

Business Analytics and Data Driven Decision Making

Session#01: Lecture#01: Foundations of Business Data Analytics

Ravi Vatrapu

Director, Centre for Digital Enterprise Analytics and Leadership (DEAL)

Loretta Rogers Research Chair in Digital Enterprise

Professor, Department of Information Technology Management

Ted Rogers School of Management

Toronto Metropolitan University, Canada

Founding Director, Centre for Business Data Analytics (bda.cbs.dk)

Copenhagen Business School, Denmark

Visiting Professor, Indian Institute of Management Visakhapatnam, India

Visiting Professor, GITAM Deemed University, India

Visiting Professor, University of Surrey, United Kingdom

Email: vatrapu@torontomu.ca

DEAL Website: <https://www.torontomu.ca/tedrogersschool/digital-enterprise-analytics-and-leadership/>

Faculty Web: <https://www.torontomu.ca/information-technology-management/faculty-research/ravi-vatrapu/>



Outline

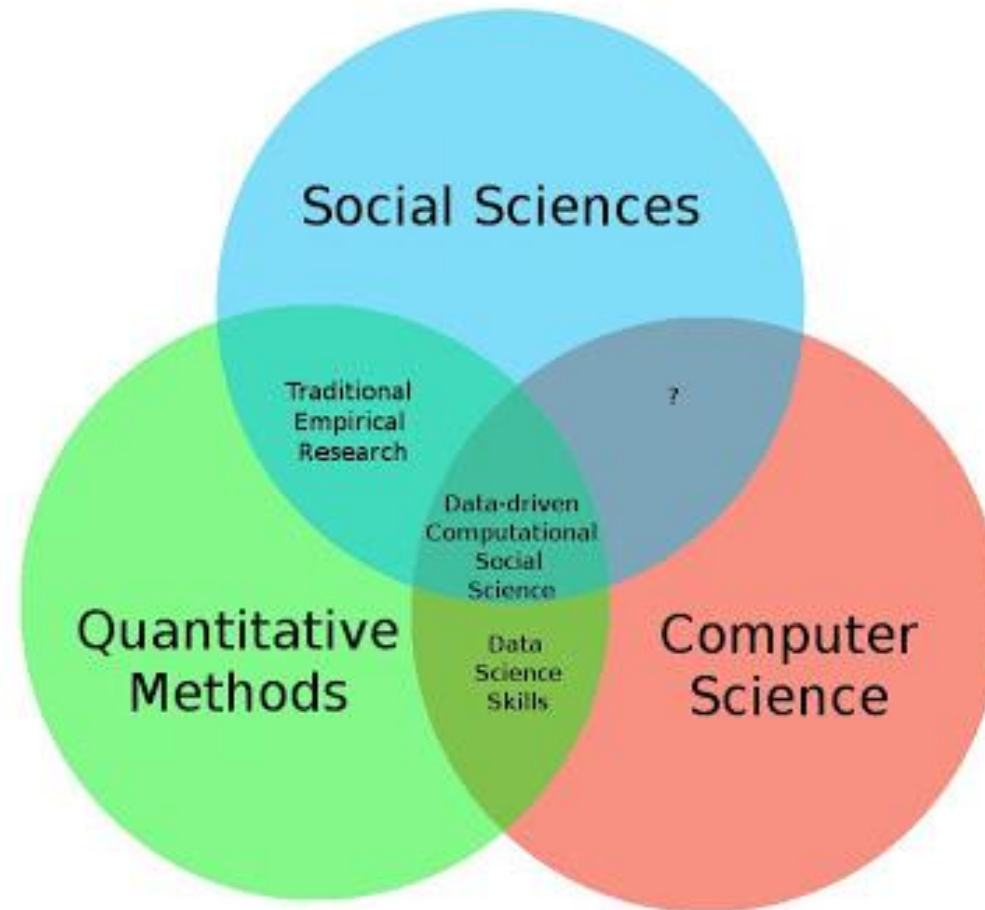
- Data Science
 - Computational Social Science
 - Internet of Events
- What is Big Data?
- What is Big Data Analytics?
- Challenges of Big Data Analytics
- Six provocations for Big Data
- Ten stories that can be told with Big Data Analytics

- Definitions
- Big Data vs. Business Data
- Big Data Analytics vs. Business Data Analytics
- Combining Big Data and Business Data Analytics

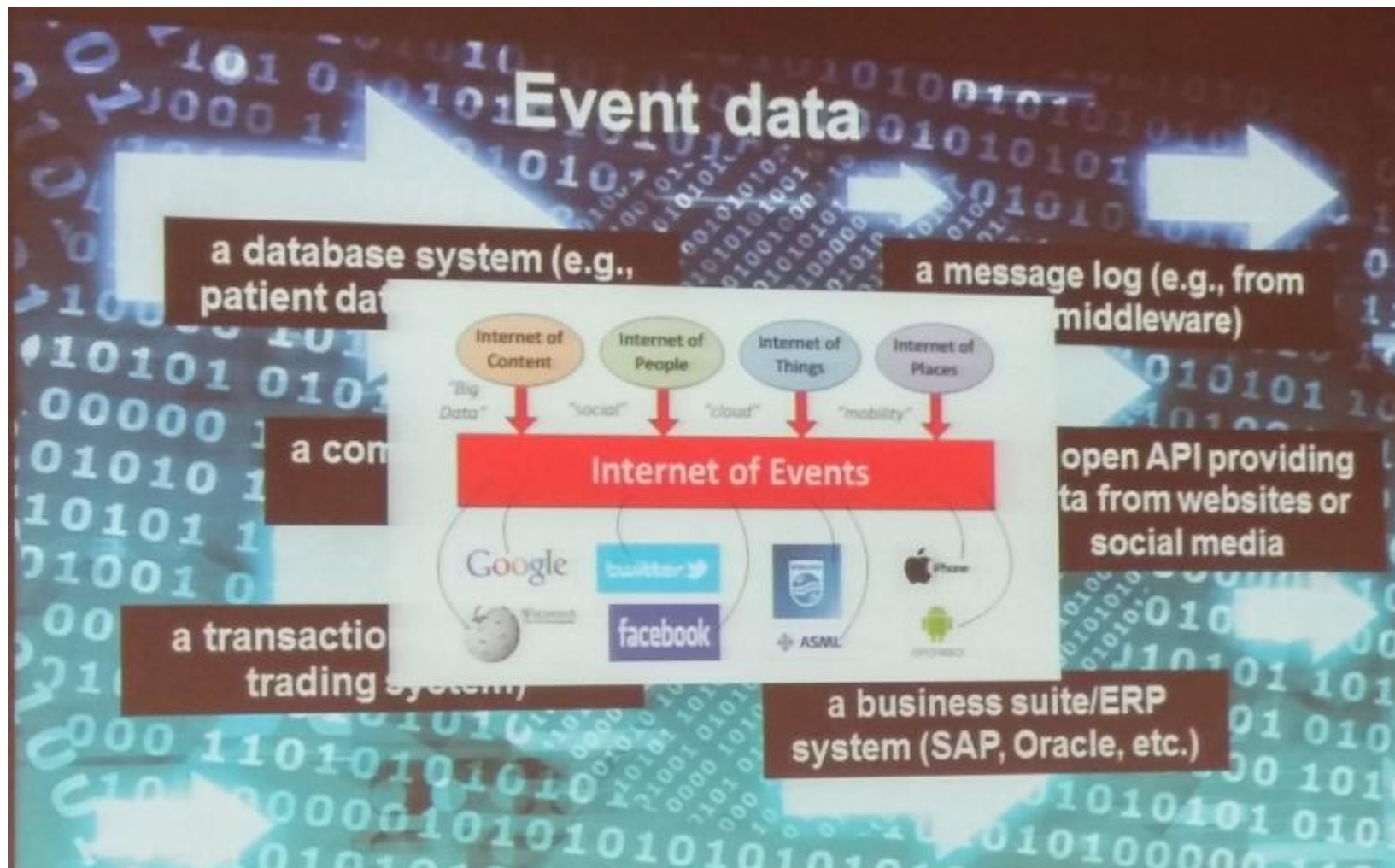
Data Science



Computational Social Science

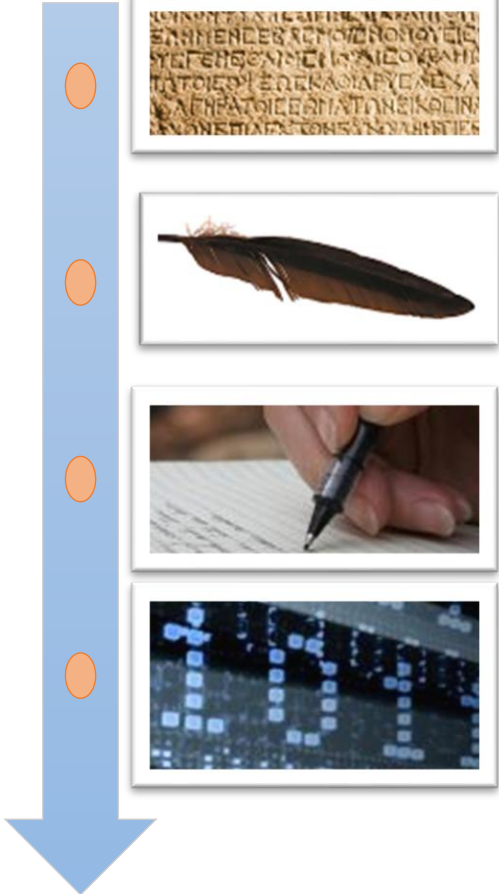


“Internet of Events”

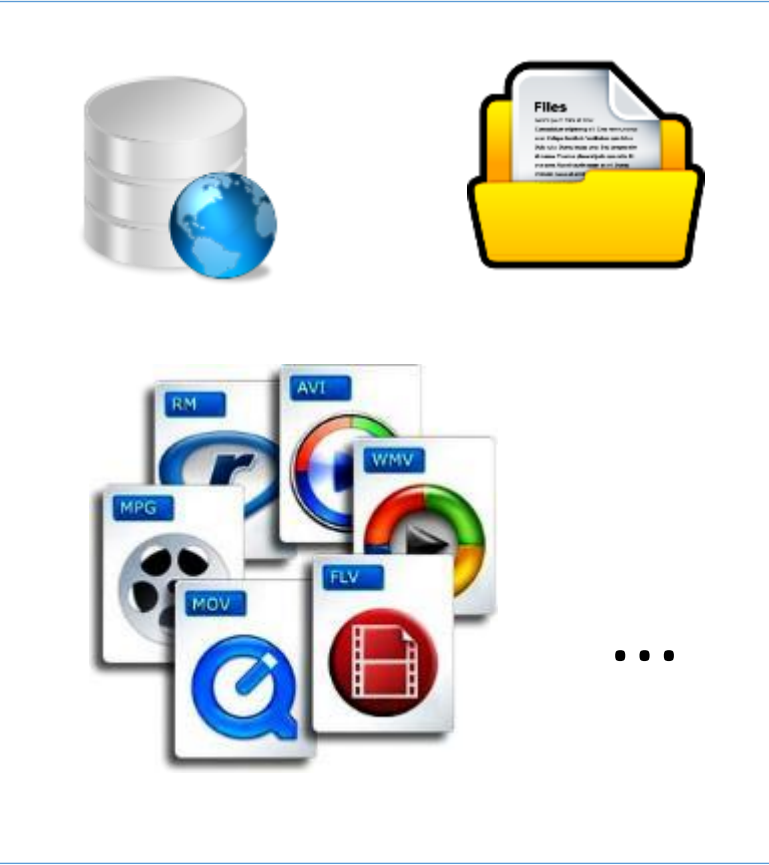


Data

Ancient



Today



Data

Data

1. Structured Data

- Database (e.g. SQL), etc.

2. Unstructured Data

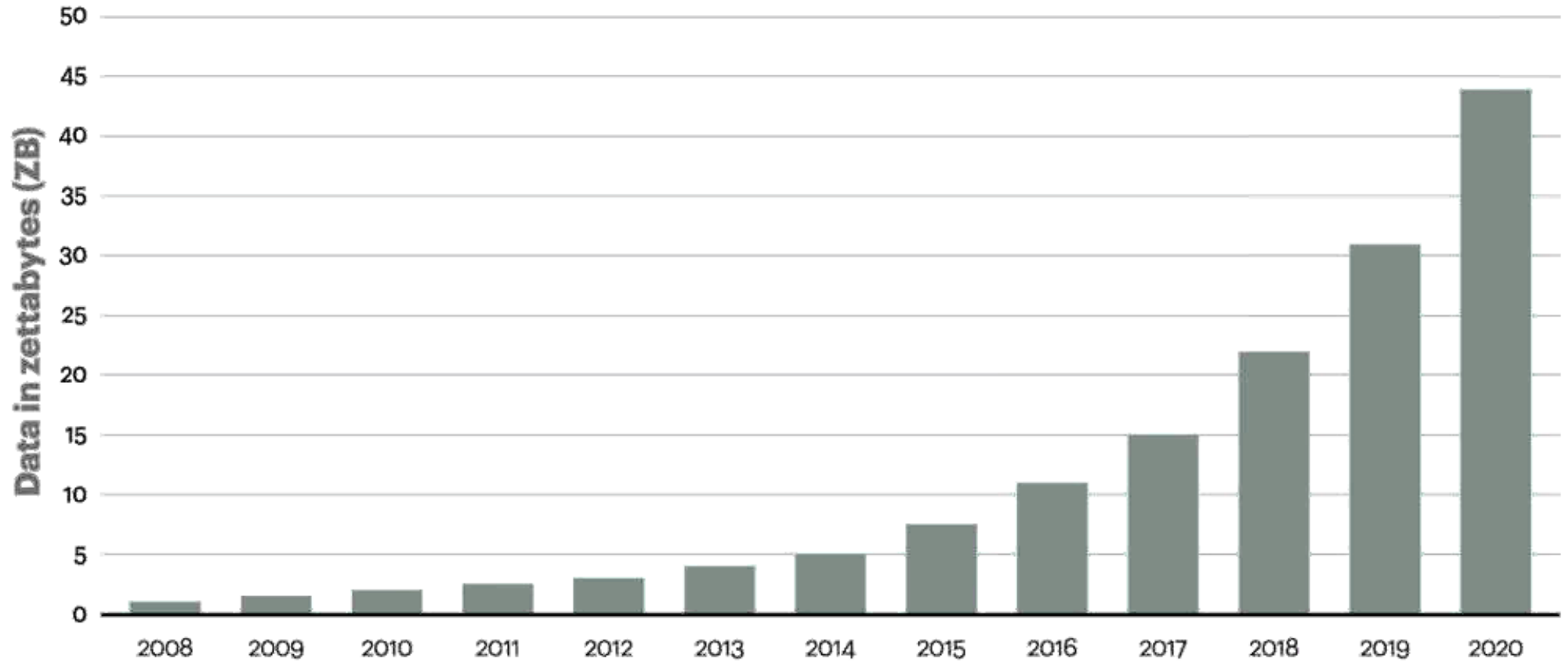
- Plain text, etc.

3. Semi-structured Data

- Tagged photos, etc.

Era of Big Data

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020

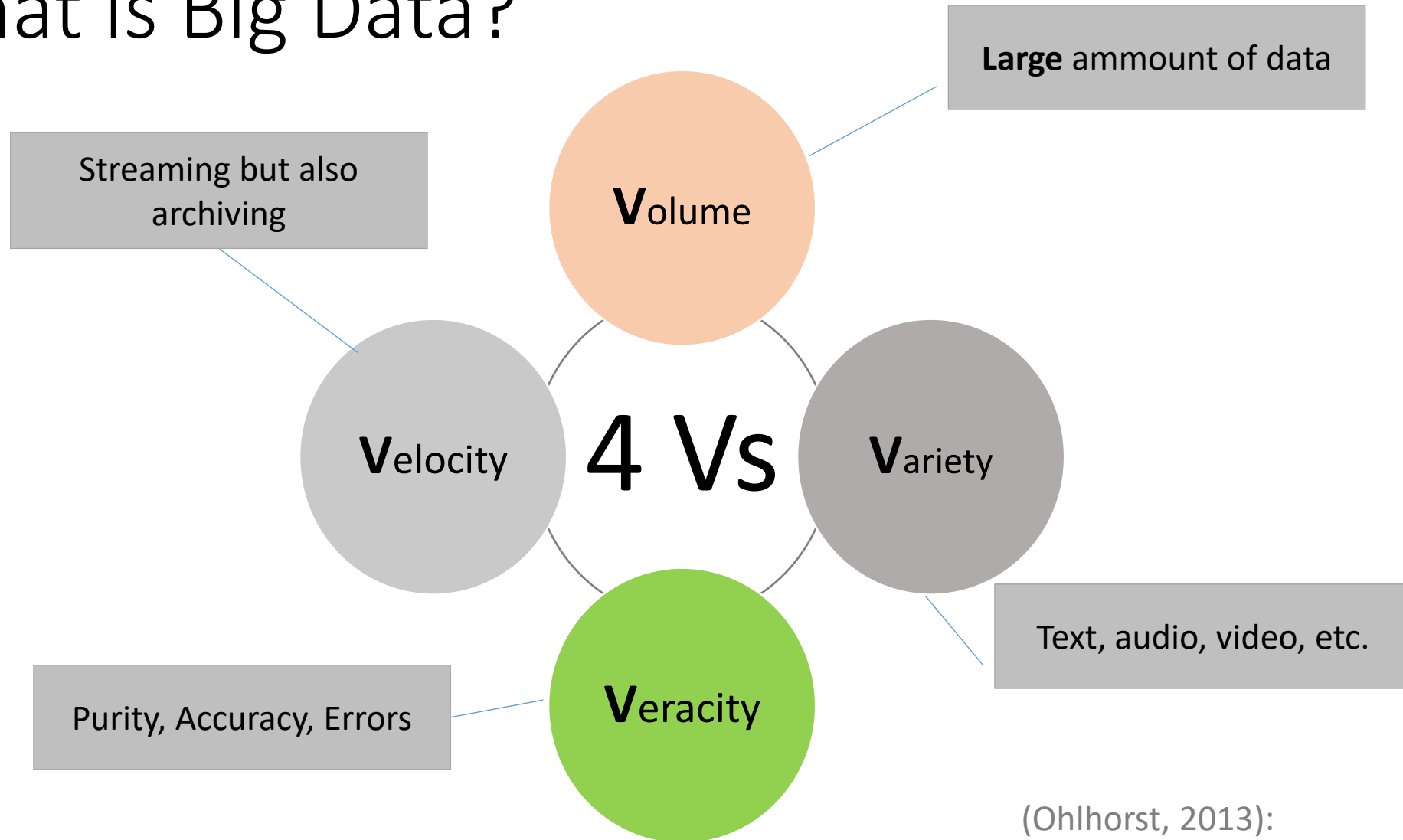


Source: Oracle, 2012

Big Data: Definition

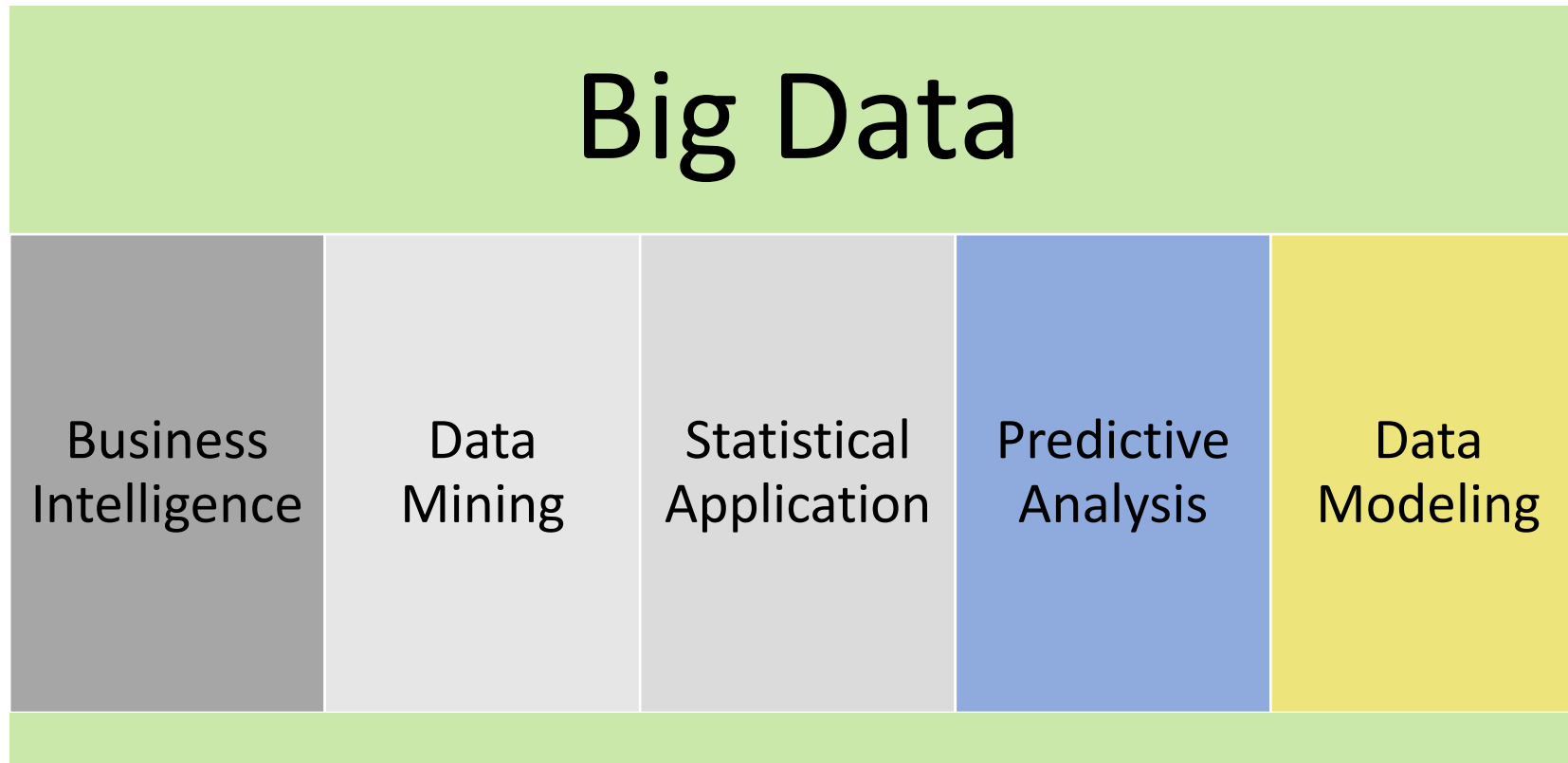
- "Big Data defines a situation in which data sets have grown to such **enormous sizes** that conventional information technologies can **no longer effectively handle** either size of the data set or the scale and growth of the data set." (Ohlhorst, 2013)

What is Big Data?



(Ohlhorst, 2013):

What is Big Data Analytics?



Big Data Analytics

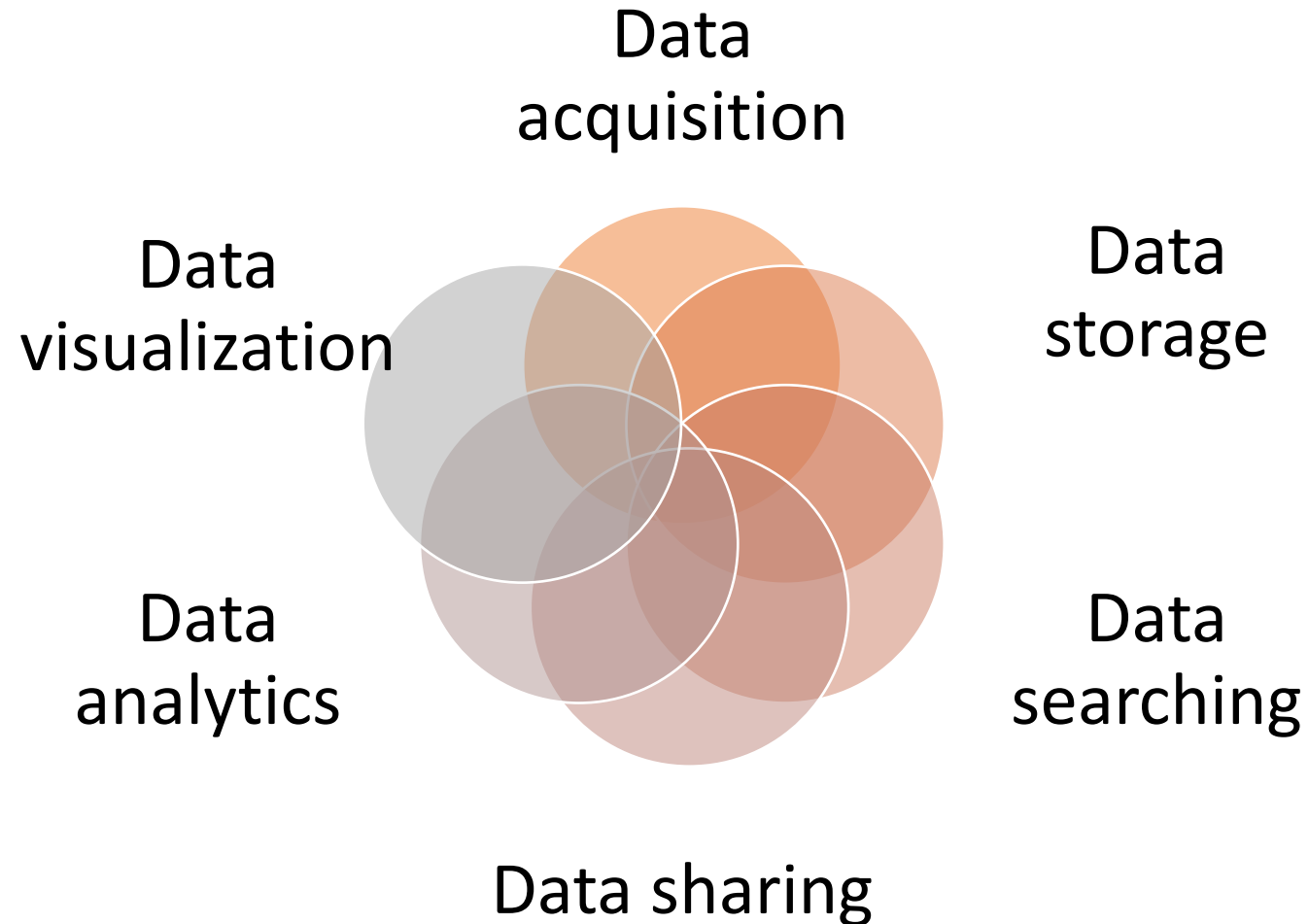


Source: [Link Analytics](#)

Big Data Analytics: Aspects

- Mayer-Schönberger and Cukier (2013)
 1. Analyze **vast amount** of data
 2. Data's real-world **messiness**
 3. Growing respect for **correlations**

Big Data Analytics: Challenges



Massive Data Analysis: Challenges

Frontiers in massive data analysis: The National Academies Press Washington, DC. (p.2)

- Dealing with highly distributed data sources
- Tracking data provenance, from data generation through data preparation
- Validating data
- Coping with sampling biases and heterogeneity
- Working with different data formats and structures
- Developing algorithms that exploit parallel and distributed architectures
- Ensuring data integrity
- Ensuring data security
- Enabling data discovery and integration
- Enabling data sharing
- Developing methods for visualizing massive data
- Developing scalable and incremental algorithms
- Coping with the need for real-time analysis and decision-making

Massive Data Analysis: Datacenters

Frontiers in massive data analysis: The National Academies Press Washington, DC. (p.27)

- Centralization of large-scale infrastructure
- Data sizes at the petabyte (10^{15} bytes) to exabyte (10^{18} bytes) scale
- Architectures that are highly fault-tolerant
- Computing that is collocated with the data for large-scale analytics

Massive Data Analysis: Trends

Frontiers in massive data analysis: The National Academies Press Washington, DC. (p.27)

- Distributed systems (access, federation, linking, etc.)
- Technologies (MapReduce algorithms, cloud computing, Workflow, etc.)
- Scalable infrastructures for data- and compute-intensive applications
- Service-oriented architectures
- Ontologies, models for information representation
- Scalable database systems with different underlying models (relation to triple stores)
- Federated data security mechanisms
- Technologies for moving large data sets

Six Provocations for Big Data (Boyd & Crawford, 2011)

1. Automating Research Changes the Definition of Knowledge
2. Claims to Objectivity and Accuracy are Misleading
3. Bigger Data are Not Always Better Data
4. Not All Data Are Equivalent
5. Just Because it is Accessible Doesn't Make it Ethical
6. Limited Access to Big Data Creates New Digital Divides

Typology of Big Data Analytics (Davenport, 2014)

Four Key Dimensions

I. Time

1. *Reporting* stories (past, descriptive analytics)
2. *Explanatory* survey stories (present, explanatory analytics)
3. *Prediction* stories (future, predictive analytics)

III. Depth

7. *CSI* stories (relatively small, ad hoc investigations)
8. *Eureka* stories (long, analytically-driven searches)

II. Focus

4. *What* stories (past, descriptive analytics)
5. *Why* stories (past+present, explanatory analytics)
6. *How to address the issue* stories (past+present+future, prescriptive analytics)

IV. Methods

9. *Correlation* stories (associations)
10. *Causal* stories (mechanisms)

Why does Big Data Analytics matter?

amazon.com[®]

Recommends products..



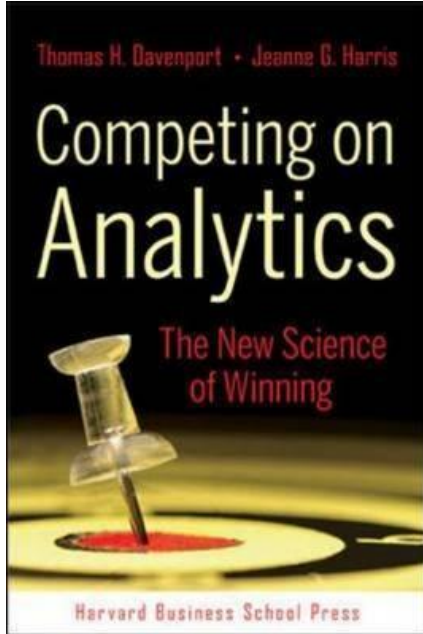
Suggests friends, products, etc.

Google

Suggests search terms, websites, etc.

The process of discovering the 5th V: **VALUE**

Understanding value-adding



“Do we think or do we know?”

Performance of any company today arises from a complex set of organizational and human dynamics.

- Understanding and making sense of this complexity is only the first step.

Analyzing and predicting the effect of today’s complexity on your tomorrow’s performance is THE MUST for building differentiated competitive advantage.

Making better evidence-based decisions

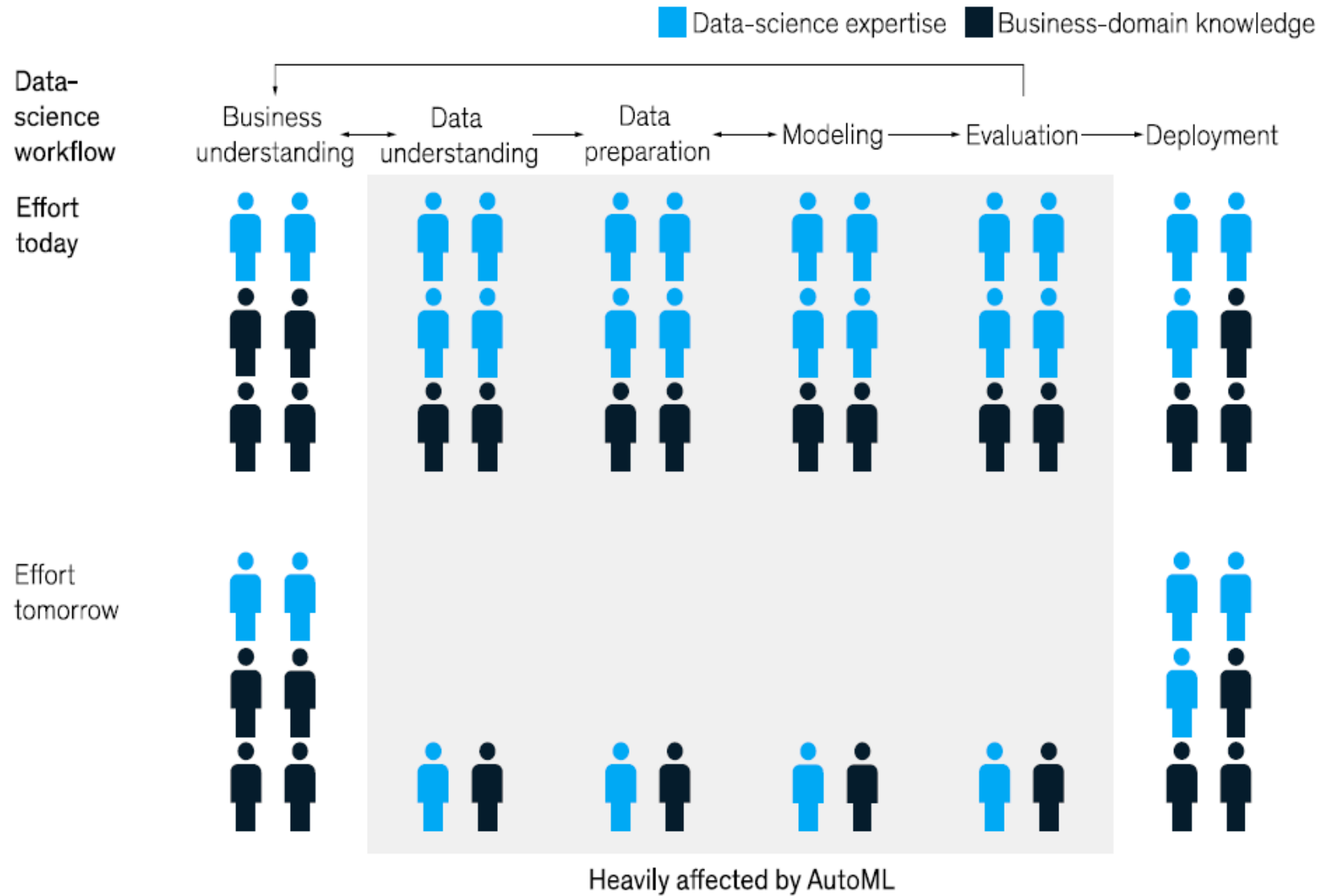
Most companies have accumulated a surfeit of business data: customer data, performance-related data, current and historic operational data, data from employees' engagement surveys, etc.

Yet, only few can make extensive use of data to drive ***evidence-based management decisions, support business development and transformation.***

Why so?

In addition to the need of owning data of a reasonable quality, developing a solid ground for evidence-based decisions requires having the right people, with the advanced analytical skills and scientific rigor in modeling and interpretation of the results.

AutoML changes the mix of talent needed.



Source: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/rethinking-ai-talent-strategy-as-automated-machine-learning-comes-of-age>

Our course

The course provides participants with *a deep understanding* of the nature of business data analytics, and a *practical toolkit* on how to transform internal (in-house) and external big datasets into business assets.

Our focus is on **exploring and exploiting** different strategic and tactical frameworks to **generate business value** from in-house and external big data sets for increasing competitiveness in the local and global marketplaces.

Definitions

Gartner IT Glossary: '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.'

- IBM introduced *veracity*, which refers to the unreliability, impreciseness, and uncertainty inherent in some sources of data.
- SAS introduced *variability* and *complexity* as two additional dimensions of Big Data. Variability refers to data flow rates that are inconsistent, and have periodic peaks and troughs. Complexity refers to the fact that Big Data is generated from a myriad of sources.
- Oracle introduced *value*, especially low value, as a defining attribute of Big Data. In its original form, Big Data has little value relative to its volume.

Data

Structured Data

- Database (e.g. SQL), etc.

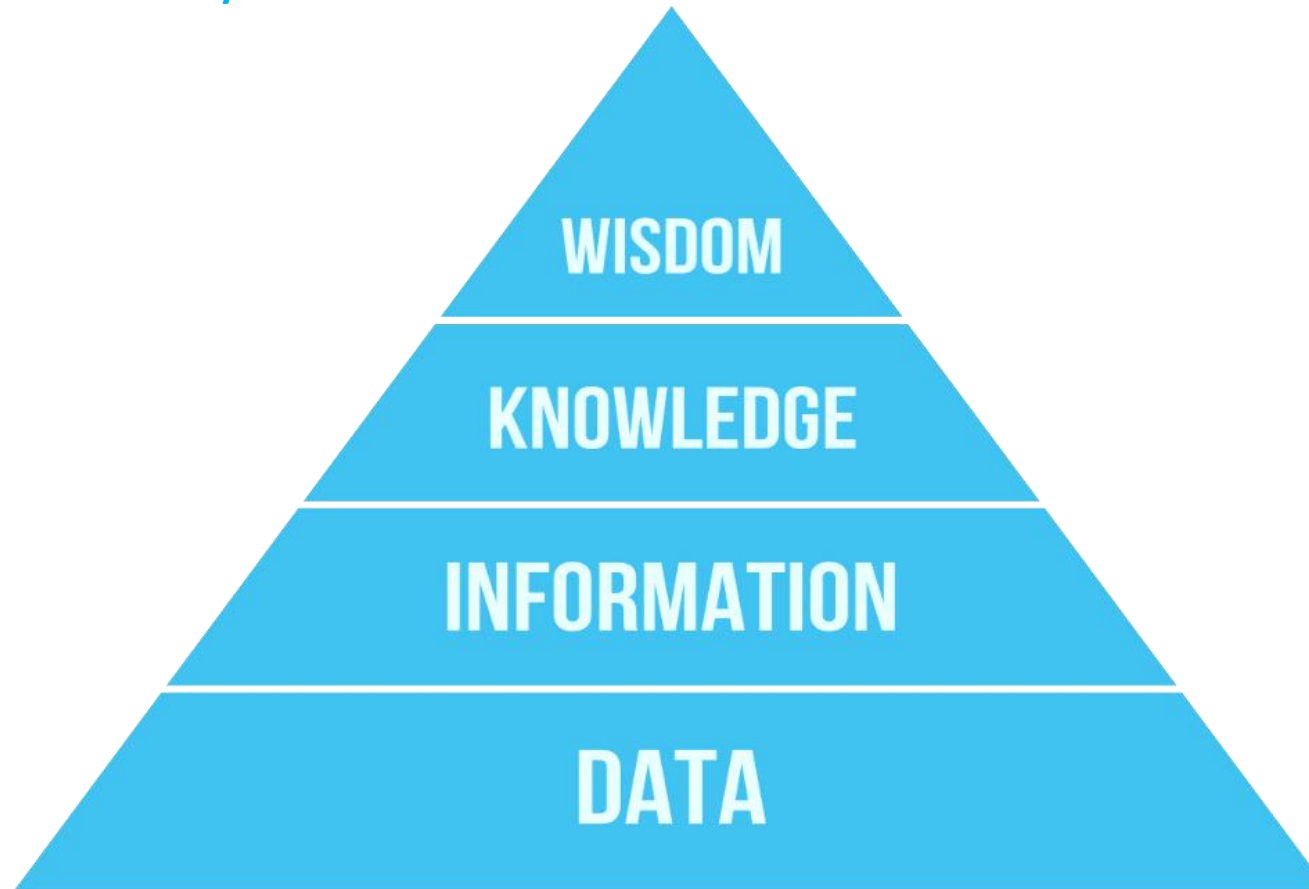
Unstructured Data

- Plain text, etc.

Semi-structured Data

- Tagged photos, etc.

DIKW Pyramid



Source: https://en.wikipedia.org/wiki/DIKW_pyramid

Programs = Algorithms + Data Structures

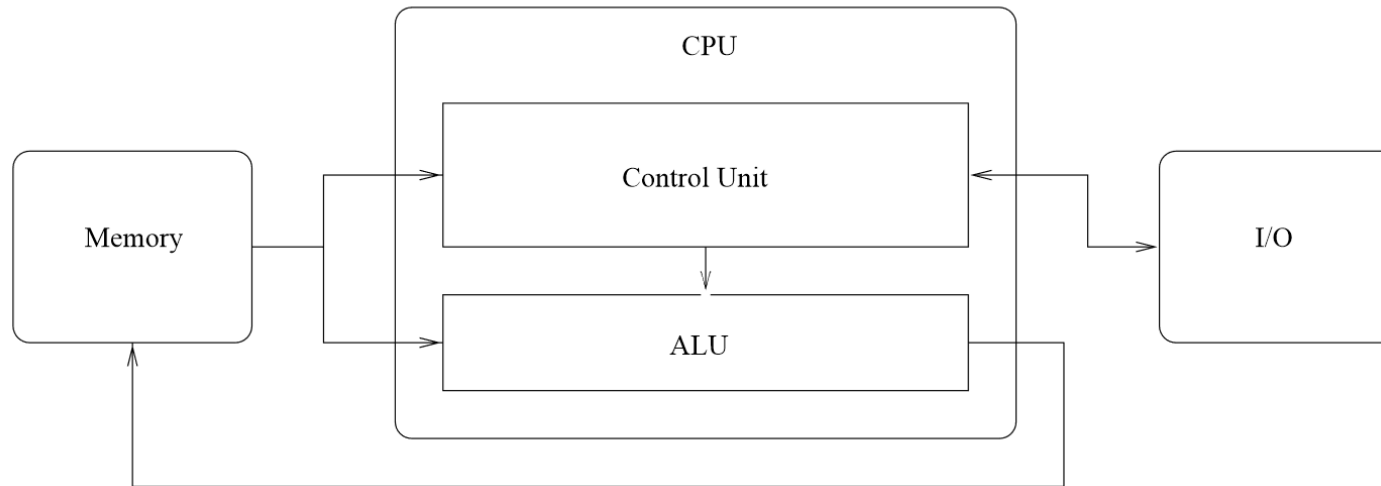
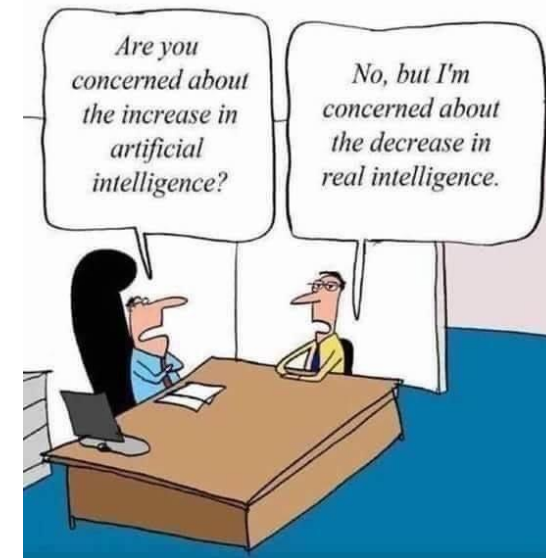


Figure 2: The basic components of a computer with a von Neumann architecture are the memory, which stores both instructions and data, the Central Processing Unit (CPU), which actually executes the instructions, and the Input/Output (I/O) devices, which provide an interface to the outside world.

<https://engineering.purdue.edu/~eigenman/reports/vN.pdf>

Algorithms need managers too

- Algorithms are extremely literal
 - <https://www.youtube.com/watch?v=hI9GyhX7yHM&t=24s>
 - Algorithms focus on the data at hand—which often pertains to short-term outcomes.
- Algorithms are black boxes
 - Garbage in, garbage out
 - Incomplete information
- First WHY, then WHAT and HOW



Group Work

Big Data vs. Business Data

Big Data Analytics vs. Business Data Analytics

Combining Big Data and Business Data Analytics

Discussion