Georgia State University

J. Mack Robinson
COLLEGE of BUSINESS

Big Data:
Characteristics and Applications

Agenda

Are we really at the tip of the ice berg wrt big data?
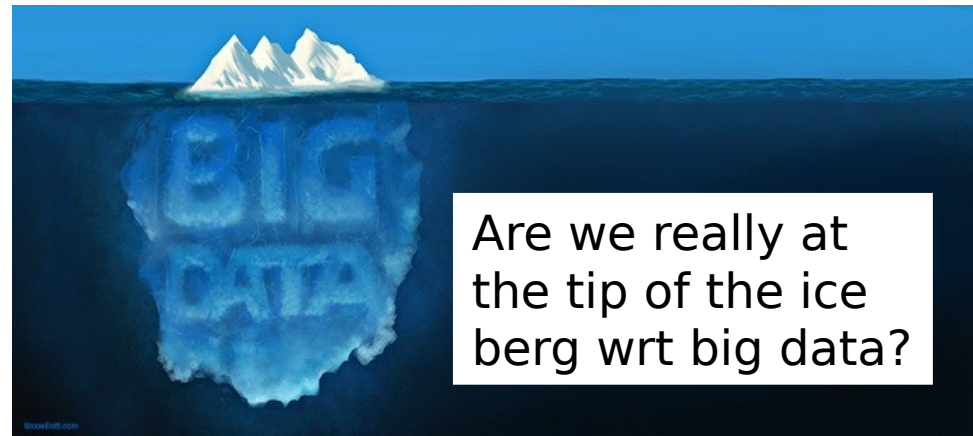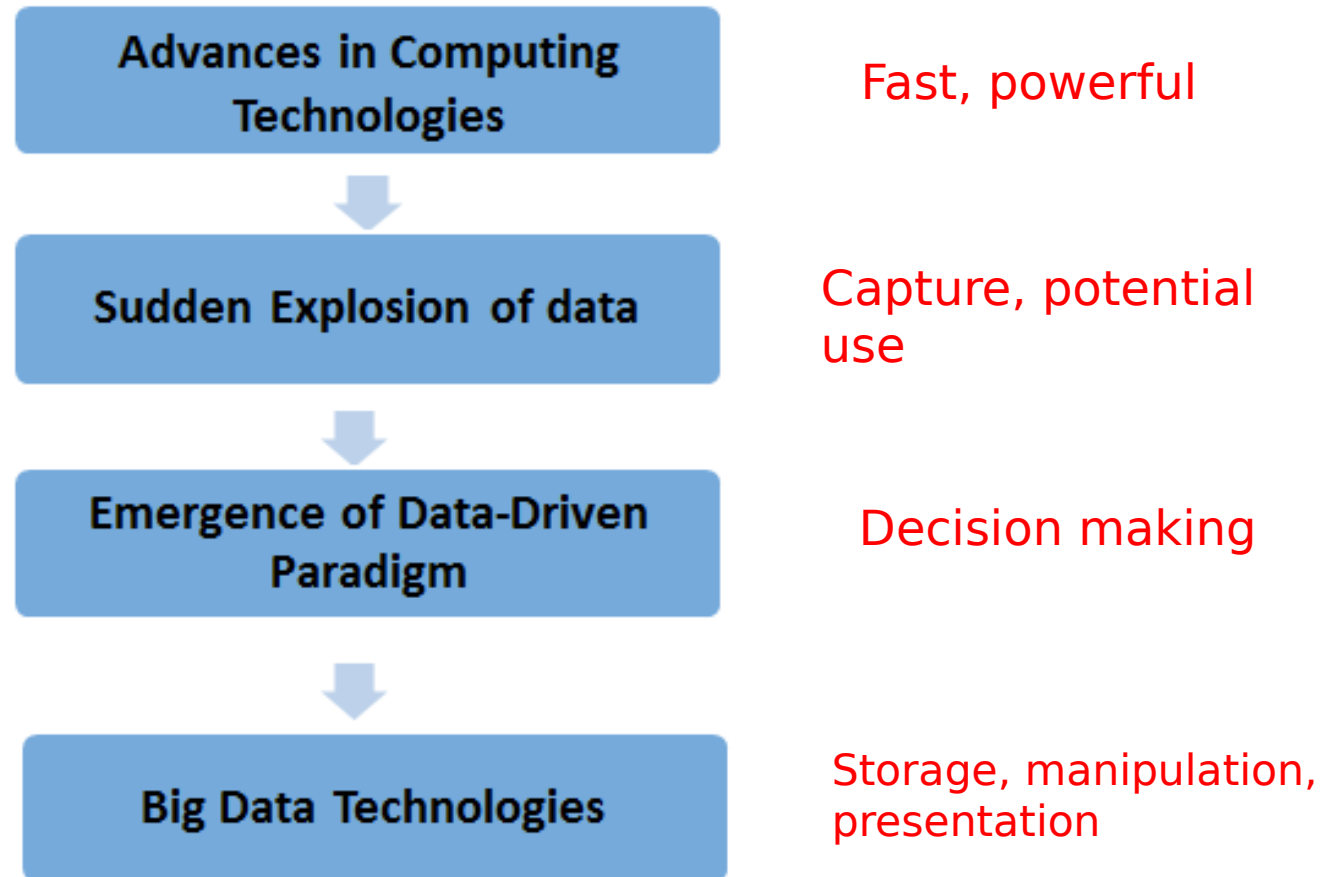
- ▹ Why big data?
- ▹ Data-driven decision making
- ▹ Characteristics of big data
- ▹ Issues
- ▹ Applications

# Why Study Big Data?

What caused the "big data" notion?

Advances in Computing Technologies	Fast, powerful
⬇	
Sudden Explosion of data	Capture, potential use
⬇	
Emergence of Data-Driven Paradigm	Decision making
⬇	
Big Data Technologies	Storage, manipulation, presentation

Big Data

(Mayer-Schonberger & Cukier)

- ▸ **Use whole data set, rather than sampling**
  - ◦ **e.g., Walmart**
- ▸ **Probability of prediction**
- ▸ **Big data outperforms experts (sometimes)**
  - ◦ **e.g., Amazon reviewers (crowd-sourced)**
- ▸ **Accept correlation**
  - ◦ **causality may not be possible**
- ▸ *Datafication*
  - ◦ **Quantify/measure as much data as possible**
  - ◦ **Let data speak for itself**
- ▸ **Movement of value**
  - ◦ **physical items ☽ brands, ideas, and intellectual rights (digital intellectual property, dIP)**
- ▸ **Implications for management?**
- ▸ **Implications for innovation?**

Forecast of Big Data market size, based on revenue: 2011 to 2027 (in billion U.S. dollars)

Big Data-related hardware, software, and professional services:
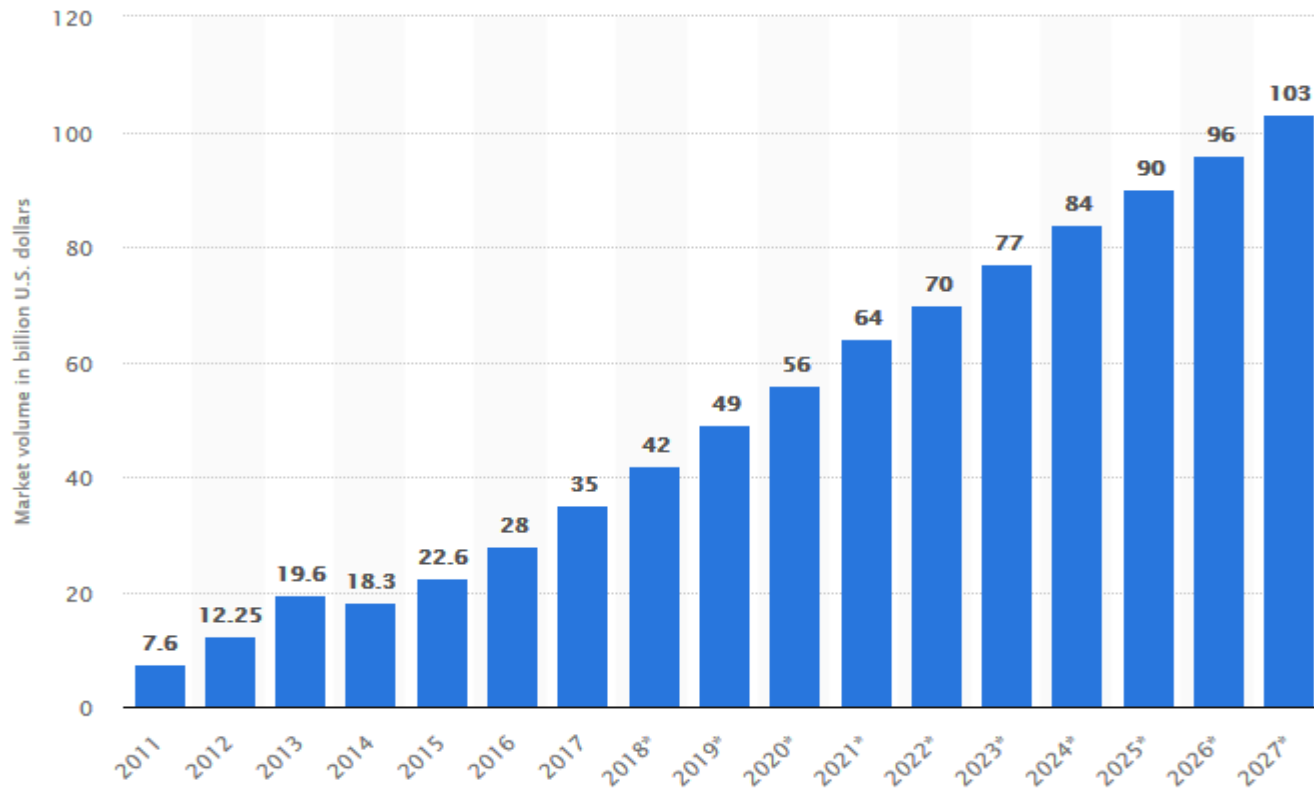
Implications?

What is Big Data? (3V)

High volume, velocity, and /or variety information assets that demand new, innovative forms of processing data for enhanced decision making, business insights or process optimization.

(Big Data: A Definition; Gartner)

Sources of Big Data (Gartner)

◦ **Operational data:** transaction systems, streaming data, sensor data;

◦ **Dark data:** own but don't use: emails, contracts, reports, etc.

◦ **Commercial data:** Structured or unstructured data purchased from industry organizations, social media, etc.

◦ **Social data:** from Twitter, Facebook, etc.

◦ **Public data:**, economic data, socio-demographic data, weather data, etc. in many formats and on many topics

◦ **Data Exhaust** -- "left over" data from core data digital transactions, collected, either intentionally or unintentionally, but for which there is no initial, specific purpose for its collection.

E.g. post on Facebook, Google searches, shopping carts

Volume: Gigabyte (10^9)
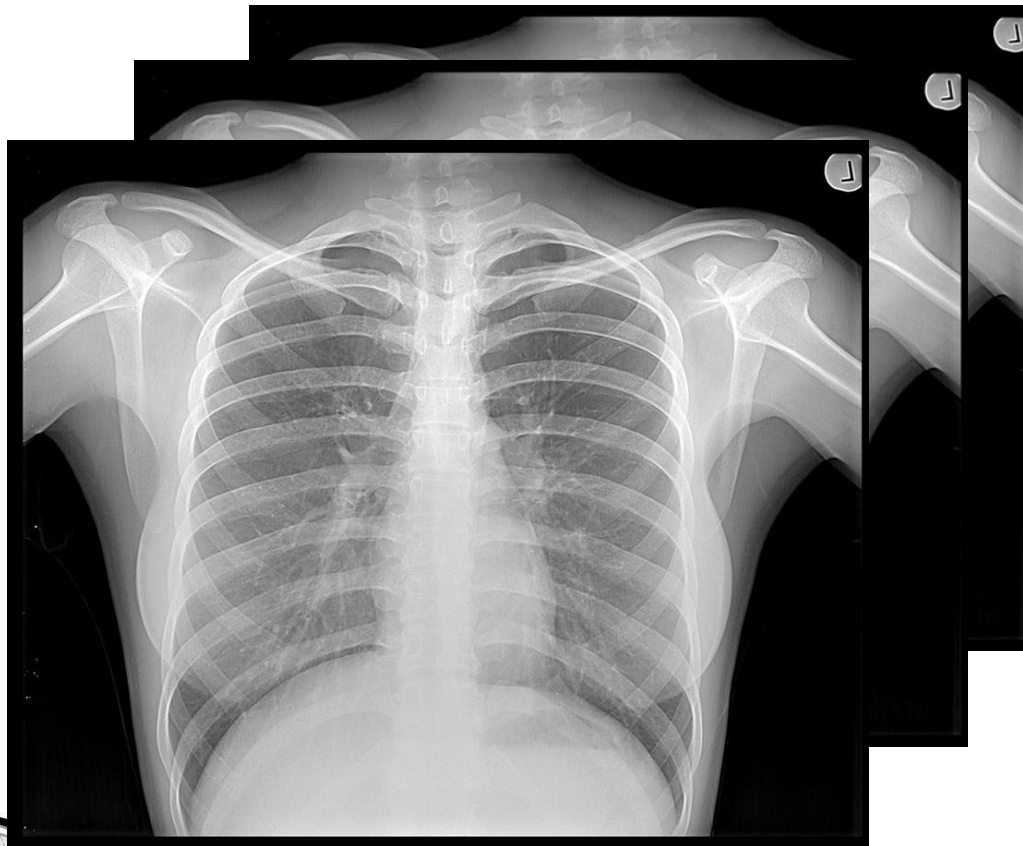
Sound wave of Beethoven's Fifth Symphony

Ref. Il-Yeol Song, Ph.D.

Volume: Terabyte (10¹²)

All the X-ray images in a large hospital

02/17/2022

Volume: Petabyte (10^{15})

10 billion Facebook photos

Recall: Wal-Mart and its Data

Really Big Data At Walmart: Real-Time Insights From Their 40+ Petabyte Data Cloud

Tagline: "Walmart – the world's biggest retailer with over 20,000 stores in 28 countries, is in the process of building the world' biggest private cloud, to process 2.5 petabytes of data every hour. Over 200 streams of internal and external data, including 40 petabytes of recent transactional data, can be modelled, manipulated and visualized.

E.g. Mine data to identify abnomalties

Ref. https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#74bdf6556c10

Volume: Exabyte (10^{18})

1/5 of the words ever spoken

Ref. Il-Yeol Song, Ph.D.

Volume: Zettabyte (10^{21})

Grains of sand on all the world's beaches

Ref. Il-Yeol Song, Ph.D.

13

Volume: Yottabyte (10^{24})

Atoms in 7,000 human bodies
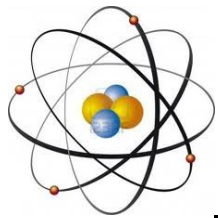
NSA data site – purportedly designed to store yottabytes of data.

Ref. Il-Yeol Song, Ph.D.
14

Business Volume Example: eBay

Revenue: $10B
Ebay Listings:
1.2 Billion
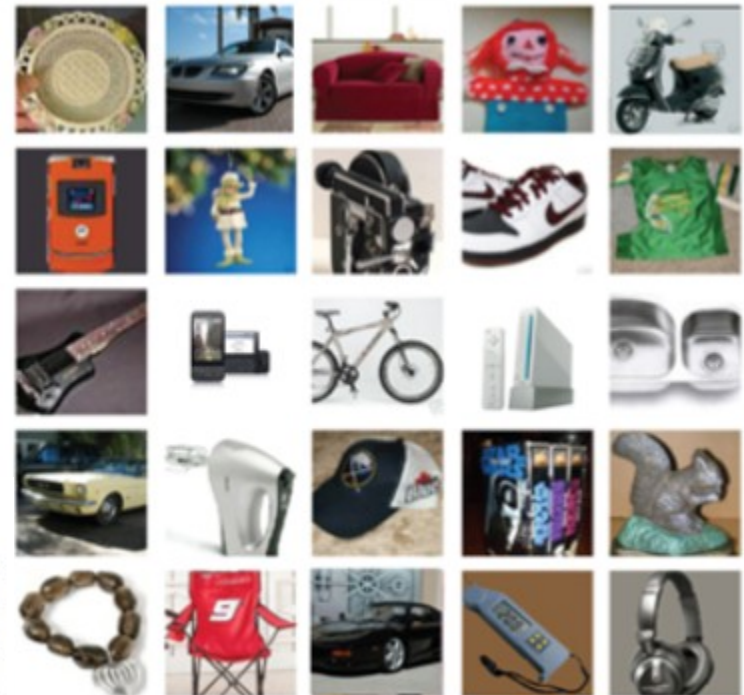
Marketplace Sales (In Billions)

	2014	2015	2016	2017	2018
ebay	$8.79	$8.59	$8.98	$9.96	$10.75
amazon.com	$88.99	$107.01	$135.99	$177.87	$232.89
eBay % of Amazon	9.9%	8.0%	6.6%	5.6%	4.6%

https://expandedramblings.com/index.php/ebay-stats/

Source: Uwe Mayer

Velocity: Ebay Sales by Country

In the United States . . .

1. A car or truck is bought via mobile every 4 minutes.
2. A pair of ladies shoes is bought via mobile every 7 seconds.
3. A ladies handbag is bought via mobile every 10 seconds.
4. A tablet is bought via mobile every 40 seconds.
5. A cell phone is sold every 4 seconds.

In the United Kingdom . . .

6. A car is bought via mobile every 2 minutes.
7. A pair of ladies shoes is bought via mobile every 6 seconds.
8. A ladies handbag is bought via mobile every 17 seconds.
9. A tablet is bought via mobile every 1 minute.

In Germany . . .

10. A car or truck is bought via mobile every 9 minutes.
11. A pair of ladies shoes is bought via mobile every 15 seconds.
12. A ladies handbag is bought via mobile every 58 seconds.
13. A tablet is bought via mobile every 3 minutes.

In Australia . . .

14. A car is bought via mobile every 20 minutes.
15. A pair of ladies shoes is bought via mobile every 1 minute.
16. A ladies handbag is bought via mobile every 2 minutes.
17. A tablet is bought via mobile every 10 minutes.

In Canada . . .

18. A pair of ladies shoes is bought via mobile every 5 minutes.
19. A ladies handbag is bought via mobile every 6 minutes.
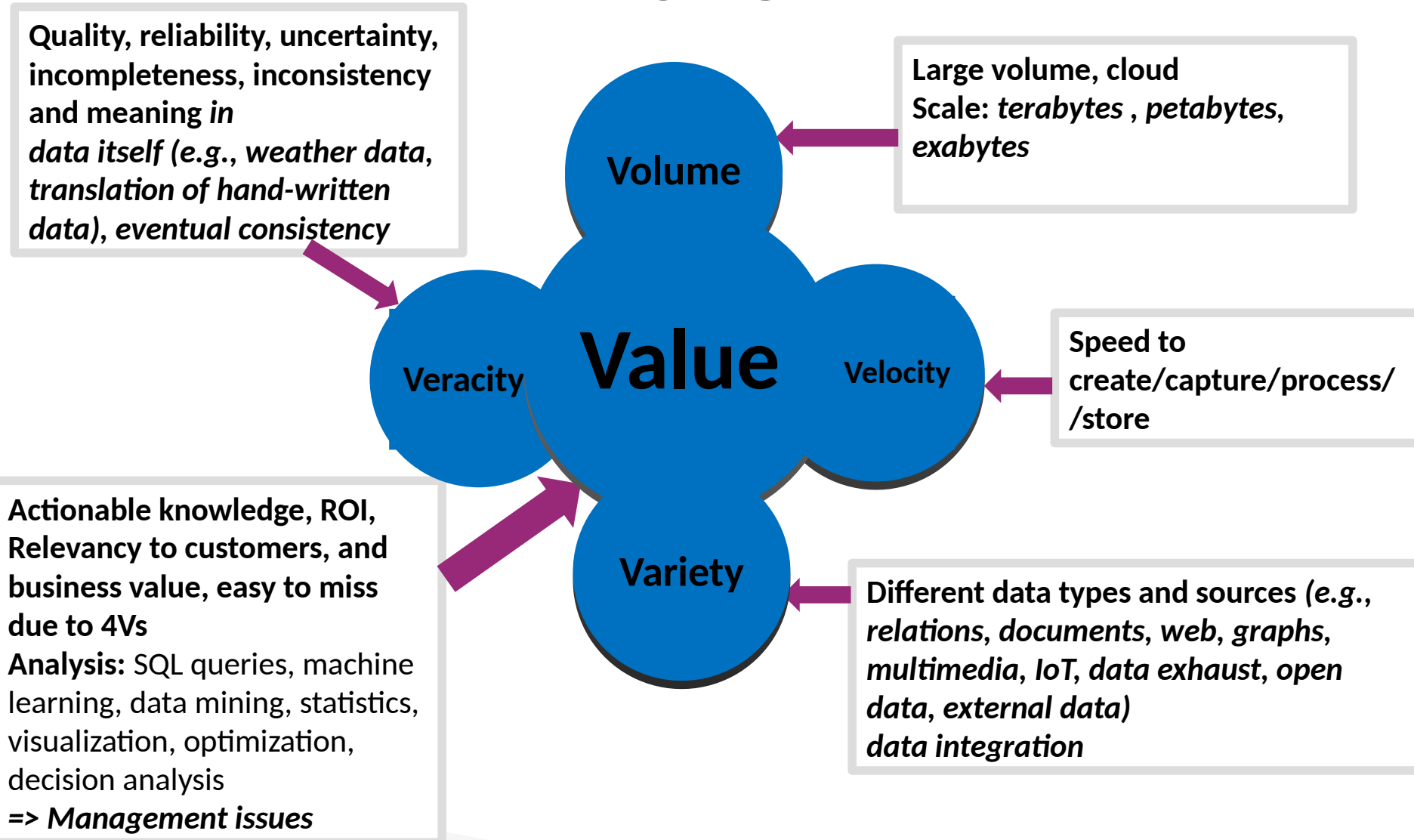20. A tablet is bought via mobile every 25 minutes.

every
9
sec.
a pair of shoes is sold

Source: Uwe Mayer

Characteristics and Challenges of Big Data

5V's

Quality, reliability, uncertainty, incompleteness, inconsistency and meaning *in data itself (e.g., weather data, translation of hand-written data), eventual consistency*

Large volume, cloud
Scale: *terabytes , petabytes, exabytes*

Volume

Veracity

Value

Velocity

Speed to create/capture/process/ /store

Actionable knowledge, ROI, Relevancy to customers, and business value, easy to miss due to 4Vs
Analysis: SQL queries, machine learning, data mining, statistics, visualization, optimization, decision analysis
=> *Management issues*

Variety

Different data types and sources (*e.g., relations, documents, web, graphs, multimedia, IoT, data exhaust, open data, external data*)
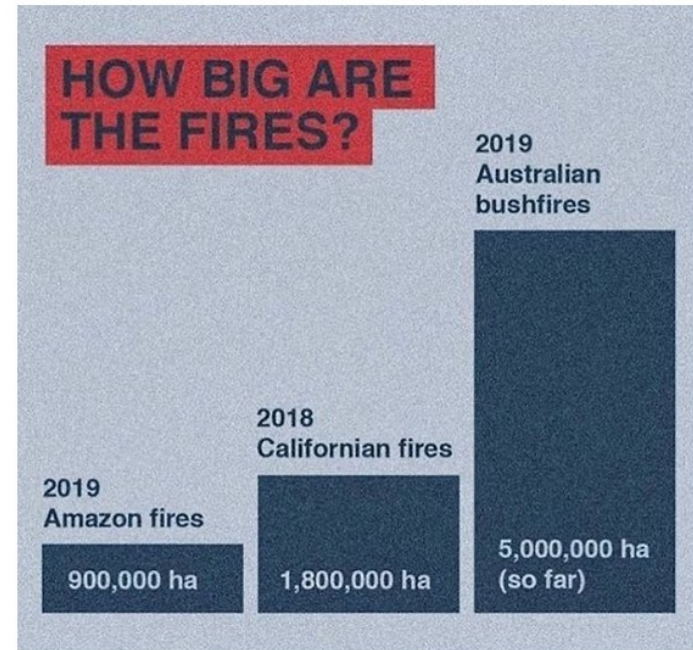*data integration*

Representation: Visualization

Visualization of Data

3D "Visualization" Of The Fires In Australia, Made From a NASA Satellite.

How Big Are The Fires?
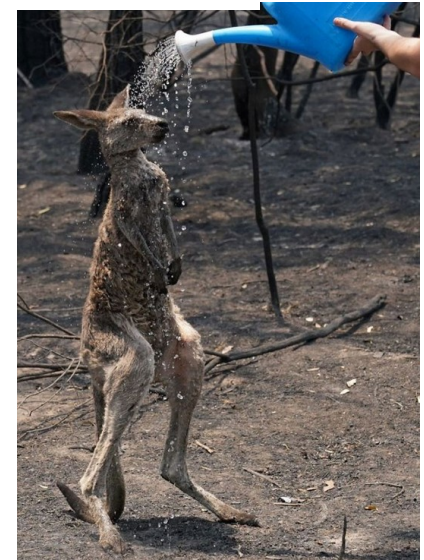

HOW BIG ARE THE FIRES?

2019 Amazon fires — 900,000 ha

2018 Californian fires — 1,800,000 ha

2019 Australian bushfires — 5,000,000 ha (so far)

6 Deaths. 2.2 Million Ares Burned. 680 Homes Destroyed. Countless Animals Lost. And No End In Sight...

Area Of Land Burnt In Australia Compared To The Size Of The Country of Ireland.

# Management Issues

- Big data –affect *how we do business*
- Succeeding with big data -- requires more than just having data.
  - *Data-based value creation* requires **identification of patterns** from which predictions can be inferred and decisions made.
  - *Value creation --* Businesses need to decide which data to use (traditional vs big)
  - *Value creation* -- requires right way of dissecting and analyzing data with right analytics.

World Economic Forum Global Technology Report Bilbao
[Via Dean Phillips]

Problem-Solving in Big Data

Data

Question **Big Data** **Techniques**

Answer
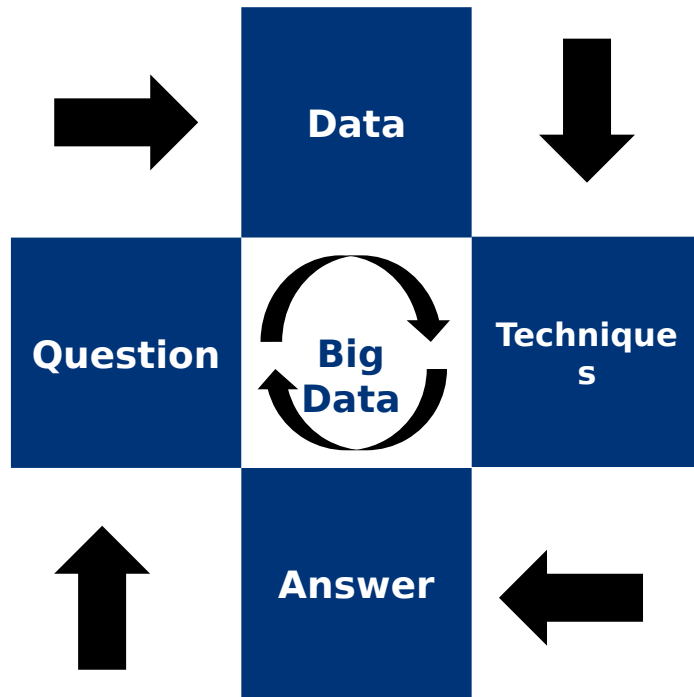
- Business situation
  - (problem / challenge / reaction)
- Input data
  - merge/combine multiple data sources
- Specify desired output data
  - whether/how can compute desired results
- Interpret results
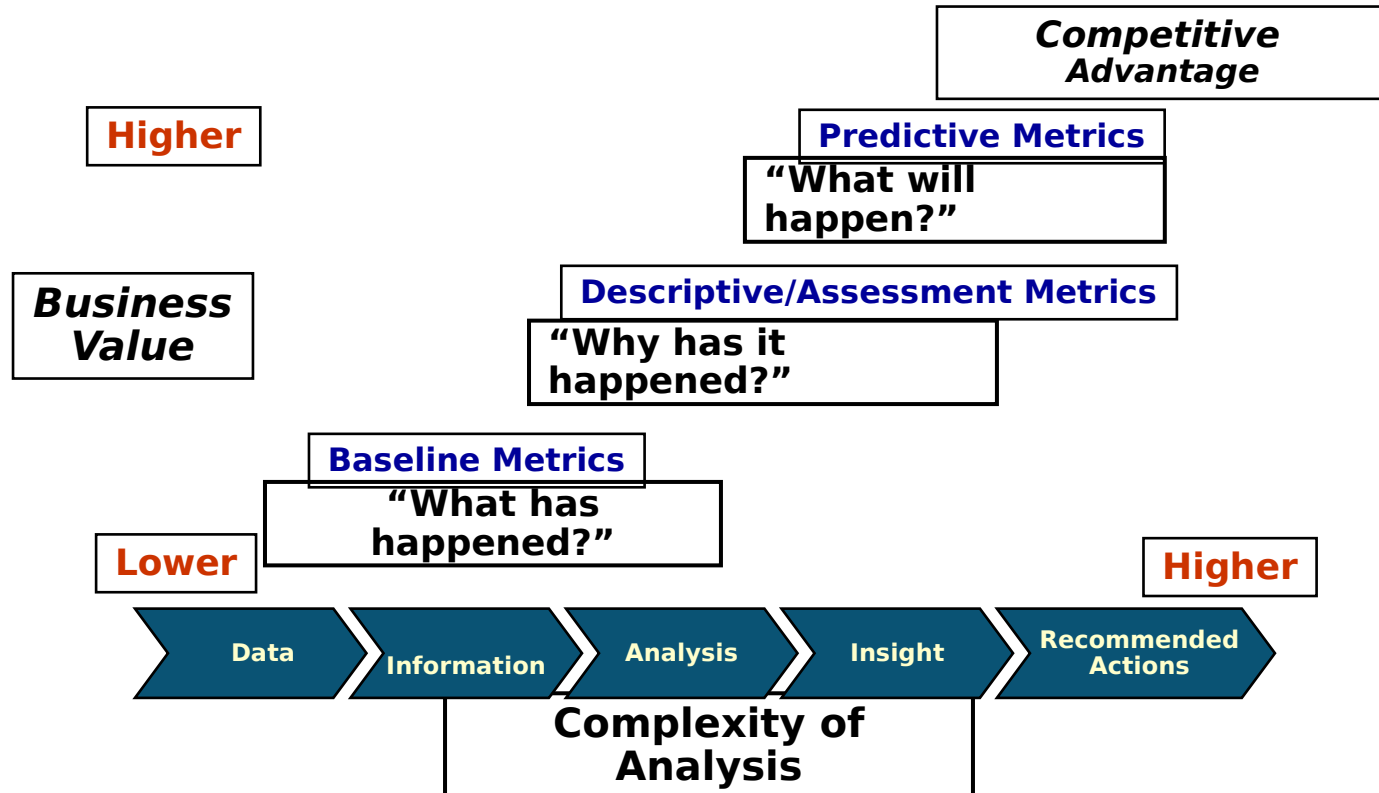 - look for new angles in **knowledge discovery / insights**

Recall: Data Mining

- **Includes:**
  - **Identifying valid, novel, potentially useful, and ultimately understandable patterns in data**
  - **Searching for relationships, patterns, and trends not known to exist or not visible**
  - **Providing answers to questions decision maker not thought to ask**
- **Requires:**
  - **Information technology**
  - **Statistics**
  - **Business knowledge**

Perhaps ROBINSON Database Management Course (or equivalent) is a start.
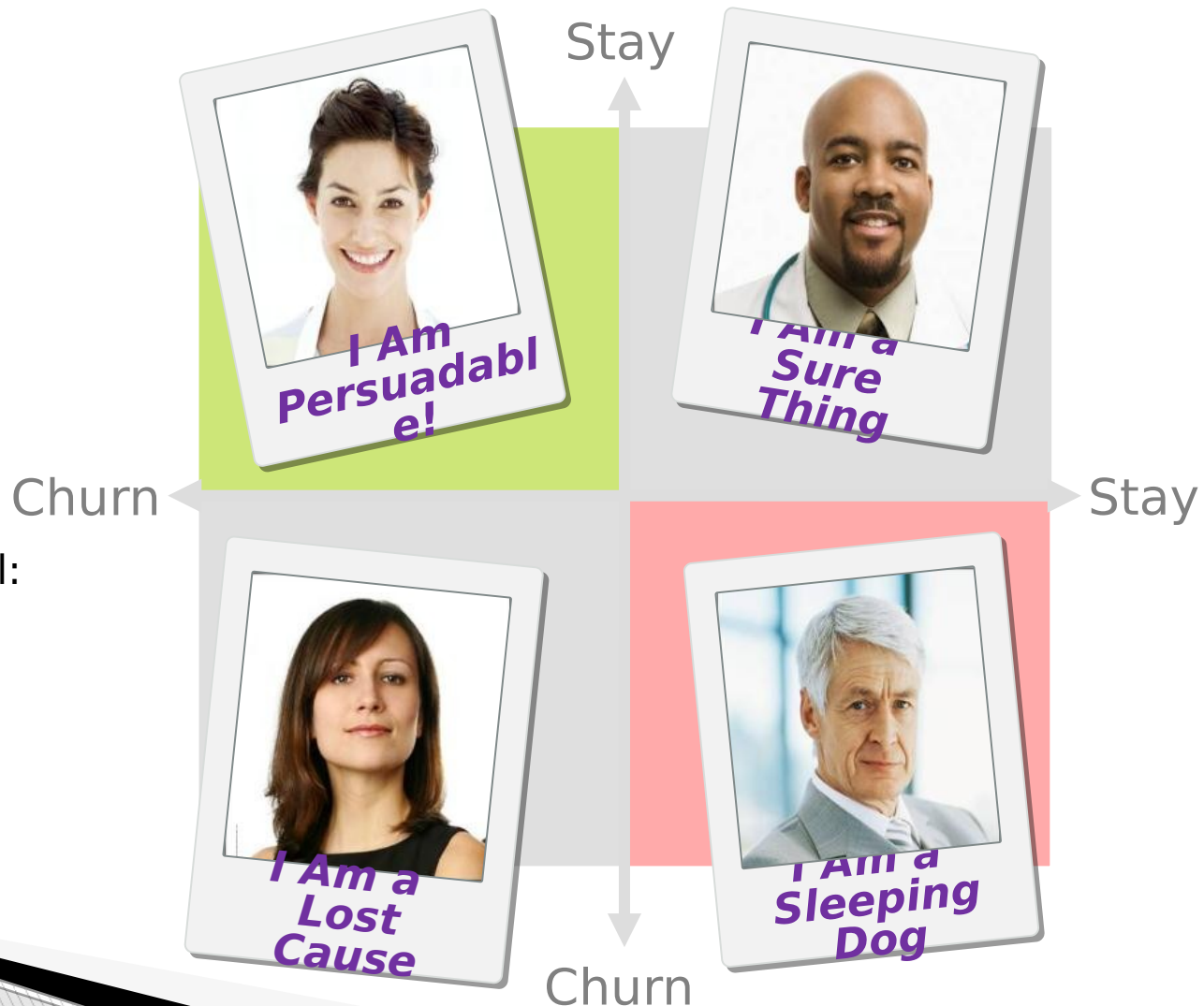
Recall: Extraction of Value

Competitive Advantage

Higher

Predictive Metrics
"What will happen?"

Business Value

Descriptive/Assessment Metrics
"Why has it happened?"

Baseline Metrics
"What has happened?"

Lower

Higher

Data → Information → Analysis → Insight → Recommended Actions

Complexity of Analysis

Recall: Predictive analytics (Herschel)
What is this market segmentation based on?

If we target, the customer will:

Stay

Churn ← → Stay

I Am Persuadable!

I Am a Sure Thing

If we **do nothing,** the customer will:

I Am a Lost Cause

I Am a Sleeping Dog

Churn

Applications of Big Data Analytics

Financial Services
- Optimize trading decisions with contextual awareness (news, weather)
- Analyze social media data to predict customer buying behavior

Utilities
- Smart meter data analysis
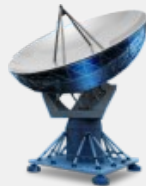- Refine wind pattern models for wind farm and individual wind turbine siting.

Transportation
- Use Streams to optimize cargo/asset location and tracking

Law Enforcement, Defense & Cyber Security
- Accurately identify suspects and victims, accounting for non-obvious relationships
- Evaluate incident patterns based on previous and current crime patterns

Telecommunications
- Detect and prevent fraudulent activities with abnormal call/data usage pattern analysis
- Predict likely customer churn by analyzing customer interaction via multiple channels
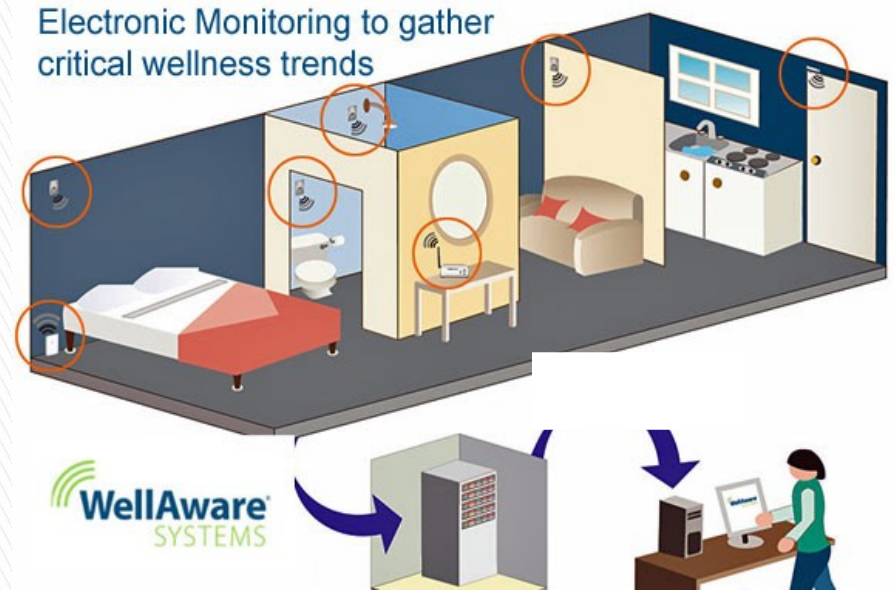
Retail
- Optimize supply chain, leveraging massive amounts of structured and unstructured data
- Create personalized promotions, utilizing massive transaction data, customer interactions, reviews, feedback, and social media

Ref. IBM?

Mobile
E.g., health care monitoring

Pascal Hitzler
communication from mobile
voice recognition may distort spelling

Emerging: Opportunities for Incorporating *Preferences* in Healthy Aging

Independent: can people be nudged toward healthier choices?

- Early detection of declining good habits (e.g., nutrition, sleeping, keeping active)
- Early intervention to suggest healthier alternatives
- Incorporate preferences in insurances, treatments, social services, activities, etc.

Electronic Monitoring to gather critical wellness trends

WellAware SYSTEMS

▸ Ref. Il-Yeol Song, Ph.D.

Big Data versus Traditional Model

Business Users
Determine what question to ask

IT
Structures the data to answer that question

IT
Delivers a platform to enable creative discovery

Business
Explores what questions could be asked

Structured & Repeatable Analytics
- Query Based -- Questions Drive Data
- Citizen Surveys
- Monthly, Weekly, Daily

Data At Rest

Versus

Iterative & Exploratory Analytics
- Autonomic -- Insight Drives Answers
- Citizen Sentiment
- Persistent and Ad Hoc

Data In Motion

Many more examples .....

- Google Flu Trends – Big data
- Google—Wikepedi-Twitter indicator of Covid19
- Delta Revenue Management – not big data, but CMR
- Linked-In – not big data (recommendations/suggestions)
- UPS – 16.3 million per day – not big data

Implications?
     For use of data?
     For management of data?

Conclusion: BIG Data

▸ Characteristics

  ○ Volume, velocity, veracity, variety, **value**

▸ Organizations

  ▫ **insights** from massive datasets

▸ Data analytics and business intelligence

  ▫ **discovery** and **innovation** (possible?)

    ▫ What questions to ask of the data?

  ▫ changing marketing (how?)

  ▫ privacy concerns (why?)

    ▫ hurt corporate reputations if mishandled