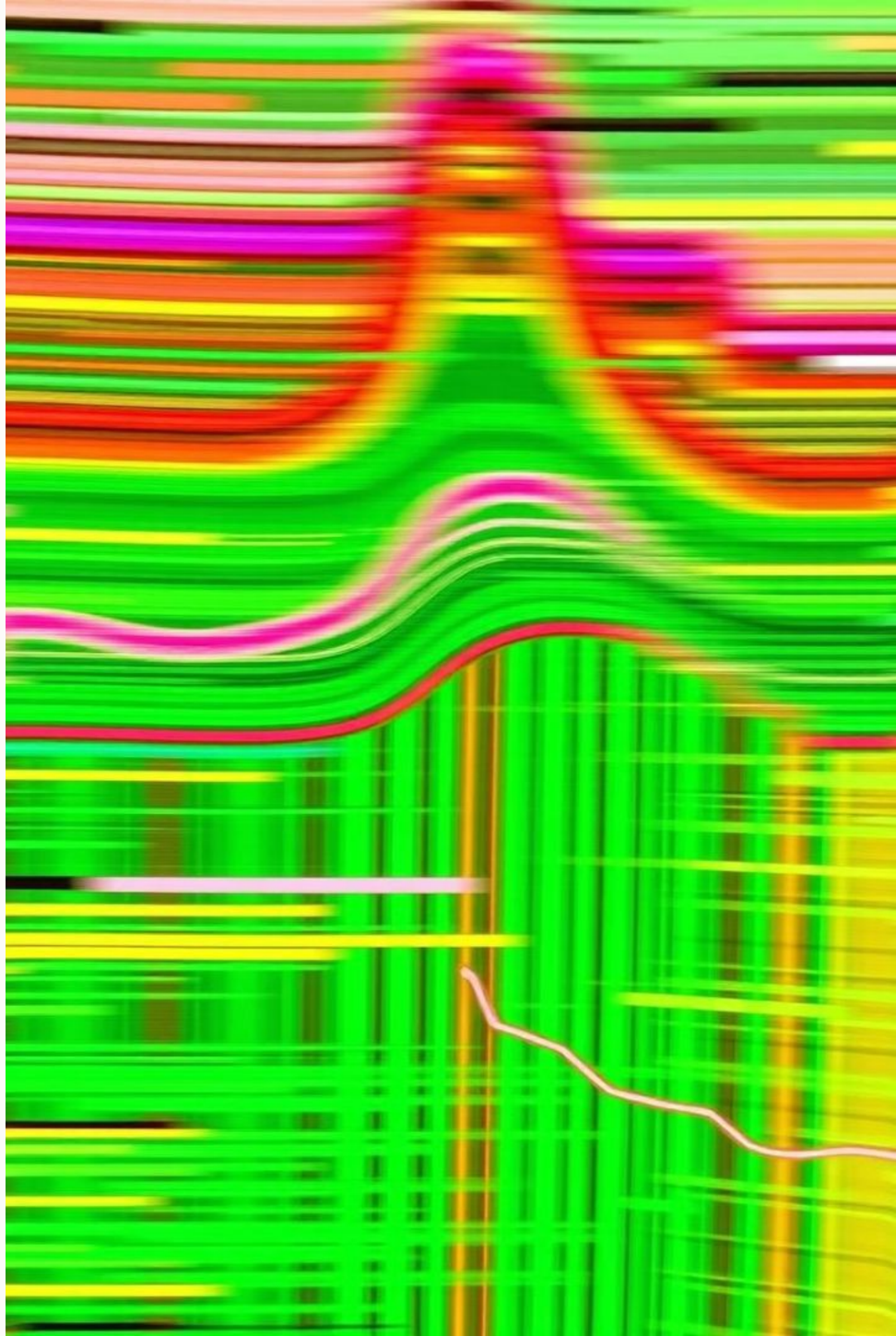
- Just think of social media messages going viral in seconds.
- Technology allows us now to analyze the data while it is being generated (in- memory analytics), without ever putting it into databases.



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 Applications

> Big Data is increasingly impacting various sectors. Key trends include its growing influence in healthcare, manufacturing, media, and the Internet of Things (IoT).

Healthcare

Manufacturing

Media

IoT

Big Data in COVID-19

Infection Identification and Risk Analysis

- Identifies infected cases by storing complete medical histories.
- Analyzes travel history to identify potential contacts and risk levels.
- Records fever symptoms to suggest necessary medical attention.
- Helps identify suspicious cases and misinformation.

Early Virus Detection and Spread Prevention

- Quickly identifies infected patients at an early stage.
- Analyzes and identifies individuals at future risk of infection.
- Tracks and monitors movement of people during lockdowns.
- Analyzes the number of people entering or leaving affected areas.

Medical Treatment and Information Management

- Effectively analyzes fast-moving diseases.
- Handles appropriate disease information.
- Collects virus information during lockdowns for health management.

Assists in fast-tracking the development of new medicines and equipment

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

Role Of Big Data In The Fight Against COVID-19

1

China’s Surveillance Infrastructure

Used to Track Exposed People

2

Mobile App

For Contact Tracing

3

Official Dashboards

Track the Virus and Outbreak Analytics

4

Big Data Analytics

Successes in Taiwan

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

Scope of Big Data

1

Increasing Demand for Data Analytics

The field needs more skilled data analysts.

2

Increasing Enterprise Adoption

More businesses are using Big Data.

3

Wide Application Across Industries

Big Data is relevant in almost every sector.

4

Huge Job Opportunities & Skill Gap

There are significant career prospects, but specific skills are required.

Challenges of Big Data

Data Growth

Dealing with ever-increasing data volumes.

Timely Insights

Generating insights rapidly.

Talent Acquisition

Recruiting and retaining skilled big data professionals.

Data Integration

Integrating diverse data sources.

Data Security

Securing large and sensitive datasets.

Organizational Resistance

Overcoming resistance to big data initiatives.

Thank You