

Business Intelligence

Week 11

Big Data and Emerging Trends

- Introduction
- Where it did come from?
- Challenges and applications of Big Data
- Characteristics of Big Data i.e Volume, Velocity, Variety and more V's

Tilahun Melak(PhD)

May, 2026



Objectives

At the end of this lecture students will be able to :

- Identify and describe the major sources of big data across different domains.
- Explain the major challenges associated with big data.
- Explain the characteristics of big data, including volume, velocity, variety.

Facts and Figures

- **Walmart** handles 1 million customer transactions per hour.
- **Facebook** handles 40 billion photos from its user base.
- **Facebook** inserts 500 terabytes of new data every day.
- **Facebook** stores, accesses, analyzes, 30+ petabytes of user-generated data, every day.
- **A flight generates** 240 terabytes of flight data in 6 to 8 hours of flight.
- **More than 5 billion people**, are calling, texting, tweeting, browsing, on their mobile phones, worldwide.
- **Decoding the human genome** originally it took 10 years to process it; now, it can be achieved in one week.
- **The largest AT&T databases** boots titles, including the largest volumes of data, in one database (312 terabytes) and the second largest number of rows (1.9 trillion), which comprises AT&T's extensive calling records.

What is making so much Data?

- Sources of big data include **people, machines, and organizations** through ubiquitous computing.
- More people are carrying data-generating devices such as **smartphones, GPS devices, and cameras**.
- The Internet continuously generates massive amounts of data every second.
- Internet live statistics show billions of searches, emails, videos, and social media activities daily.
- **Internet Data Generation**
 - **4+ Billion Internet Users**
 - **500+ Hours Video Uploaded/Minute**
 - **350+ Email Sent Daily**
 - **Millions of Google Search Every Minute**
 - **Huge Social Media Activity Worldwide**

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

What is making so much Data?

12+ TBs
of tweet data
every day



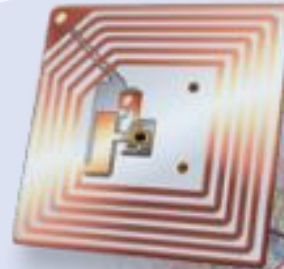
? TBs of
data every day



25+ TBs of
log data
every day



30 billion RFID
tags today
(1.3B in 2005)



4.6 billion
camera
phones
world wide



100s of millions
of GPS
enabled
devices sold
annually



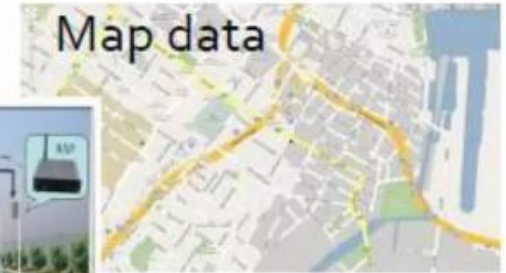
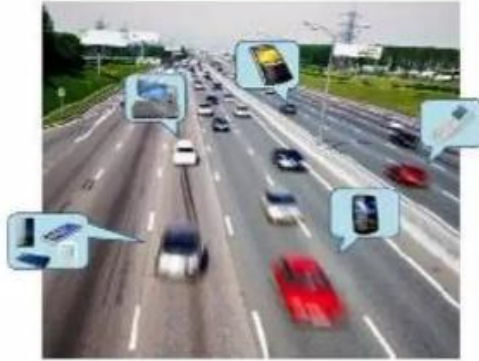
76 million smart meters
in 2009...
200M by 2014

2+ billion
people on
the Web
by end
2011



An Example of Big Data at Work

Crowdsourcing



Computing

Sensing



Real time traffic info

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Where is the Problem?

- Traditional RDBMS queries are not sufficient to gain useful information out of the huge volume of data
- To search it with traditional tools to find out if a particular topic was trending would take so long that the result would be meaningless by the time it was computed.
- Big data come up with a solution to store this data in a novel ways in order to make it more accessible, and also to come up with methods of performing analysis on it.

Challenges

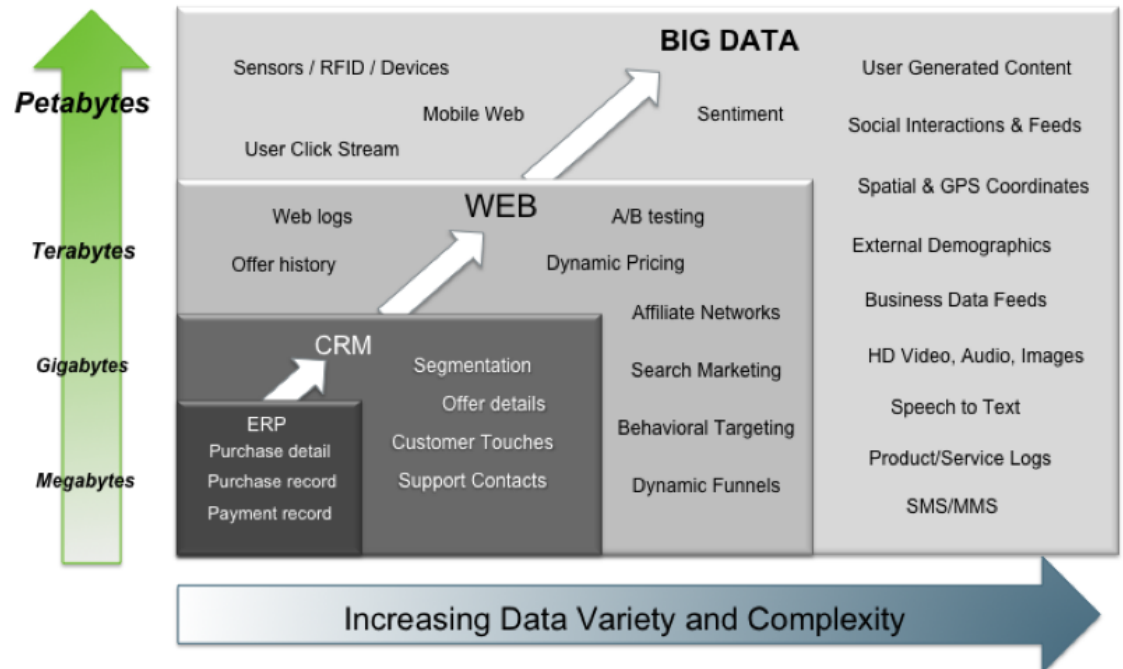
- Capturing
- Storing
- Searching
- Sharing
- Analyzing
- Visualizing

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

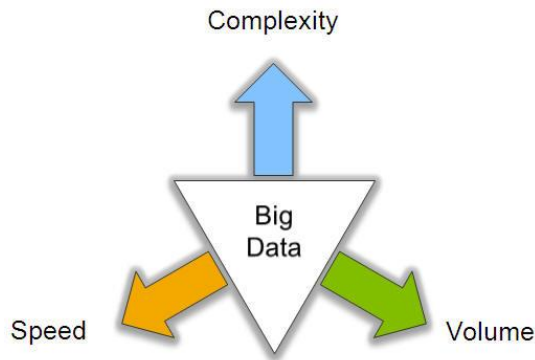
Big Data: 3V's



Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.



Volume (Scale)

- **Volume:** Enterprises are awash with ever-growing data of all types, easily amassing terabytes even Petabytes of information.
 - Turn 12 terabytes of Tweets created each day into improved product sentiment analysis
 - Convert 350 billion annual meter readings to better predict power consumption



Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

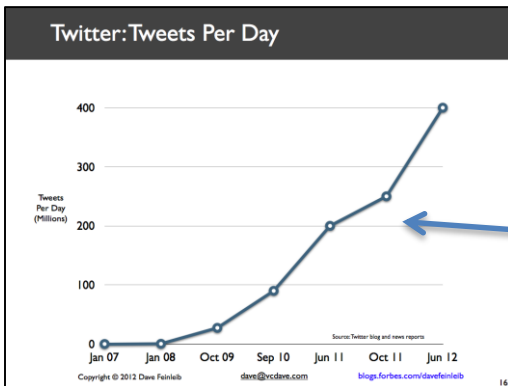
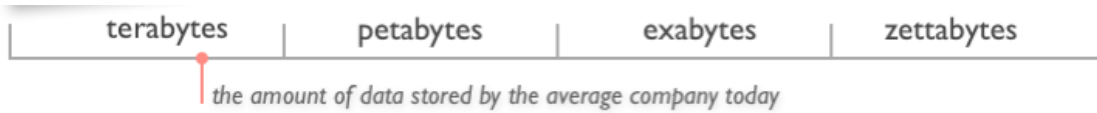
Volume (Scale)....

Data Volume

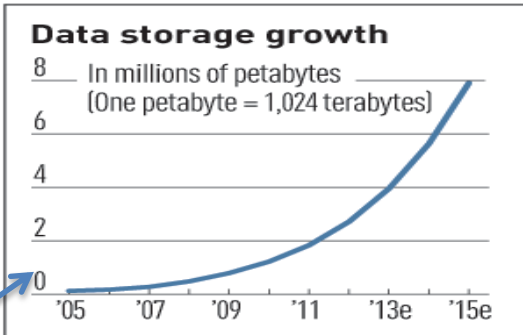
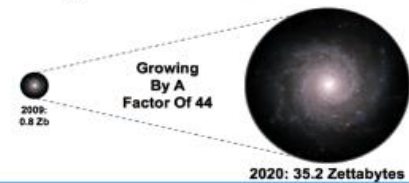
44x increase from 2009 2020

From 0.8 zettabytes to 35zb

Data volume is increasing exponentially



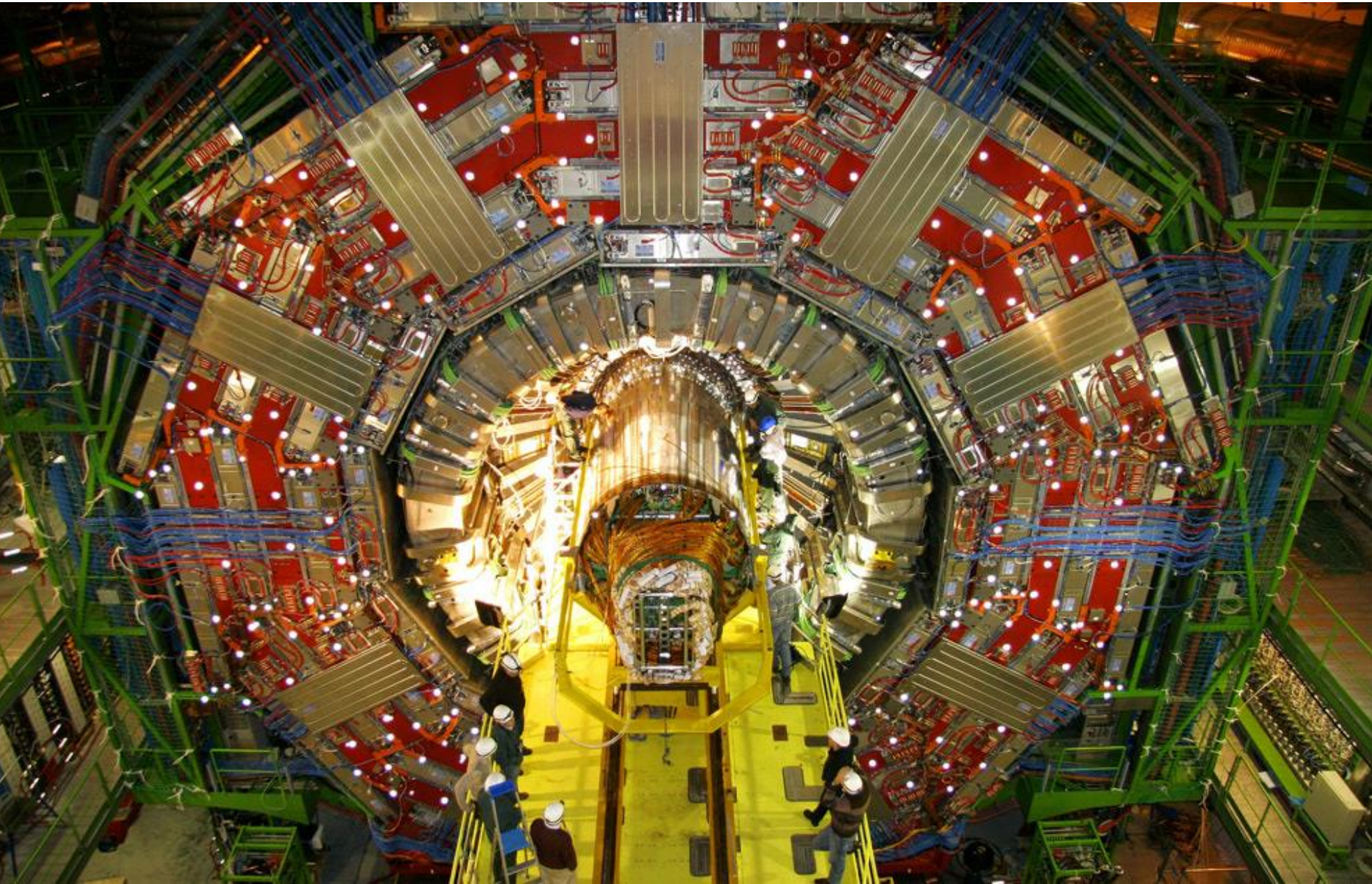
The Digital Universe 2009-2020



Exponential increase in collected/generated data

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Volume (Scale)....



CERN's Large Hadron Collider (LHC) generates 15 PB a year

Volume (Scale)

- **The Earthscope** is the world's largest science project. Designed to track North America's geological evolution, this observatory records data over 3.8 million square miles, amassing 67 terabytes of data. It analyzes seismic slips in the San Andreas fault, sure, but also the plume of magma underneath Yellowstone and much, much more.



Velocity (Speed)

- **Velocity:** Sometimes two minutes is too late For time sensitive process such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value.
 - Scrutinize five million trade events created each day to identify potential fraud.
 - Analyze five million daily call detail records in real time to predict customer churn pattern.



Velocity (Speed)...

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions → missing opportunities



- **Examples**

- **E-Promotions:** Based on your current location, your purchase history, what you like → send promotions right now for store next to you
- **Healthcare monitoring:** sensors monitoring your activities and body → any abnormal measurements require immediate reaction

Velocity (Speed)...

- **Real-time/Fast Data**



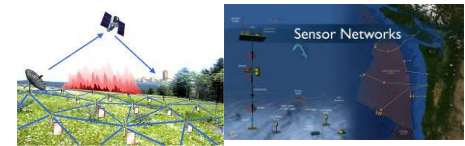
Social media and networks
(all of us are generating data)



Scientific instruments
(collecting all sorts of data)



Mobile devices
(tracking all objects all the time)



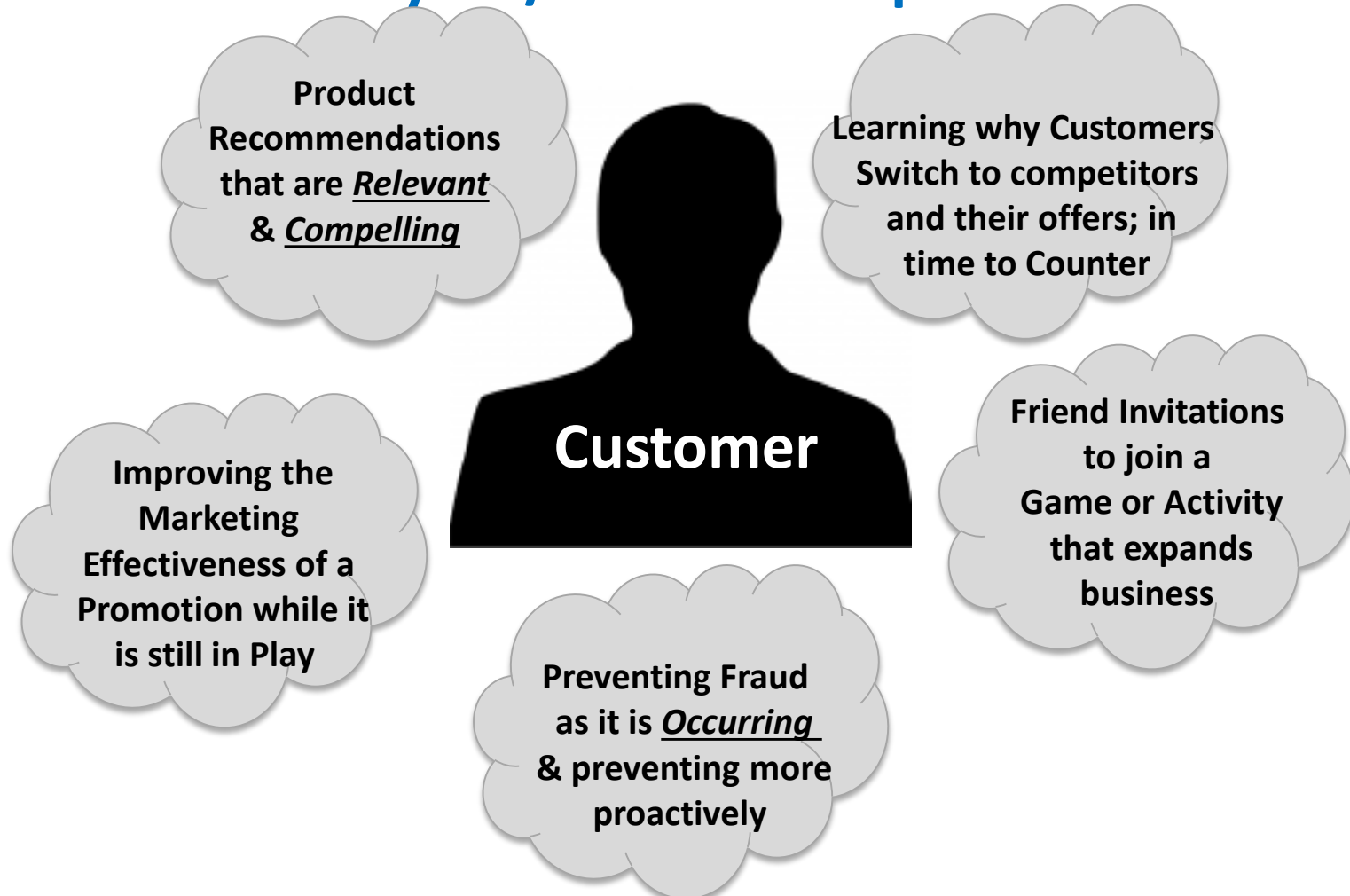
Sensor technology and networks
(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note.

Velocity (Speed)...

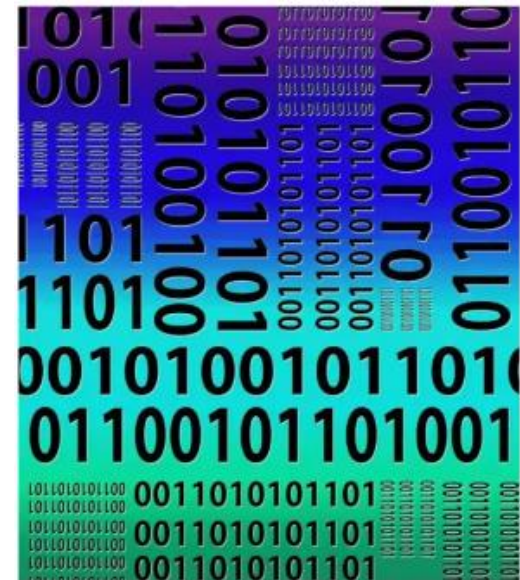
- **Real-Time Analytics/Decision Requirement**



Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note.

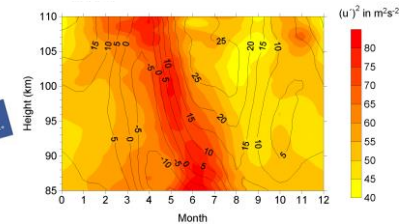
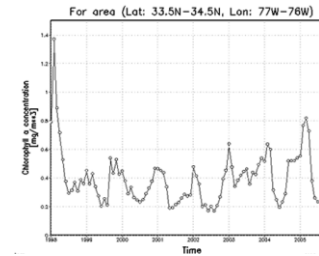
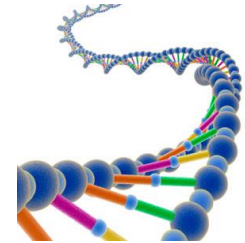
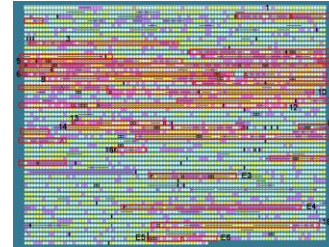
Variety (Complexity)

- **Variety:** Big data is any type of data –
 - Structured data (example: tabular data)
 - Semi Structured data (example: web data log files)
 - Unstructured data (example: text, sensor data, audio, video)



Variety (Complexity)...

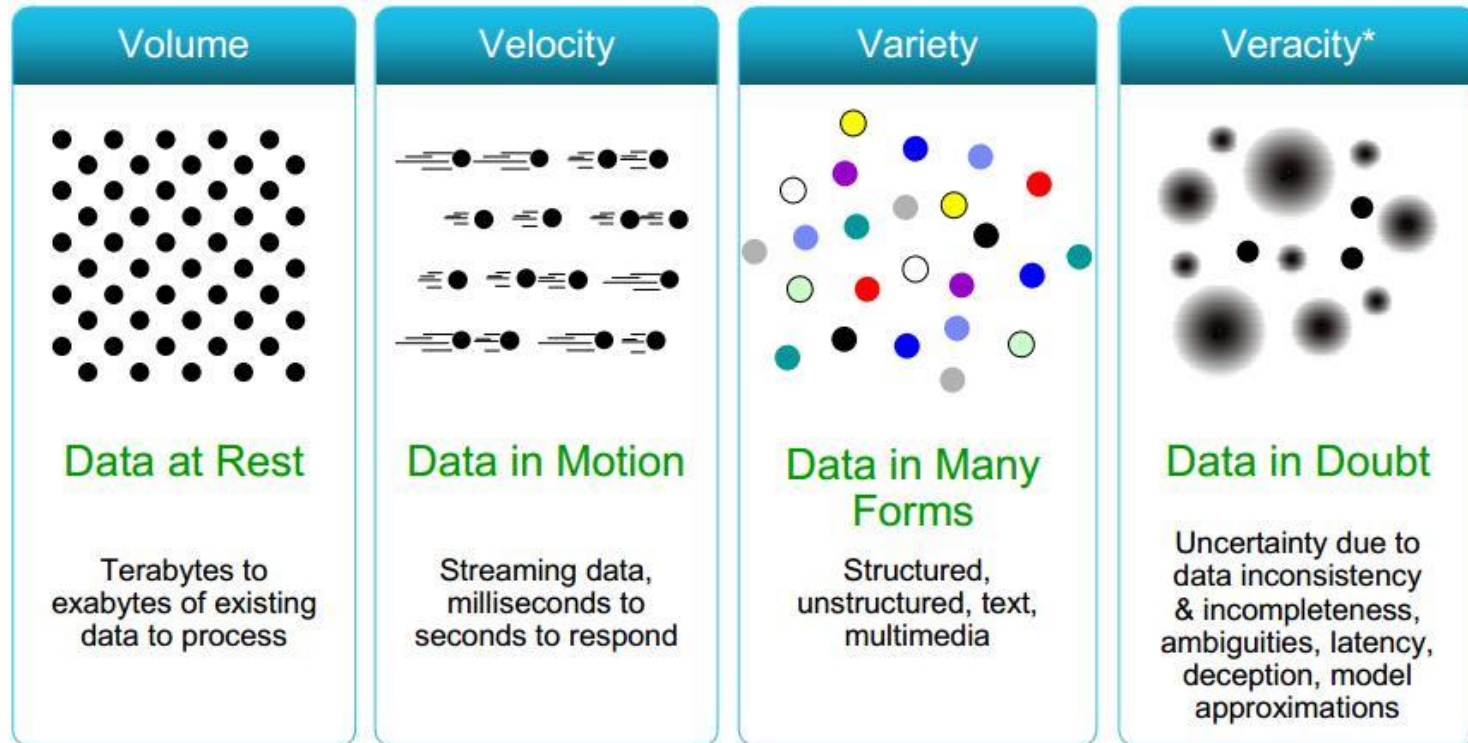
- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
 - Social Network, Semantic Web (RDF), ...
- Streaming Data
 - You can only scan the data once
- A single application can be generating/collecting many types of data
- Big Public Data (online, weather, finance, etc)



To extract knowledge → all these types of data need to be linked together

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. META Group Research Note.

Some Make it 4V's



Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. META Group Research Note.

Some Make it 4V's

- Veracity
- Validity
- variability
- Viscosity
- volatility
- Viability

....

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Value

- **Integrate Data**

- Reducing data complexity
- Increase data availability
- Unify your data systems
- All 3 above will lead to increased data collaboration



Veracity

- **Veracity** refers the biases, noises and abnormality in data, trust worthiness of data.
- 1 in 3 business leaders don't trust the information they use to make decisions.
 - How can you act on information if you don't trust it?
 - Establishing trust in big data presents a huge challenge as the variety and number of sources grow

Valence

- **Valence** refers the connectedness of big data.
- Such as the form of graph network.



Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

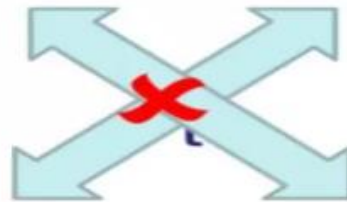
Validity

- Accuracy and correctness of the data relative to a particular use
- Example: Gauging storm intensity

• Satellite imagery

vs

Social media post



• Prediction quality

vs

human impact

Variability

- How the meaning of the data changes over time
 - Language evolution
 - Data availability
 - Sampling process
 - Changes in the characteristics of the data source

Viscosity and Volatility

- Both related to velocity
- **Viscosity:** data velocity relative to timescale of event being studied.
- **Volatility:** rate of data loss and stable lifetime of data
 - Scientific data of ten has practically unlimited lifespan, but social/business data may evaporate in finite time.

More V's

- **Viability**
 - Which data has meaningful relation to questions of interest?
- **Venue**
 - Where does the data live and how do you get it?
- **Vocabulary**
 - Metadata describing structure, content and provenance
 - Schema, semantics, ontologies, taxonomies, vocabularies
- **Vagueness**
 - Confusion what “Big Data” means

Summary

- The major sources of big data.
- The key characteristics of big data such as volume, velocity and variety.
- The major challenges associated with big data.

References

- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note.