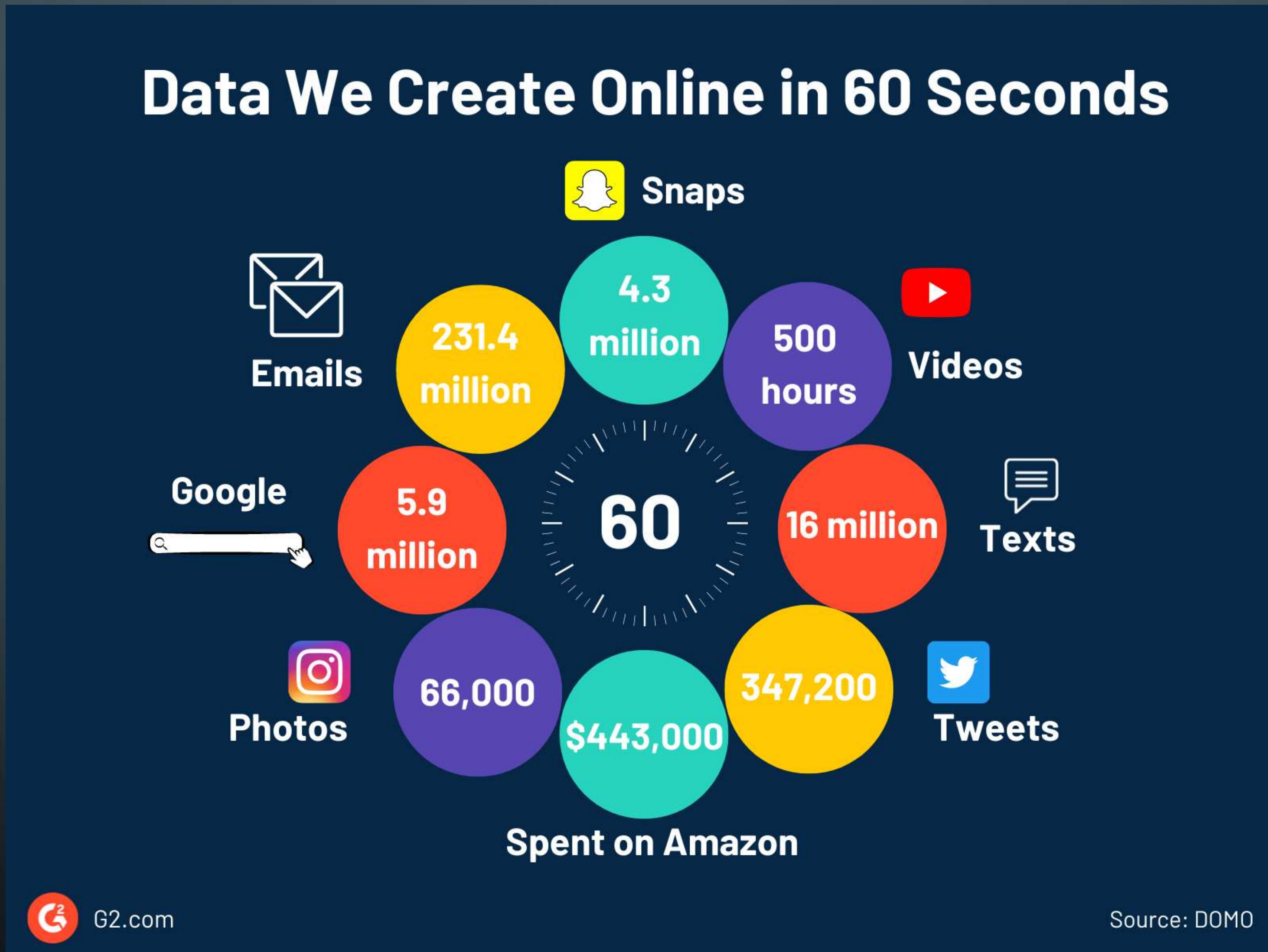
BIG DATA ANALYTICS

DR. OWAIS BHAT

Lecture – 7, 8

INTRODUCTION TO BIG DATA

- Big data refers to extremely large and complex datasets that exceed the capabilities of traditional data processing and management tools.
- Big data is generated by a variety of sources, including social media, sensors, and machines.
- Big data is growing exponentially, and it is estimated that the amount of data created in the next two years will be more than the amount of data created in the previous 50 years combined.


Data We Create Online in 60 Seconds
Snaps
4.3 million
Emails
231.4 million
Videos
500 hours
Google
5.9 million
60
Texts
16 million
Photos
66,000
$443,000
Spent on Amazon
Tweets
347,200
G2.com
Source: DOMO

CHARACTERISTICS

The term "big data" is often characterized by the "3Vs": Volume, Velocity, and Variety.

Volume

- Big data involves massive amounts of data that exceed the capacity of conventional databases and storage systems.
- The volume of data can range from terabytes to petabytes or even exabytes, generated from various sources such as social media, sensors, devices, and business transactions.

Velocity

- Big data is generated at high speeds and requires real-time or near-real-time processing and analysis.
- The data is continuously generated, streamed, or collected in rapid intervals, demanding efficient data ingestion, processing, and decision-making to capture its value.

Variety

- Structured data refers to well-organized data with a predefined schema, such as data in relational databases.
- Semi-structured data has some organization or metadata but does not adhere to a rigid schema, like JSON or XML documents.
- Unstructured data lacks a predefined structure and includes text documents, social media posts, images, videos, and sensor data.
- Big data comes from diverse sources, including social media, web logs, sensors, devices, financial systems, healthcare records, and more.

Other V's

Veracity

- Veracity refers to the quality, reliability, and trustworthiness of big data.
- Big data may contain noise, errors, or inconsistencies, making it challenging to ensure data accuracy and reliability.
- Verifying the authenticity, integrity, and quality of the data becomes crucial to derive meaningful insights and make informed decisions.

Value

- The ultimate goal of working with big data is to extract value and actionable insights from the vast amounts of data.
- By leveraging advanced analytics techniques, machine learning, and artificial intelligence, organizations can uncover patterns, correlations, and trends that were previously hidden.

IMPACT

Data Analytics are designed to explore and leverage unique data characteristics, from sequential/temporal mining and spatial mining, to data mining for high-speed data streams and sensor data.

- Analytics are formulated based on strong mathematical techniques including statistical machine learning, Bayesian networks, hidden Markov models, support vector machine, reinforcement learning and ensemble models.

Text Analytics aims at event detection, trend following, sentiment analysis, topic modelling, question-answering and opinion mining.

- Text analytics take the help of several well-researched natural language processing techniques in parsing and understanding texts.

Web Analytics aim to leverage internet-based services based on server virtualisation, scheduling, QoS monitoring, infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS) and service-level agreement monitoring, service check pointing and recovery.

Network Analytics on social networks look for link prediction, topic detection, finding influencing node, sentiment analysis, hate monger nodes and monitoring of special activities of business and security concerns.

Mobile Analytics are offered as apps on cell phones, Some of these analytics can predict presence of a person at a place at a given time, possible co-occurrence and prediction of mobility of a person. It can also perform locational service search along with mobility.

BENEFITS

- **Improved decision-making**: Big data can help businesses to make better decisions by providing them with a more complete and accurate view of their customers, operations, and markets.
- **Increased efficiency**: Big data can help businesses to improve their efficiency by automating tasks and identifying areas where costs can be reduced.
- **New opportunities:** Big data can help businesses to identify new opportunities, such as new products and services, new markets, and new ways to improve customer service.

STAGES OF ANALYTICAL EVOLUTION – BIG DATA

- The analytical evolution of big data can be categorized into three stages: Descriptive Analytics, Predictive Analytics, and Prescriptive Analytics.
- These stages represent the progression in analytical capabilities and the increasing sophistication of data-driven decision-making.

Descriptive Analytics

- Descriptive analytics focuses on understanding historical data and providing insights into what has happened.
- It involves collecting, organizing, and summarizing data to gain a retrospective view of past events and trends.
- Descriptive analytics techniques include data aggregation, data visualization, dashboards, and reporting.
- The goal is to provide a clear understanding of historical performance, identify patterns, and answer basic questions about the data.

Predictive Analytics

- Predictive analytics aims to forecast future outcomes or behaviors based on historical data patterns and statistical models.
- It leverages advanced statistical and machine learning algorithms to uncover relationships and patterns in the data.
- Predictive analytics techniques include regression analysis, time series forecasting, data mining, and machine learning algorithms.
- The objective is to generate predictions, probabilities, or likelihoods of future events to support proactive decision-making and planning.

Prescriptive Analytics

- Prescriptive analytics goes beyond predicting future outcomes and provides recommendations on the best course of action to achieve desired outcomes.
- It combines historical data, predictive models, optimization techniques, and decision science to determine the optimal decisions or actions.
- Prescriptive analytics techniques include optimization algorithms, simulation, decision trees, and prescriptive modeling.
- The goal is to offer actionable insights, scenarios, or recommendations to guide decision-making, resource allocation, and process optimization.

STATE OF PRACTICE IN ANALYTICS

- **Analytics** refers to the use of data, statistical analysis, and other quantitative methods to derive insights and inform decision-making.
- The state of the practice of analytics is constantly evolving as new technologies and techniques emerge, and organizations seek to leverage data to gain a competitive advantage.

APPLICATIONS OF ANALYTICS

- **Business analytics:** This involves the use of data to gain insights into business operations, such as sales, marketing, and finance. For example, a retailer can use analytics to identify trends in customer behavior, optimize pricing strategies, and forecast demand for products.
- **Social media analytics:** This involves the use of data from social media platforms to understand customer behavior, sentiment, and engagement. For example, a company can use analytics to track brand mentions, analyze customer feedback, and identify influencers who can help promote their products or services.

- **Predictive maintenance:** This involves the use of analytics to predict when equipment or machinery is likely to fail, allowing for proactive maintenance and reducing downtime. For example, a manufacturer can use analytics to monitor equipment performance, detect anomalies, and schedule maintenance before a breakdown occurs.
- **Fraud detection and prevention:** This involves the use of analytics to identify and prevent fraudulent activities, such as credit card fraud and insurance fraud. For example, a financial institution can use analytics to detect patterns of suspicious activity and flag potentially fraudulent transactions.

Applications of Analytics

CHALLENGES

key challenges that organizations face when implementing big data analytics:

- Data quality: One of the biggest challenges is ensuring that the data is of high quality. This means that the data is accurate, complete, and consistent.
- Data security: Big data analytics often involves sensitive data, such as customer information or financial data. It is important to protect this data from unauthorized access.
- Skill shortages: There is a shortage of skilled professionals who can implement and manage big data analytics projects.
- Cost: Big data analytics can be expensive, especially for large organizations.

SPACE OF BIG DATA - TRANSACTIONS, INTERACTIONS, OBSERVATIONS

- **Transactions** are data points that represent a single event, such as a purchase, a login, or a social media post. Transactions are typically structured data that can be easily stored and analyzed.

- **Interactions** are data points that represent a relationship between two or more entities, such as a user and a product, a user and a user, or a user and a location. Interactions are typically unstructured data that can be more difficult to store and analyze.

- **Observations** are data points that represent a measurement of an entity, such as a user's heart rate, a product's temperature, or a location's humidity. Observations are typically semi-structured data that can be stored and analyzed using a variety of methods.

Here are some examples of how big data is being used in each of these categories:

Transactions

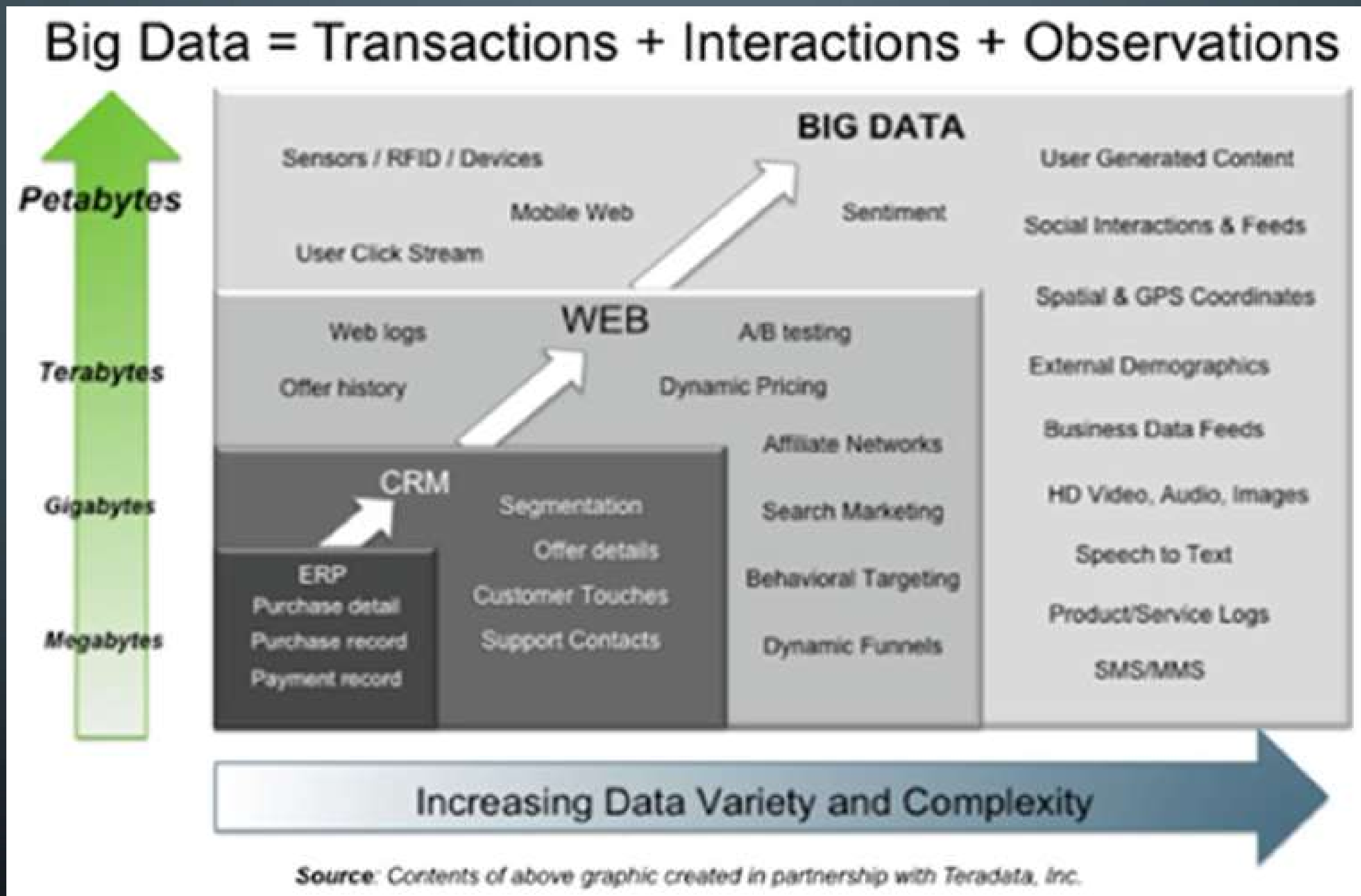
- Retailers use transaction data to track sales, identify trends, and make better decisions about inventory, pricing, and marketing.
- Banks use transaction data to detect fraud, prevent money laundering, and personalize customer experiences.
- Governments use transaction data to track tax evasion, monitor public spending, and improve public services.

Interactions

- Social media platforms use interaction data to understand user behavior, target advertising, and recommend content.
- E-commerce websites use interaction data to personalize product recommendations, improve customer service, and prevent churn.
- Airlines use interaction data to improve flight schedules, optimize staffing levels, and predict customer demand.

Observations

- Hospitals use observation data to monitor patient health, track the spread of disease, and develop new treatments.
- Environmental scientists use observation data to track climate change, monitor air quality, and predict natural disasters.
- Farmers use observation data to monitor crop health, optimize irrigation schedules, and predict crop yields.


Big Data = Transactions + Interactions + Observations
Petabytes
Terabytes
Gigabytes
Megabytes
BIG DATA
Sensors / RFID / Devices
Mobile Web
User Click Stream
Sentiment
User Generated Content
Social Interactions & Feeds
Spatial & GPS Coordinates
External Demographics
Business Data Feeds
HD Video, Audio, Images
Speech to Text
Product/Service Logs
SMS/MMS
WEB
Web logs
Offer history
A/B testing
Dynamic Pricing
Affiliate Networks
Search Marketing
Behavioral Targeting
Dynamic Funnels
CRM
Segmentation
Offer details
Customer Touches
Support Contacts
ERP
Purchase detail
Purchase record
Payment record
Increasing Data Variety and Complexity
Source: Contents of above graphic created in partnership with Teradata, Inc.

POTENTIAL USE CASES OF BIG DATA

- Customer insights: Big data can be used to gain insights into customer behavior, such as what products they buy, how often they shop, and where they live. This information can be used to improve customer service, target marketing campaigns, and develop new products.
- Fraud detection: Big data can be used to detect fraudulent activity, such as credit card fraud and insurance fraud. This information can be used to protect businesses and consumers from financial losses.
- Risk assessment: Big data can be used to assess risk, such as the risk of a natural disaster or the risk of a customer defaulting on a loan. This information can be used to make better decisions about how to allocate resources and manage risk.

- Healthcare: Big data can be used to improve healthcare, such as by tracking the spread of disease, identifying patients at risk, and developing new treatments.
- Transportation: Big data can be used to improve transportation, such as by optimizing traffic flow, predicting demand, and improving safety.
- Environment: Big data can be used to study the environment, such as by tracking climate change, monitoring air quality, and predicting natural disasters.

THANK YOU

owais@iust.ac.in