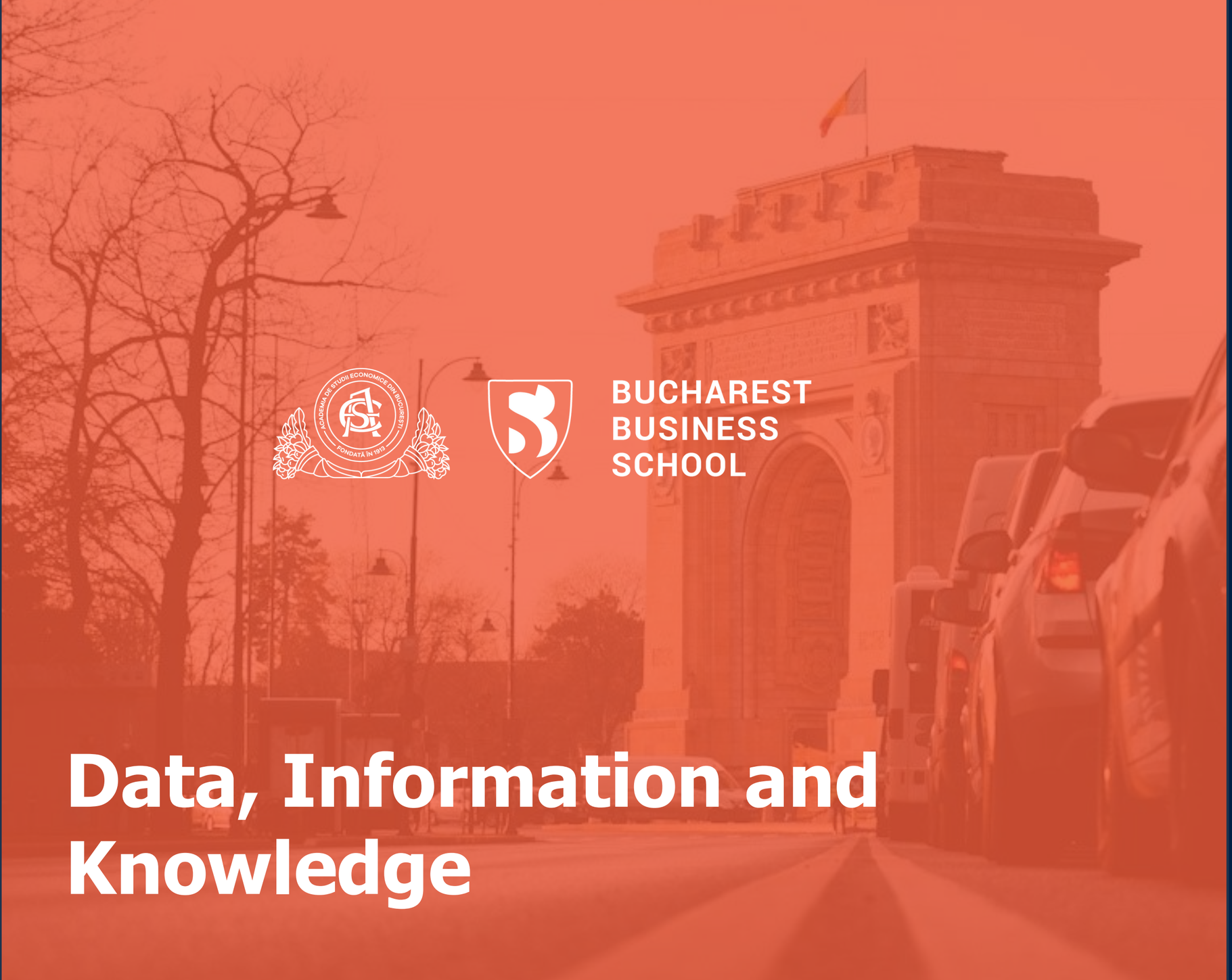


BUCHAREST BUSINESS SCHOOL

ASSOCIATION OF MBAS ACCREDITED

Leadership Through Value

Data, Information and Knowledge

DATA MINING AND BIG DATA

CATALINA CHINIE
PH.D. ASSOCIATE PROFESSOR

SUMMARY

Data Mining and Big Data

Block 1

- The value of the information
- Big data and data economy

Grade structure

Data Mining and Big Data

- 10% grant point
- 10% attendance
- 30% seminar exercises
- 50% final project assessment

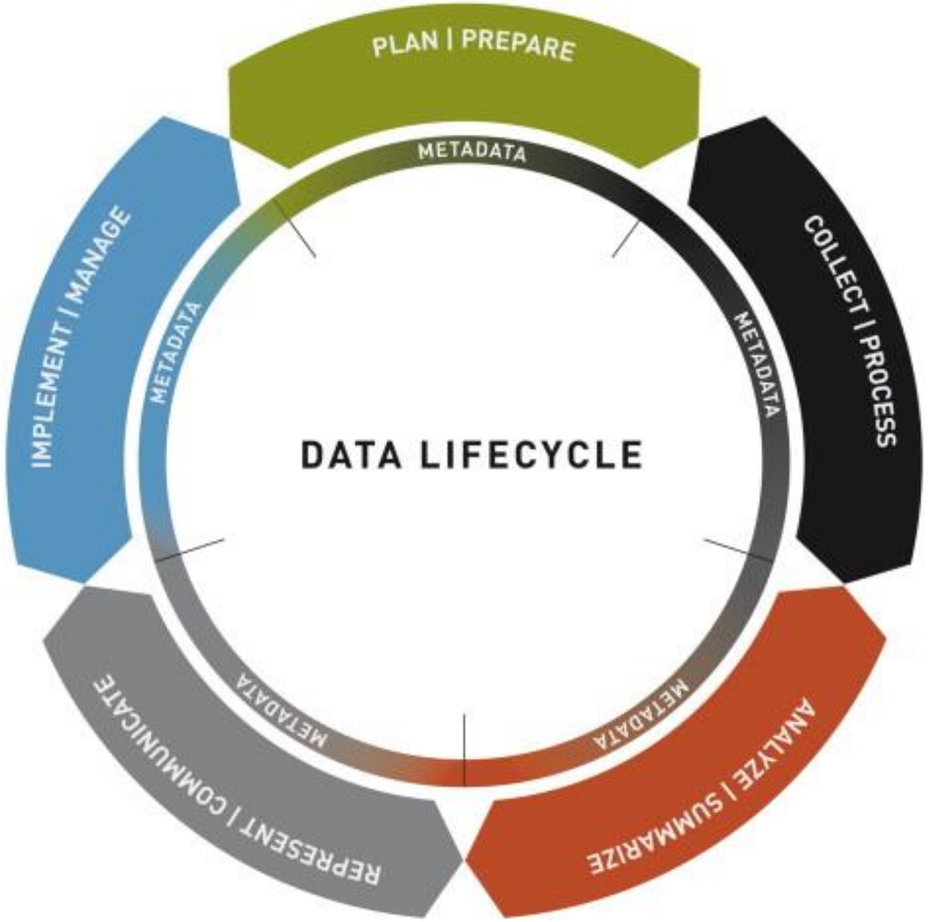
"In this world of big data, basic data literacy—the ability to analyze, interpret, and even question data—is an increasingly valuable skill"

Janice Hammond, HBR

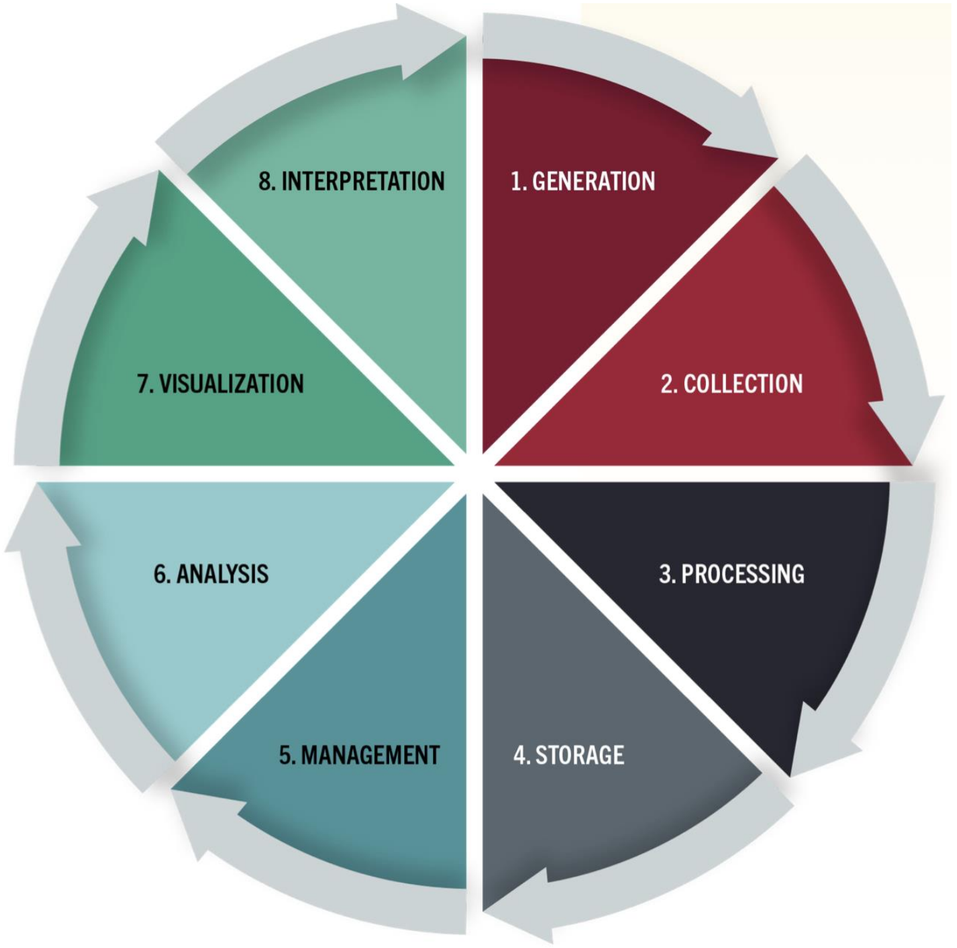
Value of information

Information results from data processing must have high level of quality. Why? Because it helps for taking decisions. Value derives from usefulness information

- *Completeness*
- *Cost-effectiveness*
- *Flexibility*
- *Relevance*
- *Accuracy*
- *Timeliness*
- *Selectivity*
- *Destination*
- *Simplicity*
- *Verifiability*

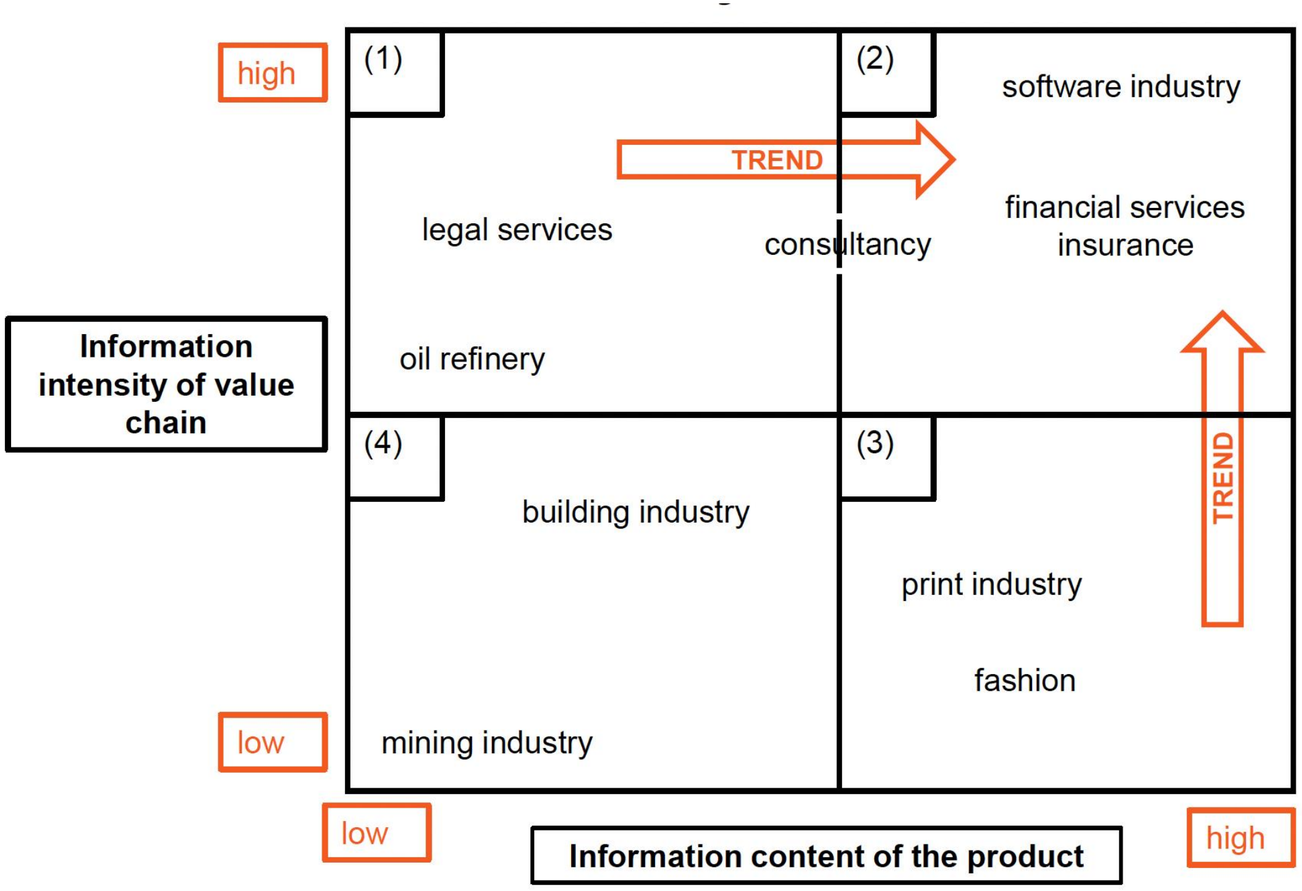


Hunter Whitney, 2013



Harvard Business School Online

Information Intensity Matrix



Big Data and Data Economy

"Big data is not about the data." – Gary King Harvard University Director, Inst. For Quantitative Social Science

IBM: "Big data refers to massive, complex data sets that traditional data management systems cannot handle. When properly collected, managed and analyzed, big data can help organizations discover new insights and make better business decisions."

Gartner: "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."

Google: "Big data refers to extremely large and diverse collections of structured, unstructured, and semi-structured data that continues to grow exponentially over time. These datasets are so huge and complex in volume, velocity, and variety, that traditional data management systems cannot store, process, and analyze them"

5 V'S OF BIG DATA

VOLUME

Massive amount of data being collected.

VELOCITY

High speed at which data is generated and processed.

VARIETY

Diverse formats and types of data.

VERACITY

The required reliability and accuracy of the data

VALUE

The meaningful insights and benefits gained from analysing the data.

What is collecting all this data?

Search Engines

- Microsoft's Internet Explorer
- Mozilla's FireFox
(Non-profit foundation,used to be Netscape)
- Google's Chrome
- Apple's Safari
- Time-Warner's AOL Explorer

- Google's
- Microsoft's
- Yahoo's
- IAC Search's

What is collecting all this data?

Smartphones & Apps

Apple's iPhone (Apple O/S)

Samsung, HTC. Nokia, Motorola (Android O/S)

RIM Corp's Blackberry (BlackBerry O/S)

Tablet Computers & Apps

Apple's iPad

Samsung's Galaxy

Amazon's Kindle Fire

What is collecting all this data?

Games Boxes and GPS Systems

Internet Service Providers

What is collecting all this data?

HDTV's and Blu-Ray Players with built-in Internet connectivity

Movie Rental Sites

NETFLIX

redbox

amazon instant video
Connecting

What is collecting all this data?

Hospitals & Other Medical Systems

Pharmacies

- Laboratories
- Imaging Centers
- Emergency Medical Services (EMS)
- Hospital Information Systems
- Doc-in-a-Box
- Electronic Medical Records
- Blood Banks
- Birth & Death Records

Banking & Phone Systems

verizon

Can you hear me now?
(Heh heh heh!)

Sprint

T-Mobile

at&t

The concept of IoT

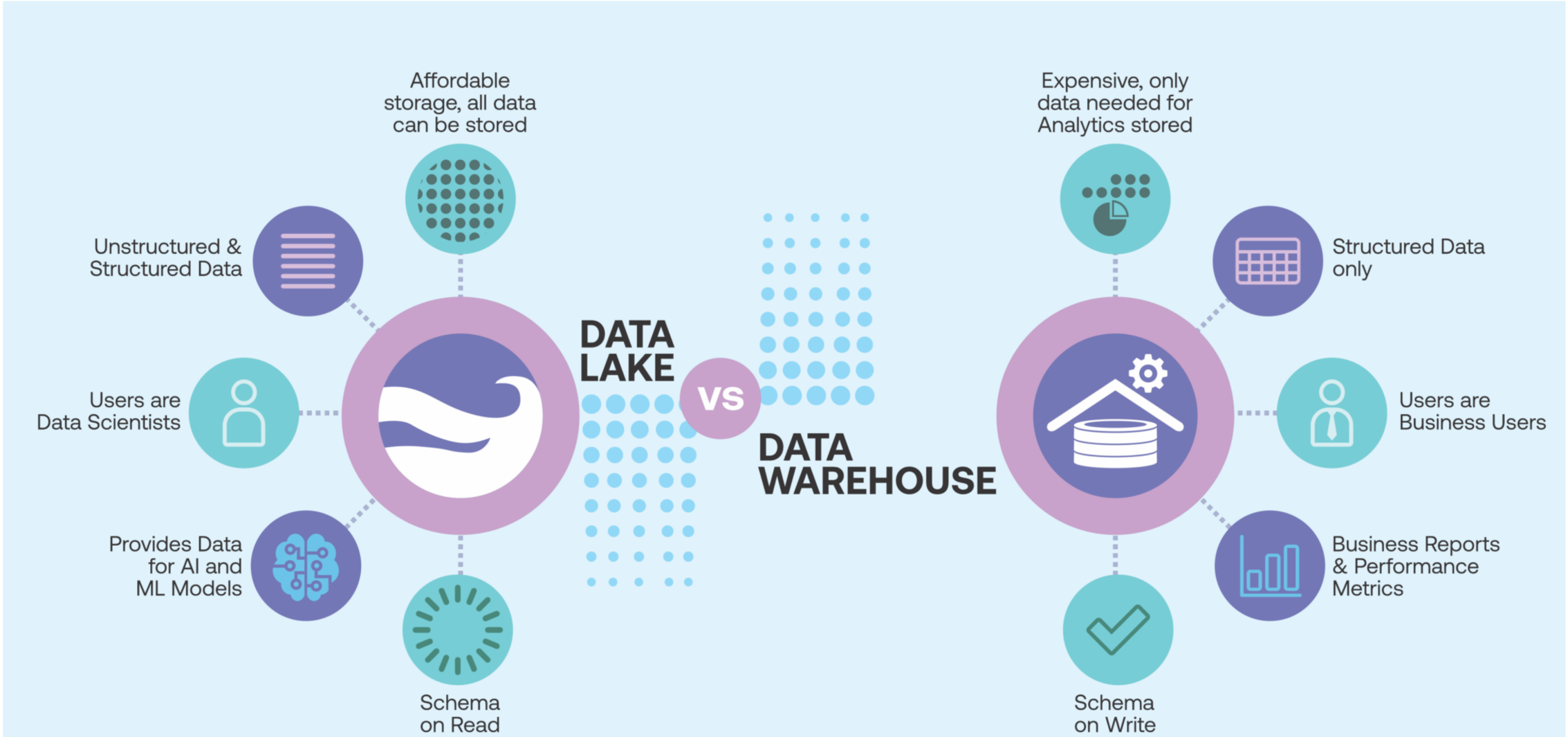
IoT is the network of everyday objects embedded with sensors and connected to the internet to collect and exchange data.

What is a *smart* product/device?

The concept of IoT

Cupcake Conveyor —
An actual cupcake ATM that Davis called a “confectionery 3D printer.” There are Sprinkles’ cupcakes ATMs in several cities in the United States, including Beverly Hills, Chicago, New York and Atlanta.
https://www.youtube.com/watch?v=mgzxHNq8Bek

Data warehouses & data lakes



	Data Warehouse	Data Lake	Data Cloud
Data Type	Structured (e.g., relational data)	Structured, semi-structured, and unstructured (e.g., logs, images)	All data types across multiple platforms and formats
Processing	Batch processing optimized for complex queries	Supports batch and real-time processing	Real-time and batch processing across distributed systems
Storage	Optimized for structured data storage	Scalable storage for vast amounts of raw data	Distributed storage across cloud platforms
Cost	Generally higher due to performance optimization	Lower storage costs; compute costs vary based on usage	Variable; depends on cloud provider pricing and resource utilization
Performance	High performance for structured queries	May require additional processing for unstructured data	Scalable performance; depends on cloud infrastructure and optimization
Scalability	Scales vertically; adding more resources to existing systems	Scales horizontally; adding more systems to distribute load	Elastic scalability across global cloud infrastructure
Security & Governance	Mature security models with robust governance	Evolving security measures; governance can be complex	Advanced security features with centralized governance tools
Primary Users	Business analysts, data engineers	Data scientists, data engineers	Cross-functional teams including analysts, engineers, and data scientists
Use Cases	Business intelligence, reporting, historical data analysis	Big data analytics, machine learning, real-time data processing	Integrated analytics, global data sharing, AI/ML applications
Examples	Amazon Redshift, Google BigQuery, Snowflake	Hadoop, Amazon S3, Azure Data Lake	Snowflake Data Cloud, Google Cloud Platform, Microsoft Azure

Why are companies collecting all this data?

Growth & Innovation: Big data uncovers new business models and revenue sources.

"Companies that embrace data-driven strategies outperform their peers by 5–6% in productivity and profitability. Big data analytics can identify underserved customer segments, enable personalized offerings, and create entirely new business models."

(McKinsey & Company, 2015)

Identify new patterns	Use novel data sources	Adapt dynamically
New algorithms	Social media posts	Spot changing trends
Nonlinear patterns	Product review	Anticipate changes
New methods for identifying clusters	Satellite data	Dynamic pricing
	Sensors data	

Source: After Barton Paulson

Why are companies collecting all this data?

Better Decisions: Predictive analytics reduces uncertainty and supports proactive strategies.

Efficiency Gains: Real-time monitoring and predictive maintenance lower costs and downtime.

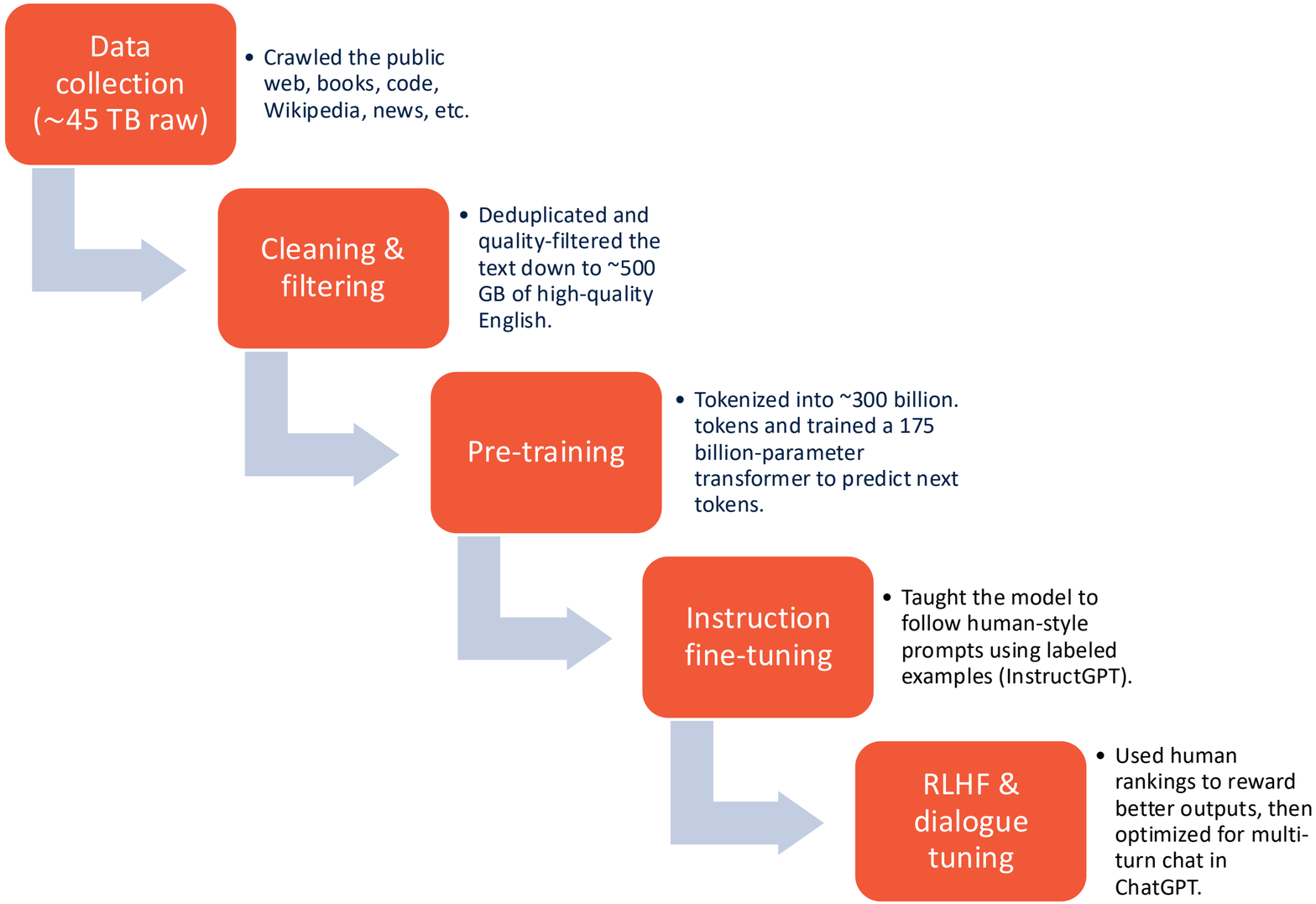
Customer Focus: Personalized experiences boost sales and loyalty.

Risk Management: Analytics-driven surveillance strengthens fraud detection and regulatory adherence.

Big Data makes Generative AI possible

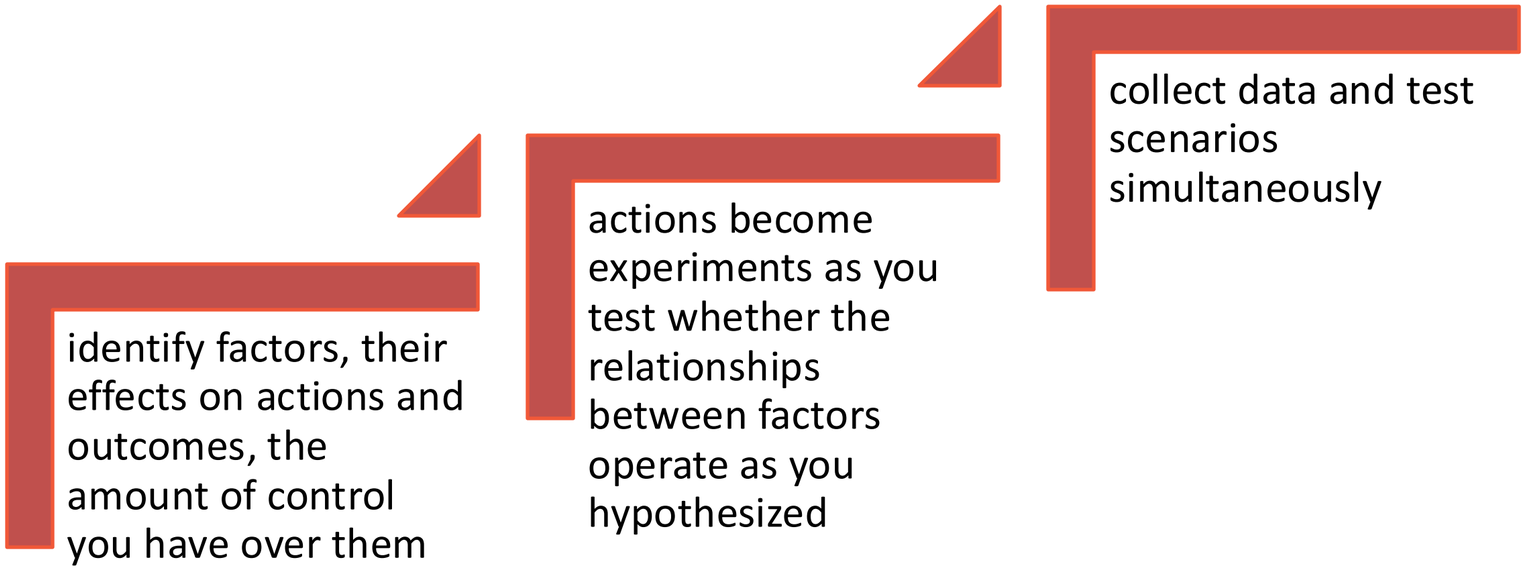
Not our actual logo obviously ☺

How did Big Data contribute to GenAI?

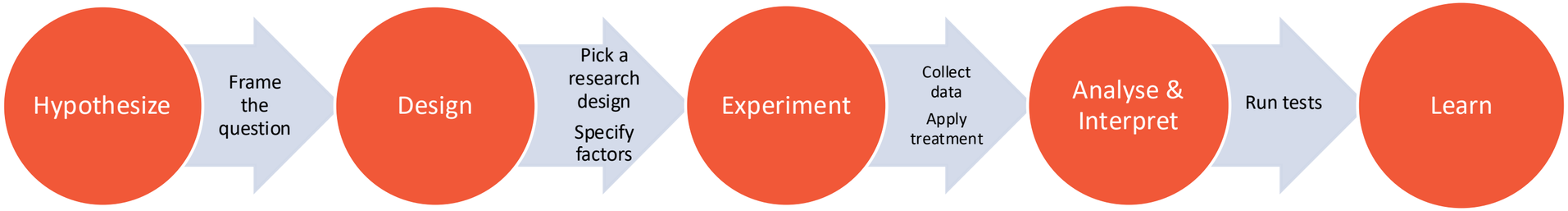


In the information economy, companies employ a Dynamic Customer Strategy

DSC is a skillset and toolset for capturing market value through agile marketing, which includes experimental design, data management, statistics, and more. (Tanner, 2014)



How does Big Data drive personalized experiences for customers?



Why are companies collecting all this data?

Growth & Innovation: Big data uncovers new business models and revenue sources.

"Companies that embrace data-driven strategies outperform their peers by 5–6% in productivity and profitability. Big data analytics can identify underserved customer segments, enable personalized offerings, and create entirely new business models." (McKinsey & Company, 2015)

Better Decisions: Predictive analytics reduces uncertainty and supports proactive strategies.

Efficiency Gains: Real-time monitoring and predictive maintenance lower costs and downtime.

Customer Focus: Personalized experiences boost sales and loyalty.

Risk Management: Analytics-driven surveillance strengthens fraud detection and regulatory adherence.

BUCHAREST
BUSINESS
SCHOOL