MMPA

Day 1

Introduction to Information Systems

MGT 1272 – MANAGEMENT INFORMATION SYSTEMS

Behram Faroogh

Introduction

Professor: Behram Faroogh
behram.faroogh@utoronto.ca
Retired PwC Partner; Vice President at Board of Director for ISACA Toronto; Professor at Seneca College; Guest lecturer at York University, Durham College, Conestoga College; Think Tank member at University of Waterloo

TA: Belal Raifai
belal.raifai@mail.utoronto.ca

All of You: tell us a bit about yourself: your goals, fears, accomplishments and aspirations

Course Schedule

Session	Date		Delivery	Topic
1	Thu	Sep 5, 2024	In person	Introduction
2	Wed	Sep 11, 2024	In person	Data Analytics
3	Fri	Sep 13, 2024	In person	Data Wrangling
4	Wed	Sep 18, 2024	In person	Data Analyst for a Day
5	Mon	Sep 23, 2024	Online	Practical Assignment
6	Thu	Nov 14, 2024	In person	Business Process Life Cycles
7	Fri	Nov 22, 2024	In person	Governance, Risk & Controls
8	Wed	Nov 27, 2024	In person	Business Analyst for a Day
9	Fri	Nov 29, 2024	In person	Group Project & Presentation
10	Mon	Dec 16, 2024	In person	Final Exam

Introduction to Information Systems

Key Dates to Remember

Date	Evaluation	Weight
Sep 18, 2024	Quiz 1 - Data Analyst	10%
Sep 23, 2024	Practical Assignment	20%
Nov 27, 2024	Quiz 2 - Business Analyst	10%
Nov 29, 2024	Group Project & Presentation	20%
Dec 16, 2024	Final Examination	40%

What will we learn today?

1. Information Systems Engineering
2. Information Systems Classification
3. Information Systems Infrastructure
4. Information Systems Architecture
5. All about Data

1. Information Systems Engineering

Information Systems Engineering

Information Systems Engineering is a data-oriented methodology to create and maintain a conceptual blueprint of any organization's desired structure for Information Systems. It is a top-down generic planning methodology for aligning business objectives and strategies with technology to obtain a broad understanding of the Information Systems needs of the entire organization. **Planning** involves 3 steps:

1. Identify strategic planning factors
   a. Organizational Goals
   b. Critical Success Factors (CSFs)
   c. Problematic areas
2. Identify corporate planning structure
   a. Organizational units
   b. Organizational locations
   c. Business functions
   d. Entity types
3. Develop an Enterprise Model
   a. Functional Decomposition
   b. Entity-Relationship Diagram
   c. Planning matrix

Information Systems Planning

1. Identify strategic planning factors

a. Organizational Goals – what we HOPE to accomplish

b. Critical Success Factors (CSFs) – what MUST work in order for us to survive

c. Problematic areas – what challenges we HAVE as identified by our SWOT analysis

Strength . . .	Weaknesses . . .
Opportunities . . .	Threats . . .

Assess environment
for opportunities or threats

↓

Develop objectives (goals)
based on a SWOT analysis

↓

Define
Critical Success Factors
CSFs

Information Systems Planning

2. Identify corporate planning structure

a. Organizational units
– departments, cost centers, profit centers, etc.

b. Organizational locations
– by city, state, region, country, etc.

c. Business functions
– distinct groups of business processes

d. Entity types
– the things we are trying to model

Define
Critical Success Factors
CSFs

↓

Define
Corporate Strategy
to achieve goals and CSFs

↓

Implement
Business Processes
tied to Org. Structure

↓

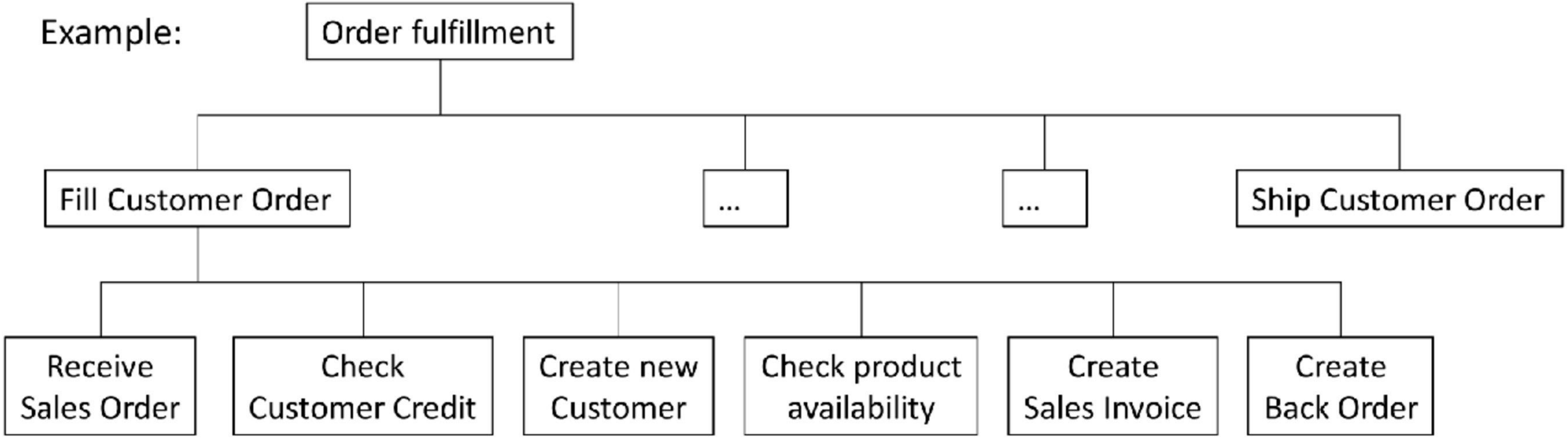
Identify
Key Performance Indicators (KPIs)
to demonstrate achievement of CSFs

Information Systems Planning

3. Develop an Enterprise Model

a. Functional decomposition – breaking large tasks into smaller tasks.
   It is an iterative process, breaking system description down into fine details.

Example:



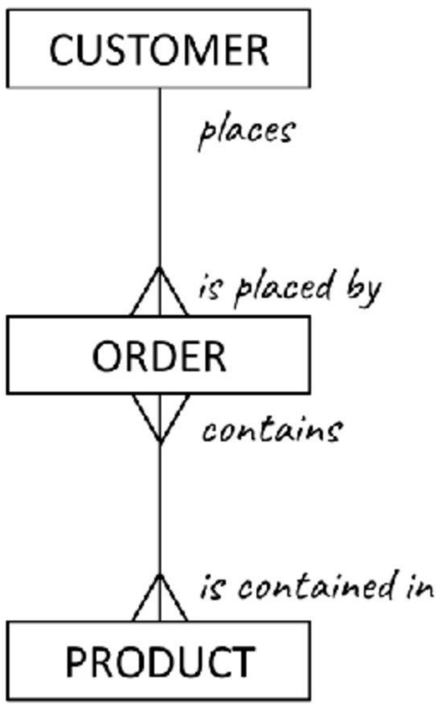
Information Systems Planning

3. Develop an Enterprise Model

b. Entity-Relationship (ER) Diagram – shows relationship between different entities.

Example: a CUSTOMER places multiple ORDERs
multiple ORDERs may be placed by a CUSTOMER

an ORDER may contain multiple PRODUCTs
a PRODUCT may be in multiple ORDERs



Information Systems Planning

3. Develop an Enterprise Model

c. Planning matrix – describes relationship between planning objects in an organization. The matrix can be for anything being planned for.

Example:

- Business Functions to Data Entity Types

Business Functions \ Data Entity Types	Customer	Product	Raw Material	Order	Work Center	Work Order	Invoice	Equipment	Employee
Business Planning	x	x						x	x
Product Development		x	x		x			x	
Materials Management		x	x	x	x	x		x	
Order Fulfillment	x	x	x	x	x	x	x	x	x
Order Shipment	x	x		x	x		x		x
Sales Summarization	x	x		x			x		x
Production Operations		x	x	x	x	x		x	x
Finance and Accounting	x	x	x	x	x		x	x	x

Other Examples:

- Business Locations to Business Functions
- Business Units to Business Functions
- etc.

2. Information Systems Classification

Accounting Information Systems

An **Accounting Information System (AIS)** is used by companies to collect, store, manage, process, and report financial data. AIS can be used by accountants, consultants, business analysts, managers, chief financial officers, auditors, and regulators.

The internal and external controls of an AIS are critical to protecting a company's sensitive data. A good AIS uses hardware and software to effectively store and retrieve data. It helps the different departments within a company work together.

Following are 10 key elements of AIS:

1. Technology
2. Databases
3. Reporting
4. Control
5. Business Operations
6. Events Processing
7. Management Decision Making
8. Systems Development and Operation
9. Communications
10. Accounting and Auditing Principles

Accounting Information Systems

10 Key Elements

1. Technology – provides core foundation; critically important
2. Databases – imperative for data collection, storage and retrieval
3. Reporting – data analytics, report writers and ad-hoc queries
4. Control – making sure the intended actually happens
5. Business Operations – process whereby work is performed, e,g, hire employees, purchase goods, etc.
6. Events Processing – recording a ‘transaction’ when an event occurs in business operation, e.g. sales
7. Management Decision Making – information used must be tailored to both the audience and decision type
8. Systems Development and Operation – design, implement, operate
9. Communications – ability to communicate effectively
10. Accounting and Auditing Principles – must adhere to proper accounting standards, policies and procedures; and must understand the audits to which the accounting information will be subjected

Database Management Systems (DBMS)

A **Database Management System (DBMS)** is at the heart of AIS.

Businesses typically use a **Relational DMBS (RDBMS)** where different tables are related to each other based on the Entity-Relationship defined in an ER diagram.

An RDBMS may be identified as under:

- Data in RDBMS is logically organized into 2-dimensional tables. The tables themselves fall into two categories: (i) master data tables, and (ii) transactional detail lines tables.
- An RDBMS is able to handle complex queries.
- An RDBMS traditionally allowed only for text and numerical data to be stored.
- An RDBMS typically does not like inclusion of complex objects, such as graphics, audio and video; although it will happily point to it.

Relational Database Management Systems (RDBMS)

Core Elements

- Tables - a place to store data
- Queries - a way to retrieve data
- Forms - on-screen presentation of data
- Reports - printed lists and data summaries

Transactional Processing Systems

A **Transactional Processing System** is nothing more than a RDBMS that includes the 10 Elements of AIS. These are more commonly known as ERPs - where **ERP** stands for **Enterprise Resource Planning**.

Some of the well-known ERPs are from vendors like: SAP, Microsoft, Oracle, Workday, etc.

ERPs are characterized as under:

- ERP does **end-to-end** information management for an organization.
- ERP is driven by an **integrated** suite of software modules that support the basic internal processes of the company.
- ERP is **customizable**, extendible and tailored to specific industries.
- ERP contains **shared master data** across all applications, such as for: Customers, Vendors, Employees, etc.
- Within an ERP, each business function/department has their own module, and can **share information** with the rest of the company's business modules.

Decision Support Systems (DSS)

What makes up a DSS?

- Systems that assist management with decision making
- Systems that possess interactive capabilities
- Systems that can answer ad-hoc queries
- Systems that provide data modeling faculties, e.g. spreadsheets
- Systems that support non-recurring, relatively unstructured decision making

Executive Information Systems (EIS)

What constitutes EIS?

- Information Systems, often considered a subset of DSS, that combine information gathered from the organization and the environment.
- Organizes and analyzes the information in a form suitable for management to make appropriate and timely business decisions

3. Information Systems Infrastructure

Information Systems Infrastructure

There are several modalities for setting up an Information Systems Infrastructure. When doing so, one must decide whether to:

- **Delegate:**
  Inhouse (using your own employees), OR
  Outsource (using the best available talent)

- **Relegate:**
  On-shore (contributing to local employment), OR
  Off-shore (taking advantage of cheap labor in distant land, e.g. China, India, etc.), OR
  Near-shore (take advantage of cheap labor closer to home)

Information Systems Infrastructure

Cloud Computing

Cloud Computing refers to accessing your programs and storing your data **over the internet** or having computing delivered to you like a utility. It is **location-independent** delivery of IT management services that provides significant evolution of IT service delivery to organizations. Cloud-based service delivery empowers organizations to **simplify** their infrastructure and potentially reduce costs via standardized platforms, combined with new skills and management practices.

Today most ERPs are resident in the cloud, not on customer premise. This results in significant savings for IT infrastructure and staffing; as it also avoids the common problem of under-utilization of IT staff.

Key features of ERP on Cloud:

- Subscription-based pricing; no upfront cost
- Always on the current version; no need to upgrade
- Instant access to computing power as and when needed
- No customization required whatsoever
- Rapid deployment; only need web browser and high-speed internet connection

Information Systems Infrastructure

Customer Concerns with Cloud Computing

- **Security**
  data privacy and user confidentiality is threatened, especially if data is hosted in foreign countries
- **Reliability**
  lack of control with system failures and performance issues with poor internet connectivity
- **Synergy**
  substantial investment may be required to make existing systems cloud ready
- **Customization**
  multi-tenancy disallows software customization, as other organizations may also be on same system
- **Transparency**
  limited information about cloud service provider’s operations, processes and controls

4. Information Systems Architecture

Information Systems Architecture

Delivery Models for Cloud Computing

- **SaaS**
  in a SaaS (Software-as-a-Service) model, the cloud service provider controls the (i) hardware, (ii) network, (iii) operating system, and (iv) business applications. Customers access applications remotely.

- **PaaS**
  in a PaaS (Platform-as-a-Service) model, the cloud service provider controls the (i) hardware, (ii) network, and (iii) operating system. Customers provide access applications remotely.

- **IaaS**
  in an IaaS (Infrastructure-as-a-Service) model, the cloud service provider controls only the (i) hardware, and (ii) network. Customers manage and run their own applications and operating systems.

Information Systems Architecture

Deployment Models for Cloud Computing

- **Public** Cloud – deployments pass the responsibility and associated risk for an IT infrastructure from the business to a third-party in this model.
- **Private** Cloud – is a popular choice for large organizations and governments. IT infrastructure is provided virtually via the internet, but from designated facilities, whether owned by the customer or the vendor.
- **Hybrid** Cloud – mix and match of public and private clouds, where private cloud calls certain services from the public cloud.

5. All about Data

Data vs. Information

Data is an accumulation of facts and figures collected in raw form. It represents the measurement of observations of objects and events. To be useful to a decision maker, data must be transformed into Information.

Information is data presented in a form that is useful in a decision making activity. Information holds value for the decision maker because it reduces uncertainty and increases knowledge about a particular area of concern.

While data is an unsystematic fact or detail about something, information is a systematic and filtered form of data, which is useful. When any 'data' is processed and transformed in such a way that it becomes useful to the users, it is known as 'information'.

Thus, Information is described as that form of data which is processed, organized, specific, structured, and presented to a decision maker. It assigns meaning and improves the reliability of the data, thus ensuring understandability and reduces uncertainty. When the data is transformed into information, it is free from unnecessary details or immaterial things, which then makes it valuable to the information consumer.

The function of MIS is to capture and transform data into useful information.

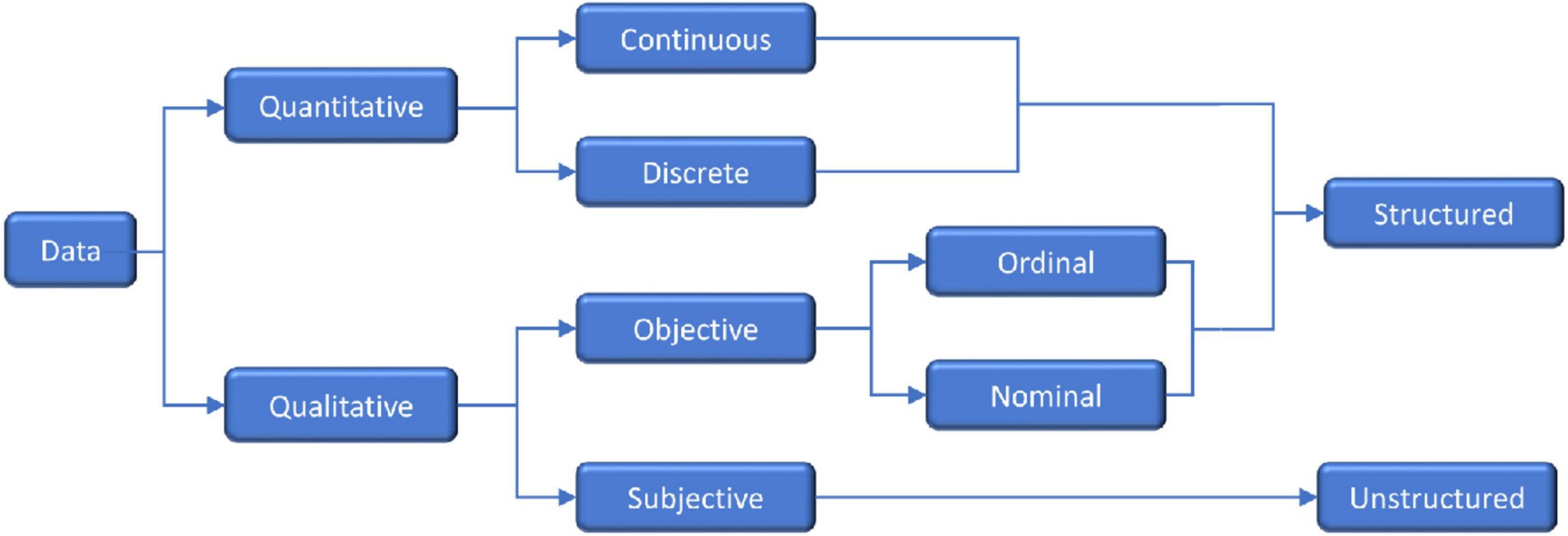
Data vs. Information

Key Differences

- Raw facts gathered about a condition, event, idea, entity or anything else which is bare and random, is called data. Information refers to facts concerning a particular event or subject, which are refined by processing.
- Data is simple text and numbers, while information is processed and interpreted data.
- Data is in an unorganized form, i.e. it is randomly collected facts and figures which are processed to draw conclusions. On the other hand, when the data is organized, it becomes information, which presents data in a better way and gives meaning to it.
- Data is based on observations and records, which are stored in computers or simply remembered by a person. As against this, information is considered more reliable than data, as a proper analysis is conducted to convert data into information by a researcher or investigator.
- Data collected by a researcher, may or may not be useful to him, as when the data is gathered, it is not known what the data is about or what it represents? Conversely, information is valuable and useful to the researcher because it is presented in the given context and so readily available to the researcher for use.
- Data is not always specific to the need of the researcher, but information is always specific to his or her requirements and expectations, because all the irrelevant facts and figures are eliminated during the transformation of data into information.
- When it comes to dependency, data does not depend on information. However, information cannot exist without data.

Data

Data is facts, figures and statistics that are collected together and used as a basis for calculation, reasoning or discussion. Data is a very broad term and most people categorize it into either Quantitative or Qualitative. Further classification may vary depending on whom you are talking to. For our intent, we shall classify it as under:



Data Classification

Quantitative – anything that can be expressed numerically, e.g. Count, Average, Sum, etc.

- **Continuous** – any number along a sliding scale or continuous range, e.g. when you say a person is 6' tall … the number 6 here represents a data point along an expected continuous range or sliding scale. Obviously you're not going to have a 40' tall person, but a person's height could be measured in any of an infinite number of decimals anywhere along a sliding scale within an expected range of say 4' to 10'.
- **Discrete** – any number within a finite number of choices, e.g. when you ask how many wheels are on a vehicle … the answer will generally be 2 for a scooter, 3 for a rickshaw, 4 for a car and 6 for a dump truck. The key here is: when you have a finite number of choices, it becomes discrete data.

Structured Data

Data Classification

- **Qualitative** – anything that is expressed non-numerically, e.g. an observation or a piece of news.
  - **Objective** – any observation from a set of data parameters.
    - **Ordinal** – any thing from a logical sequence of possible answers defined in a specific order, e.g. Small, Medium, Large; or Happy, Neutral, Sad.
    - **Nominal** – any thing from a random pile of possible answers, e.g. the type of pets in my city … the answer could be Dog, Cat, Goldfish, etc.

    (Objective, Ordinal, Nominal) → Structured Data
  - **Subjective** – a piece of information that doesn't collate neatly, e.g. when a fast-food restaurant asks for feedback on its burgers in a 'comment questionnaire', people are going to respond with a variety of pieces of information. It is still data but not something that can be easily grouped or parsed out.

    (Subjective) → Unstructured Data

Unstructured Data

Most often referred to as qualitative data, unstructured data is usually subjective opinions and judgments in the form of text, audio, video, picture, file, etc. which most data collection software can't collate. This makes unstructured data difficult to gather, store, and organize in a typical ERP system. It is also difficult to examine unstructured data with standard data analysis methods and tools like regression analysis and pivot tables.

Since you can't use standard data analysis methods and tools to pull insights from unstructured data, you can either manually analyze it or use more advanced data analytics tools (e.g. Alteryx, Power Query, etc.) to examine unstructured data. This requires you to have a high degree of technical expertise. We will examine Power Query on Day 3.

Key Attributes of Unstructured Data:

- Qualitative in nature – is primarily descriptive rather than numerical (e.g. emails, surveys, etc.)
- Subjective in opinion – is judgmental and can be non fact-based (e.g. blogs, news, fake news, etc.)
- Non-restrictive in format – can be shared or provided in any format (e.g. comments, audio, video, etc.)

The Internet is a great source for unstructured data. When a company can successfully extract meaningful insights from unstructured data, it develops a deep understanding of its customer's preferences and their sentiment toward its brand.

Unstructured Data

Examples

- **Survey Responses**
  Any time you gather feedback from anyone, you're collecting unstructured data. Electronic surveys with text responses are a classic example of unstructured data. While such data can not be clearly formatted or collected in a database, it is still valuable information you can use to inform business decisions.

- **Email Responses**
  Similar to survey responses, email responses can also be considered unstructured data. The feedback you give or receive via email is important information, but it can't necessarily be collected in a database.

- **Social Media Comments**
  If you've ever given or received feedback through social media comments, likes, shares, tweets, etc. you've dealt with unstructured data. Such data provides valuable insights into consumer behavior, sentiment and market trends.

- **Web Scraping**
  Web scraping is a common means of capturing data from the web browser. This can be done through search engines, data analysis tools or dedicated software for Robotic Process Automation (*aka* RPA, which we will discuss on Day 7).

- **Phone Call Transcripts**
  Customer service reps are always collecting unstructured data in their phone call conversations with you. Since these calls may include some feedback or critique for their company, it is important information for them to collect.

Structured Data

Most often referred to as quantitative data, structured data is objective facts and figures that most data collection software can collate, making the data easier to collect, import, export, store, and organize in a typical ERP system.

Even though structured data is just numbers or words packed in a database or stored in an ERP system, you can easily extract insights from structured data by running it through data analysis methods and tools like regression analysis and pivot tables. This is the most valuable aspect of structured data.

Key Attributes of Structured Data:

- Quantitative in nature – is primarily made up of numbers and lists of data (e.g. Customer list, Total Sales, etc.)
- Objective in opinion – is fact-based and can be supported with hard evidence (e.g. total sales orders in a day, etc.)
- Restrictive in format – data must be shared or provided in a prescribed format (e.g. ordered lists, data entry forms, etc.)

Structured vs Unstructured Data

The key difference between structured and unstructured data is that:

- **structured** data is objective facts and numbers that can be collected and stored in an orderly fashion, while
- **unstructured** data is usually subjective opinions and judgments that may be easily repudiated.

Other differences include:

Structured Data:

- highly organized and clearly defined data types
- could be included in databases or spreadsheets
- easily searchable by simple search engine algorithms

Unstructured Data:

- usually text heavy
- may contain video, audio, data, numbers or facts
- may be contradictory across different sources

Data Unit

A Bit is the smallest unit of measure for data. It represents either a 1 or 0, for ON or OFF.

A Byte is made up of 8 Bits and it represents any single character that you press on the keyboard, e.g. "a".

Form here on, we measure in multiples of 1024 (2^{10}) bytes. For simplicity sake, we round 1024 down to 1000.

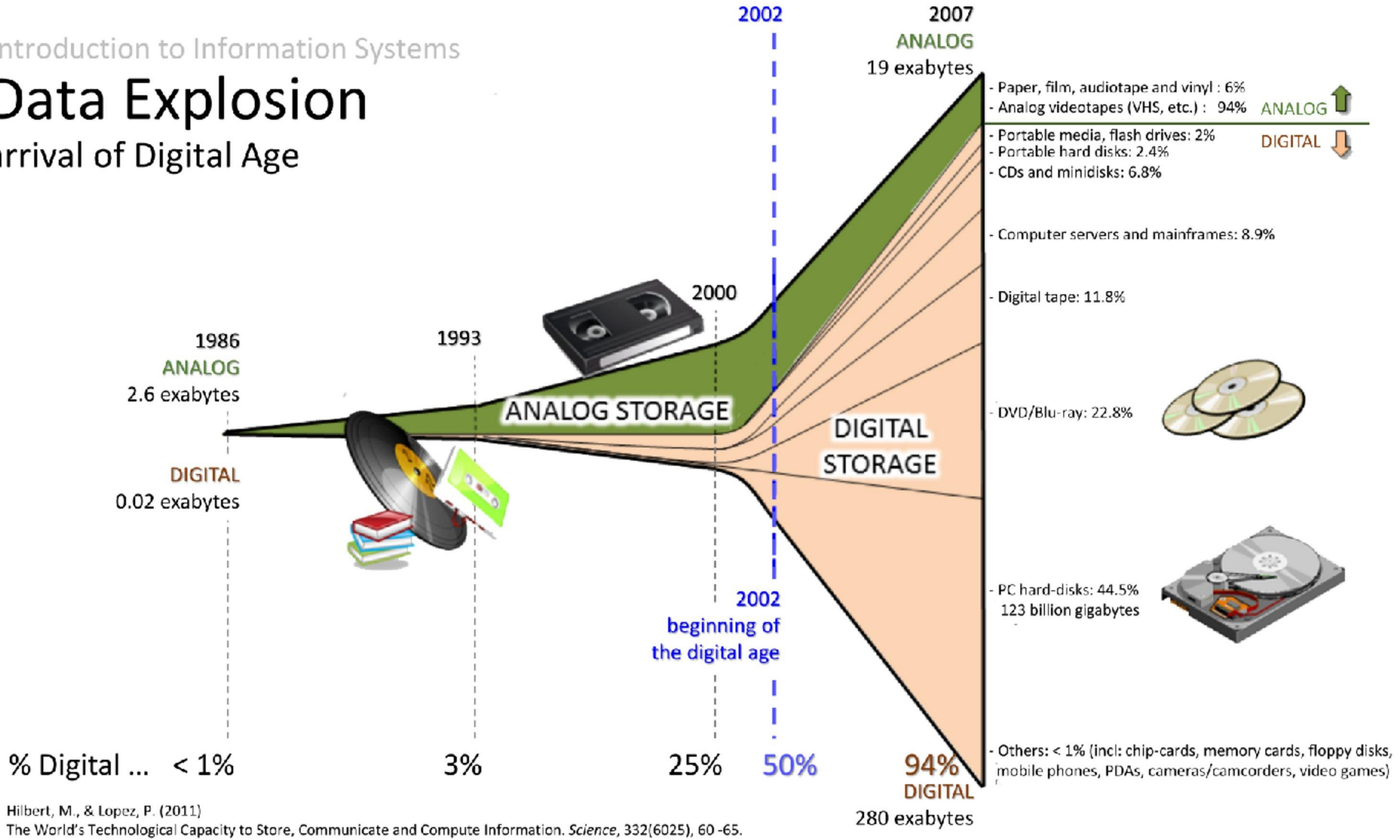
1 Kilobyte (KB)	= 1000 bytes	= 2^{10} bytes	… roughly 2 to 3 paragraphs of text
1 Megabyte (MB)	= 1000 KB	= 2^{20} bytes	… roughly 4 books of 200 pages each
1 Gigabyte (GB)	= 1000 MB	= 2^{30} bytes	… roughly 350 images or 650 web pages
1 Terabyte (TB)	= 1000 GB	= 2^{40} bytes	… minimum recommended size of a hard-disk in 2020
1 Petabyte (PB)	= 1000 TB	= 2^{50} bytes	
1 Exabyte (EB)	= 1000 PB	= 2^{60} bytes	
1 Zettabyte (ZB)	= 1000 EB	= 2^{70} bytes	
1 Yottabyte (YB)	= 1000 ZB	= 2^{80} bytes	

Fun Fact:
In 1982 I operated the most advanced computer available back then. It was called Commodore 64. It's entire capacity was 64KB. Compare this to your laptops and smartphones of today to appreciate how technology has evolved.

As of today, there is no approved standard for any unit of measure larger than a yottabyte.
However, talks are underway to adopt a Brontobyte to be equivalent to 1000 YB, or 2^{90} bytes.

Data Explosion

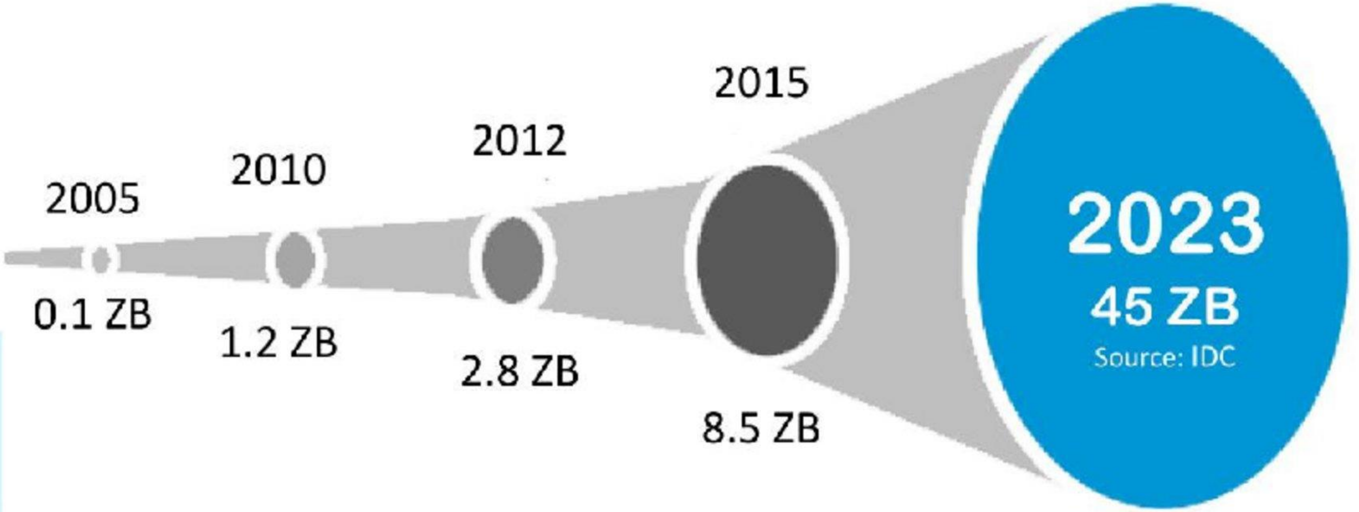
arrival of Digital Age

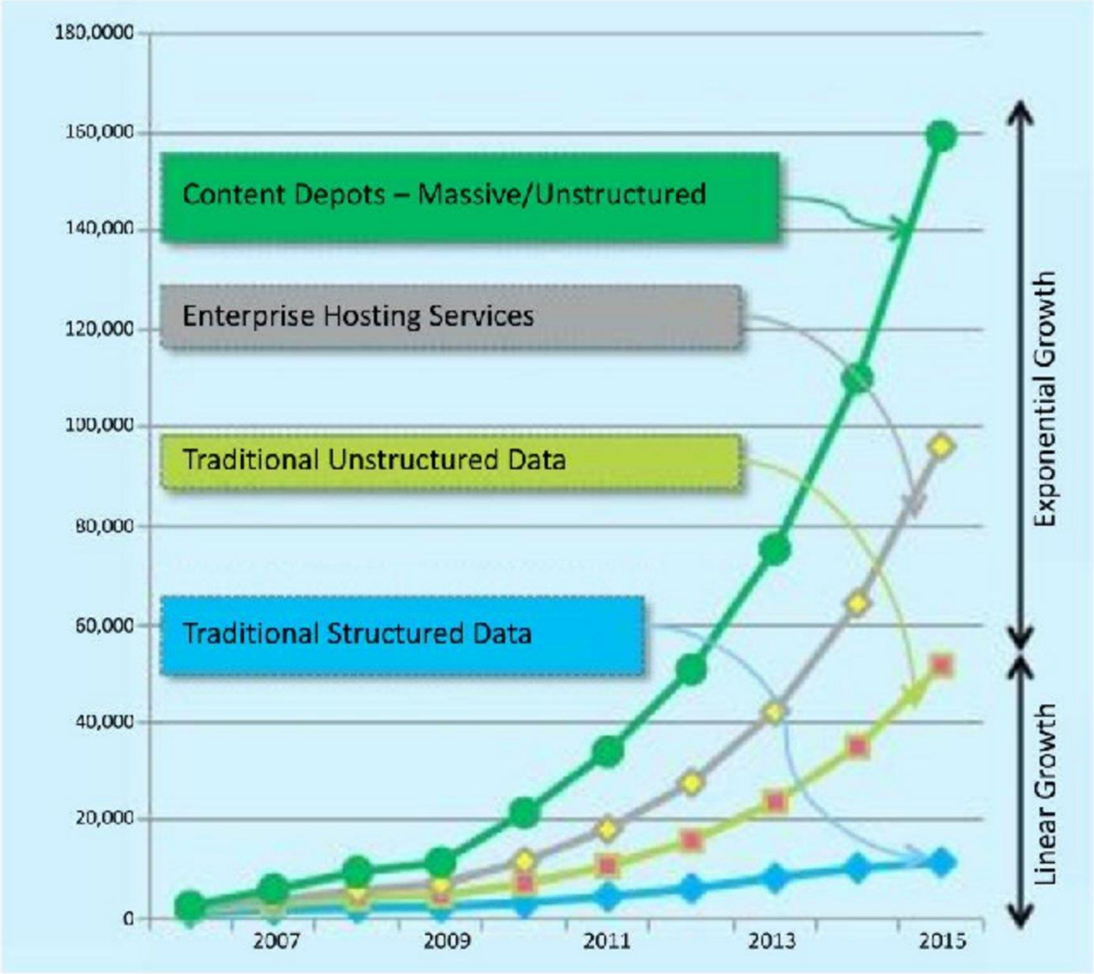


Source: Hilbert, M., & Lopez, P. (2011)
The World's Technological Capacity to Store, Communicate and Compute Information. *Science*, 332(6025), 60 -65.

Data Tsunami

birth of Big Data





Over **24 Petabytes** of data processed by Google everyday in 2011

4 Billion
pieces of content shared on Facebook everyday by June 2011

250 Million
Tweets per day in October 2011

5.5 Million
legitimate emails sent every second in 2011

Between the birth of Earth and 2003, there were 5 EB of data created. We now create 5 EB every single day.

Big Data

Big Data is defined as an accumulation of very large data sets that is too large and complex for processing by traditional database management tools but may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

Since the turn of the century, the amount of data in our world has been fast exploding.

The increasing volume and detail of information captured by organizations and large enterprises, coupled with the rise of multimedia, social media, satellite imagery and the Internet of Things (*aka* IoT, which we will discuss on Day 7) have fueled this exponential growth in real time data volume and complexity of large data sets; called Big Data.

Leaders in every industry are grappling with the implication of Big Data. Analyzing Big Data has become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer demand.

Key Attributes of Big Data:

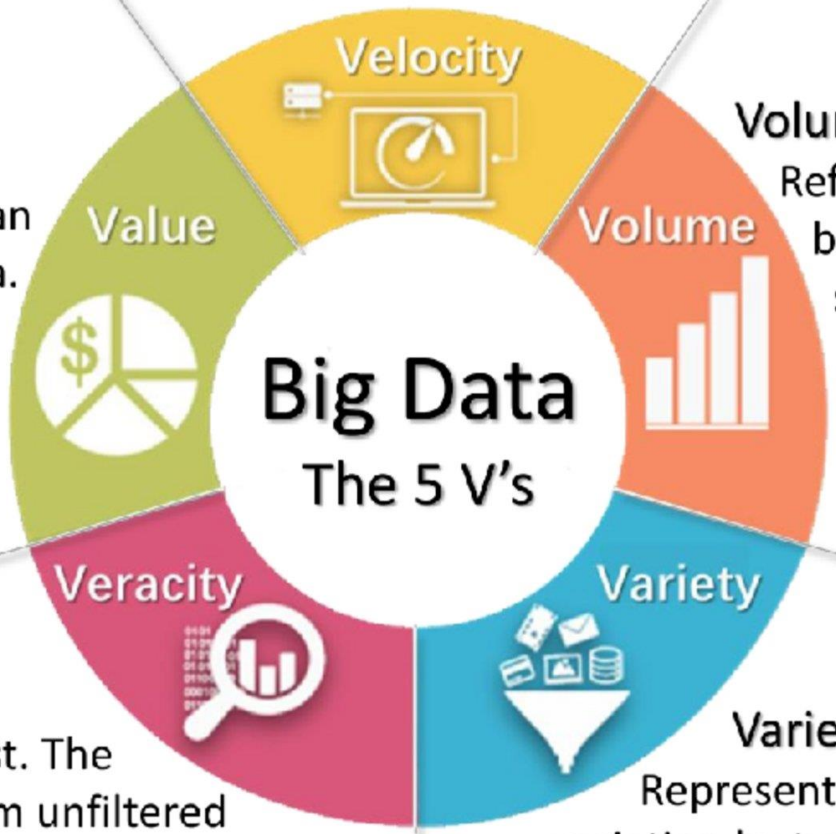
The 5 V's …
1. Velocity - speed of data
2. Volume - size of data
3. Variety - diversity of data
4. Veracity - reliability of data
5. Value - utility of data

Big Data

Key Attributes

Velocity
Speed at which data is generated and processed to meet demands as well as changes occurring between diverse data sets.

Value
Refers to the utility that can be extracted from big data. Having access to big data is useful, but that's only good if we can turn it into value.



Volume
Reflects the sheer volume of data being generated and stored every second. The size of data determines the value and potential insight, and whether it can be considered as big data or not.

Veracity
Refers to data reliability and trust. The exponential increase of data from unfiltered and constantly flowing data sources often leads to data quality issues. One must sift through all the noise in data.

Variety
Represents the diversity of data by nature and variation by type (social, media, text, etc.) and is a combination of structured and unstructured data.

Data Lake

A **data lake** is a system or repository of data stored in its natural and raw format.

A data lake is usually a single store of all enterprise data including raw copies of source system data and transformed data used for tasks such as reporting, visualization, advanced data analytics and machine learning.

A data lake can include:

- structured data from RDBMS (rows and columns),
- semi-structured data (CSV files, logs, XML, JSON),
- unstructured data (emails, documents, PDFs) and
- binary data (images, audio, video).

A data lake can be established "on prem" (within an organization's data centers) or "in the cloud" (using cloud services from vendors such as Amazon, Google and Microsoft).

A **data swamp** is a deteriorated and unmanaged data lake that is either inaccessible to its intended users or is providing little to no value.

A **data graveyard** is created when companies dump every imaginable piece of data into a data lake, in the hopes of doing something with it down the road; but then they lose track of what’s there.

Data Warehouse

A **data warehouse** is a repository of an organization's electronically stored data.

The primary intent and definition of Data Warehouse focuses on data storage.
However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Therefore, many people refer to data warehousing in this broader context.

A data warehouse is a system used to facilitate reporting and data analysis, and is considered a core component of business intelligence.

Organizations build data warehouses because the information in the RDBMS is not organized in a way that makes it readily accessible, requiring queries that are too complicated or resource-consuming.

Data stored in the data warehouse is uploaded from transactional processing systems (such as ERP). The data may pass through an operational data store and may require data cleansing using data analysis tools (such as Power Query) to ensure data quality before it is released and used in the data warehouse for reporting.

While transactional processing systems and ERPs are designed to be constantly updated, data warehouses are read only and are primarily designed to access large groups of related records for reporting purposes.

Data Warehouse

Benefits

- Integrate data from multiple data sources into a single database and data model, enabling a central view across the enterprise. This benefit is always valuable, but particularly so when the organization has grown by merger.
- Mitigate the problem of database lock contention in transaction processing systems caused by attempts to run large, long-running analysis queries in transaction processing databases.
- Maintain data history, even if the source transaction systems do not.
- Improve data quality, by providing consistent codes and descriptions, flagging or even fixing bad data.
- Present the organization's information consistently.
- Provide a single common data model (*aka* single source of truth) for all data of interest regardless of the data's source.
- Restructure the data so that it makes sense to the business users.
- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.
- Add value to operational business applications, notably customer relationship management (CRM) systems.
- Make decision-support queries easier to write.
- Organize and disambiguate repetitive data.

Data Mart

A **data mart** is a process-oriented database that meets the demands of a specific group of business users. Data marts accelerate business processes by allowing access to information in a data warehouse or operational data store within days as opposed to months or longer.

The data mart is a subset of the data warehouse and is usually oriented to a specific business line or team. It is the access layer of a data warehouse environment that is used to get data out to the users.

Whereas data warehouses have an enterprise-wide depth, the information in data marts pertain to a single department.

A data mart is basically a condensed and more focused version of a data warehouse that reflects the regulations and process specifications of each business unit within an organization. Data marts improve end-user response time by allowing users to have access to the specific type of data they need to view most often, by providing the data in a way that supports the collective view of a group of users.

Each data mart is dedicated to a specific business function or region. This subset of data may span across many or all of an enterprise's functional subject areas. It is common for multiple data marts to be used in order to serve the needs of each individual business unit (different data marts can be used to obtain specific information for various enterprise departments, such as accounting, marketing, sales, etc.).

Data Ethics

90% of digital data in the world today has been created in past 2 years, which is driving the machine learning explosion. Given this pace, many companies have not paused to consider if they are using this newly created data ethically.

Since the dawn of Internet, the sheer quantity and quality of data has dramatically increased, and is continuing to do so exponentially, giving birth to Big Data. Recent innovations in research and healthcare, such as high-throughput genome sequencing, electronic medical patient records and a plethora of internet-connected health devices have triggered a deluge of personal data that is becoming increasingly difficult to handle. As the quantity of data increases, Data Ethics becomes increasingly relevant because of the scale of its impact.

Data Ethics refers to systemizing, defending, and recommending concepts of right and wrong conduct in relation to data in general, and personal data in particular.

Business are interested in adopting data ethics to:

- Comply with regulations.
- Prove themselves trustworthy.
- Ensure fair and reasonable data usage.
- Minimize biases and social inequities.
- Develop a positive public perception.

Data Ethics

10 Key Principles

- Ownership: individuals own their own data.
- Consent: individuals should have control over how their data is to be used. They need to provide informed and explicitly expressed consent for what personal data moves to whom, when, and for what purpose.
- Transparency: if an individual's personal data is used, they should have transparent access to the algorithm used to generate aggregate data sets, with clarity of intent and any underlying ethical and societal implications.
- Availability: aggregate data sets should be freely available to individuals.
- Privacy: if data transactions occur, all reasonable efforts should be made to preserve the privacy of an individual's data.
- Currency: individuals should be aware of any financial transactions resulting from the use of their personal data and the scale of these transactions.
- Trustworthiness: sourcing of data and any related process should be truthful, reliable and consistent.
- Accountability: data custodians and data generators should be held accountable for the outcome of actions performed on data and the decisions derived from it.
- Benevolence: data should not be used to create harm: physically, reputationally, monetarily or emotionally.
- Value: data should not only contribute to knowledge and innovation, but also have significant social impact.

Data Ethics

Regulation

Transparency	Accountability	Right to Privacy	Freedom of expression	Equality & Non-discrimination	Data Protection & User Control	Economic Rights
Regulators require entities to manage data in the enterprise lifecycle in a transparent and explainable way.	Increasing mention in regulations on questions of liability, where data decisions lead to lawsuits and are impacted by common law and torts.	Implications for the enterprise when you have data on individuals or subjects who are under the purview of different privacy laws.	Regulation related to filtering the content related to expression, managing hateful content and automatic content regulation by algorithms.	Anti discrimination laws in context of the outcomes of algorithms. Bias testing and performance needs in light of evolving legislations.	Legal rights of the data owners in terms of consent and approval for usage of their data in processing it towards an outcome.	Ensuring that the use of data does not exploit the economic rights of a group, e.g. targeted interest rates for a certain class of people.