# Information and Big Data

## MODULE 3 / UNIT 10 / 0.5

MOISES M. MARTINEZ
FUNDAMENTALS OF COMPUTER ENGINEERING

2025/2026

# Data foundations

# 01

# Data foundations

## Data vs Information

Data consists of unorganized and unrefined raw facts.

Information is the organized and interpreted form of those raw facts (context).

# Data foundations

The data is categorized based on the way it is stored and organized, utilizing a specific structure and/or labelling mechanism.

# Data foundations

The data is categorized based on the way it is stored and organized, utilizing a specific structure and/or labelling mechanism.

| | | |
|---|---|---|
| Structured | Information is organized using a well-defined data model or schema, where each element has a precise meaning.<br>This ensures clear and consistent data interpretation, storage, and retrieval. | Databases |

# Data foundations

The data is categorized based on the way it is stored and organized, utilizing a specific structure and/or labelling mechanism.

| | | |
|---|---|---|
| Structured | Information is organized using a well-defined data model or schema, where each element has a precise meaning.<br>This ensures clear and consistent data interpretation, storage, and retrieval. | Databases |
| Semi-structured | Information is organized using tags or markers that define semantic attributes, allowing for more flexible and hierarchical data organization.<br>Information does not conform to the tabular structure typical of relational databases or other forms of structured data tables. | CSV, XML, HTML, JSON, … |

# Data foundations

The data is categorized based on the way it is stored and organized, utilizing a specific structure and/or labelling mechanism.

| | | |
|---|---|---|
| Structured | Information is organized using a well-defined data model or schema, where each element has a precise meaning.<br>This ensures clear and consistent data interpretation, storage, and retrieval. | Databases |
| Semi-structured | Information is organized using tags or markers that define semantic attributes, allowing for more flexible and hierarchical data organization.<br>Information does not conform to the tabular structure typical of relational databases or other forms of structured data tables. | CSV, XML, HTML, JSON, … |
| Unstructured | Information is not organized according to a predefined data model or schema.<br>This type of data does not fit neatly into traditional databases or structured formats, making it more challenging to analyse and process. | Plain text, multimedia files, etc |

# Data foundations

Different storage systems are utilized depending on the specific data format and storage requirements.

| | |
|---|---|
| Files | Plain text files store raw data.<br>The information within these files is often messy and unstructured, lacking a predefined format or organization, which can make it challenging to process and analyse. |
| Distributed files | Mass storage in structured data files is typically managed within a distributed environment.<br>The information within these files follows a specific structure or schema, allowing for efficient organization, retrieval, and analysis of the data across multiple systems or locations. |
| Databases | SQL (Structured Query Language) databases store data in a highly structured format using tables with predefined schemas.<br>NoSQL databases offer more flexibility by allowing storage of unstructured, semi-structured, or structured data without requiring a fixed schema. |

The choice of storage system is influenced by factors such as the type of data being stored (structured, semi-structured, or unstructured), the need for scalability, access speed, security, and the desired level of data organization and retrieval efficiency.

# Databases

# 02

# Databases

A database is a structured collection of information stored on a computer or computer system, designed for easy accessibility, retrieval, and modification.

- Databases efficiently manage and store vast amounts of data, making it possible to handle extensive datasets.
- Databases support quick and adaptable access to information, allowing users to retrieve specific data through queries tailored to their needs.
- Databases provide structured frameworks that enable the interconnection of various types of information, allowing for complex relationships between data points.
- Databases simplify the processes of printing and distributing information in multiple formats, ensuring that data can be shared effectively across different platforms and applications.

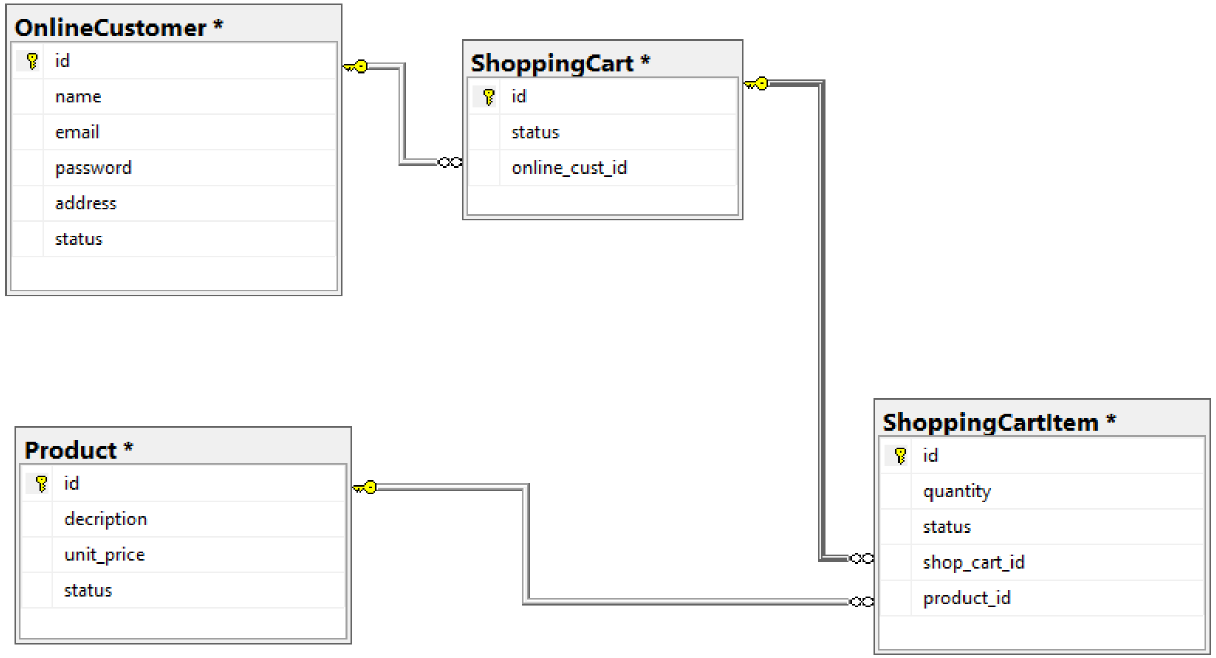
# Databases

There are different types of databases, categorized based on how information is organized and how it is stored at the physical level:

- Based on How Information is Organized:
  - SQL Databases
  - NoSQL (Not only SQL) Databases
- Based on How Information is Stored at the Physical Level:
  - Centralized Databases
  - Distributed Databases
  - Cloud Databases

## Relational Databases

A relational database (RDB) is a system for organizing information into tables, where data is structured in rows and columns. An RDB has the capability to create connections, or relationships, between different pieces of information by joining tables.




A relational database is often colloquially referred to as an SQL database.

## Relational Databases

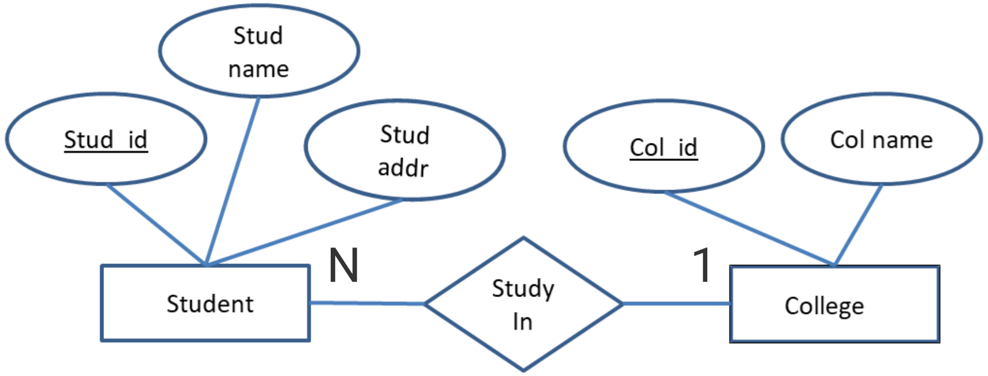
Relational databases are often represented using an Entity-Relationship (ER) diagram, which consists of four fundamental components:

- Entity: An entity represents a distinct object, person, place, unit, or any item for which information is stored. For example, in a user database, **User** would be an entity.
- Attribute: Attributes are specific pieces of information or characteristics associated with an entity, serving as properties of that entity. For example, the **first name** of a user is an attribute of the **User** entity.
- Primary Key: A primary key is a unique attribute that uniquely identifies each record within an entity. It ensures that each entry in a table can be uniquely distinguished from others. For example, a **User ID** could serve as the primary key for the **User** entity.
- Relationship: A relationship describes the associations or links between entities. For instance, a relationship might show that a particular user can rate multiple movies, indicating a connection between the **User** and **Movie** entities.

## Relational Databases

Entity-Relationship (ER) diagrams follow specific rules when being designed:

- Entities: Entities are visually represented by rectangles. Each rectangle represents a distinct entity in the database.
- Attributes: Attributes are depicted by ovals connected to their respective entities. Attributes that make up a primary key are underscored to indicate their uniqueness in identifying records within the entity.
- Relationships: Relationships between entities are symbolized by diamonds. These diamonds are connected to the related entities, indicating how the entities are associated with each other.
  - A "1" at one end of a relationship line signifies that the relationship involves a single entity on that side (one-to-one or one-to-many relationship).
  - An "N" (or "M") at one end signifies that the relationship involves multiple entities on that side (many-to-one or many-to-many relationship).



## Relational Databases

There are different types of relationship between entities:

- A one-to-one (1:1) relationship: An entity in A relates only to an entity in B and vice-versa. This relationship is not very common, because often one of the entities is defined as an attribute of the other. For example, each car has a unique number plate and each number plate belongs to only one car.
- An One-to-many (1:N) relationship: An entity in A is related to zero or many entities in B. But an entity in B relates to only one entity in A. For example, a customer can place any number of orders. But each specific order is placed by only one customer.

There is a many-to-many (N:M) relationship which exists conceptually. This is often transformed into an intermediate entity which incorporates attributes that serve as the primary keys of the entities involved in the relationship.

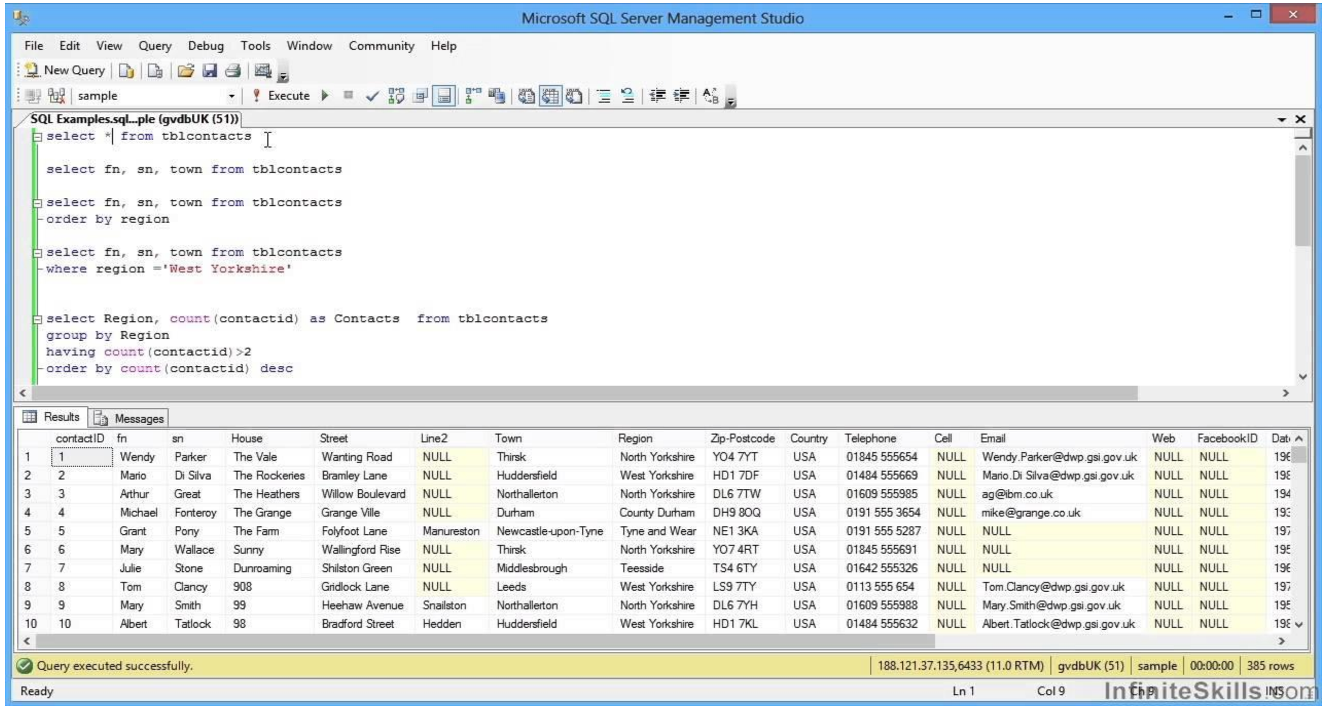
## Relational Databases

A Database Management System (DBMS) is application software designed to store, retrieve, query, and manage data efficiently. DBMSs offer a range of functionalities that facilitate the management of both the database and its data, including:

- DBMSs enable the creation, modification, and removal of definitions that structure the organization of data within the database.
- DBMSs support the insertion, modification, querying, and deletion of data.
- DBMSs provide tools for database administration, which includes registering and monitoring users, enforcing data security measures, ensuring data integrity, monitoring database performance, managing concurrency control, and recovering data that has been corrupted due to events like unexpected system failures.

## Relational Databases

SQL (Structured Query Language) is a domain-specific programming language specifically designed for managing and retrieving information from relational database management systems (RDBMS).




https://dev.mysql.com/doc/mysql-tutorial-excerpt/8.0/en/examples.html

## Relational Databases

SQL (Structured Query Language) is a domain-specific programming language specifically designed for managing and retrieving information from relational database management systems (RDBMS).

```
CREATE TABLE shop (
 article INT(4) UNSIGNED ZEROFILL DEFAULT '0000' NOT NULL,
 dealer  CHAR(20)                 DEFAULT ''     NOT NULL,
 price   DOUBLE(16,2)             DEFAULT '0.00' NOT NULL,
 PRIMARY KEY(article, dealer));



INSERT INTO shop VALUES
(1,'A',3.45),(1,'B',3.99),(2,'A',10.99),(3,'B',1.45),
(3,'C',1.69),(3,'D',1.25),(4,'D',19.95);

INSERT INTO shop(article, dealer, price) VALUES
(1,'A',3.45);
```

```
SELECT * FROM shop;

+---------+--------+-------+
| article | dealer | price |
+---------+--------+-------+
|    0001 | A      |  3.45 |
|    0001 | B      |  3.99 |
|    0002 | A      | 10.99 |
|    0003 | B      |  1.45 |
|    0003 | C      |  1.69 |
|    0003 | D      |  1.25 |
|    0004 | D      | 19.95 |
+---------+--------+-------+
```

## Relational Databases

SQL (Structured Query Language) is a domain-specific programming language specifically designed for managing and retrieving information from relational database management systems (RDBMS).

```
CREATE TABLE shop (
 article INT(4) UNSIGNED ZEROFILL DEFAULT '0000' NOT NULL,
 dealer  CHAR(20)                 DEFAULT ''     NOT NULL,
 price   DOUBLE(16,2)             DEFAULT '0.00' NOT NULL,
 PRIMARY KEY(article, dealer));



INSERT INTO shop VALUES
(1,'A',3.45),(1,'B',3.99),(2,'A',10.99),(3,'B',1.45),
(3,'C',1.69),(3,'D',1.25),(4,'D',19.95),(5,'R',1.45),;


INSERT INTO shop(article, dealer, price) VALUEA(1,'A',3.45);
```

```
SELECT * FROM shop;

+---------+--------+-------+
| article | dealer | price |
+---------+--------+-------+
|    0001 | A      |  3.45 |
|    0001 | B      |  3.99 |
|    0002 | A      | 10.99 |
|    0003 | B      |  1.45 |
|    0003 | C      |  1.69 |
|    0003 | D      |  1.25 |
|    0004 | D      | 19.95 |

+---------+--------+-------+
```

There are different ways to insert data in a DDBB, but these are the most common. Although, syntax can change according to the DBMS.

# How to create a DDBB for an online store?

## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?

## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?
  - Clients, Products and orders.

## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?
  - Clients, Products and orders.
- Which information do we need to store client data?

## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?
  - Clients, Products and orders.
- Which information do we need to store client data?
  - Client: Name, Surname, DNI, billing address, shipping address and more …
  - Product: Reference, name, description, price, available units and more …
  - Order: Order_id, Date, client_id, product_id, state and more …

## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?
  - Clients, Products and orders.
- Which information do we need to store client data?
  - Client: Name, Surname, DNI, billing address, shipping address and more …
  - Product: Reference, name, description, price, available units and more …
  - Order: Order_id, Date, client_id, product_id, state and more …
- Which relationships do we need?

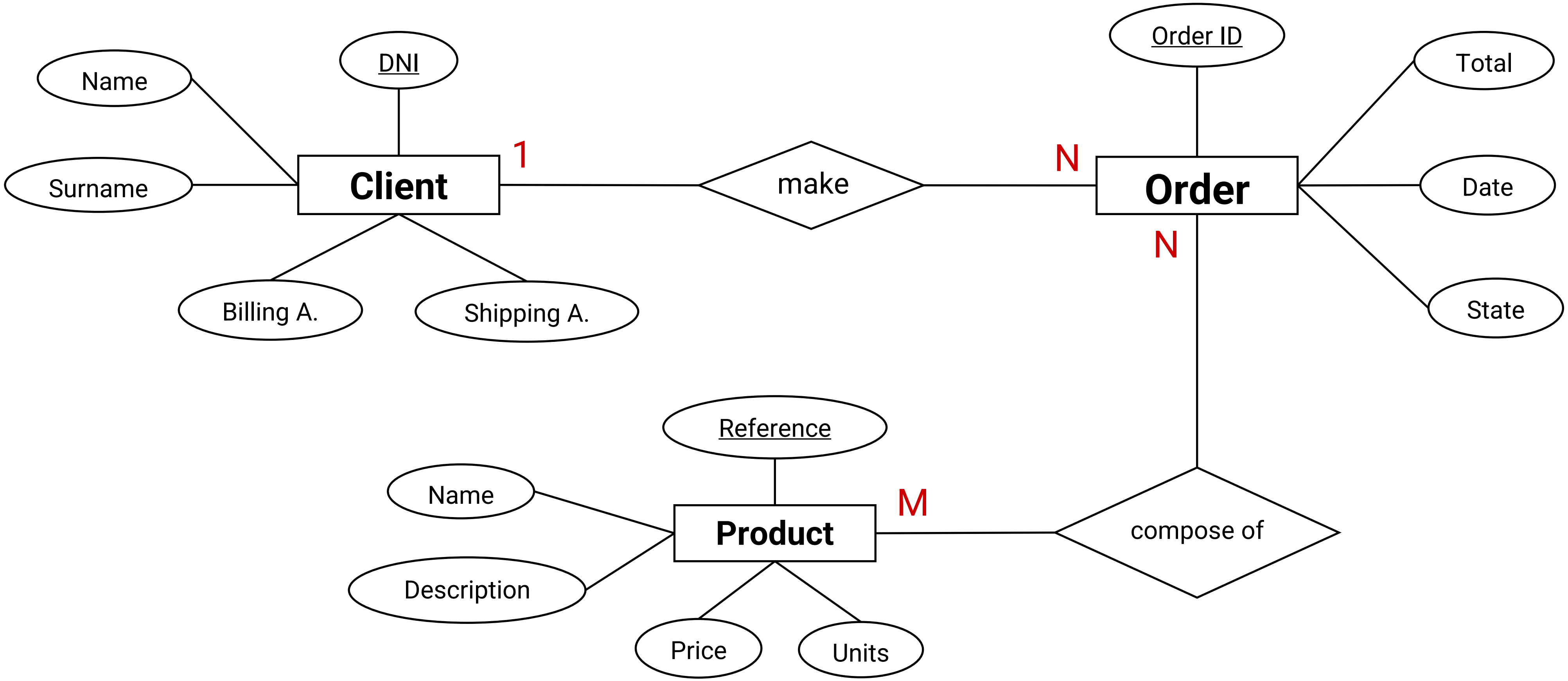
## Relational Databases

How to create a DDBB for an online store?

- Which entities do we need to store the basic data?
    - Clients, Products and orders.
- Which information do we need to store client data?
    - Client: Name, Surname, DNI, billing address, shipping address and more …
    - Product: Reference, name, description, price, available units and more …
    - Order: Order_id, Date, client_id, product_id, state and more …
- Which relationships do we need?
    - 1:N $\rightarrow$ client:order
    - N:M $\rightarrow$ product:order

## Relational Databases

How to create a DDBB for an online store?



## Relational Databases

How to create a DDBB for an online store?

1. Definition of the information

   The different entities are converted into tables, where each attribute is a field of the table.

   - Fields within a database exhibit specific data types such as varchar, int, double, char, among others.
   - Each distinct constituent within a table, often referred to as a row, is denoted as a record.
   - Primary keys are fashioned by employing fields that necessitate uniqueness to uniquely identify each individual element within the dataset

## Relational Databases

How to create a DDBB for an online store?

2. Definition of the relationships' cardinalities

   The relationship (make) 1-to-many (1:N) is included to describe when a client have many orders but each specific order is only related to one client.

   - This association involves the integration of the primary key from the singular entity side into the multiple-entity side.
   - The DNI field, serving as the primary key of the client, is incorporated as a distinct field within the order table. This new field is called Foreign Key.

A **FOREIGN KEY** is a field (or collection of fields) within one table, establishing a reference to the **PRIMARY KEY** present in another table.

## Relational Databases

How to create a DDBB for an online store?

2. Definition of the relationships' cardinalities

   The relationship (compose-of) many-to-many (N:M) is included to describe when an order is composed of some product units.

   - This relationship generates a new table into the DB. It can include other attributes like units, price, etc.
   - This table will incorporate the **PRIMARY KEYS** from the tables involved in the relationship.

| Order id | Reference |
|---|---|
| 12324 | 364834034843784 |

# Big Data

# 02

OFF-EXAM CONTENT


https://www.youtube.com/watch?v=bAyrObl7TYE

## “Big Data” definition

The term "Big Data" describes the large volume of data, both structured and unstructured, that is created daily by today's society.

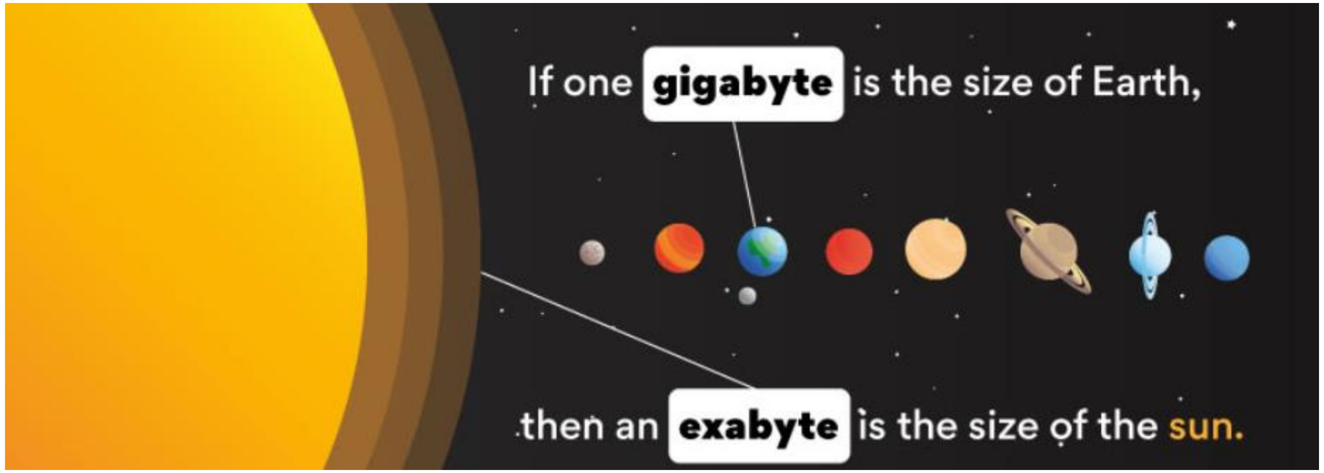
## “Big Data” definition

The term "Big Data" describes the large volume of data, both structured and unstructured, that is created daily by today's society.

## My “Big Data” definition

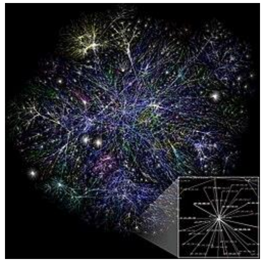
It is the process of collecting, storing and subsequent analysis and manipulation of data at a massive level in order to extract value.

From cuneiform writing, the earliest known writing system to date, to contemporary data centres, human civilization has consistently amassed information. Moreover, prognostications indicate that by the year 2030, our society will produce several yottabytes of information annually.




The **yottabyte** is currently the largest recognized unit of data storage for devices and cloud services.

# Big Data



1.3 ZB of network traffic by 2016.

The CERN Large Hadron Collider generates 1Petabyte per second.

Brontobyte (BB)

Zettabyte (ZB)

Petabyte (PB)

Gigabyte (GB)

$10^6$ $10^9$ $10^{12}$ $10^{15}$ $10^{18}$ $10^{21}$ $10^{24}$ $10^{27}$

Megabyte (MB)

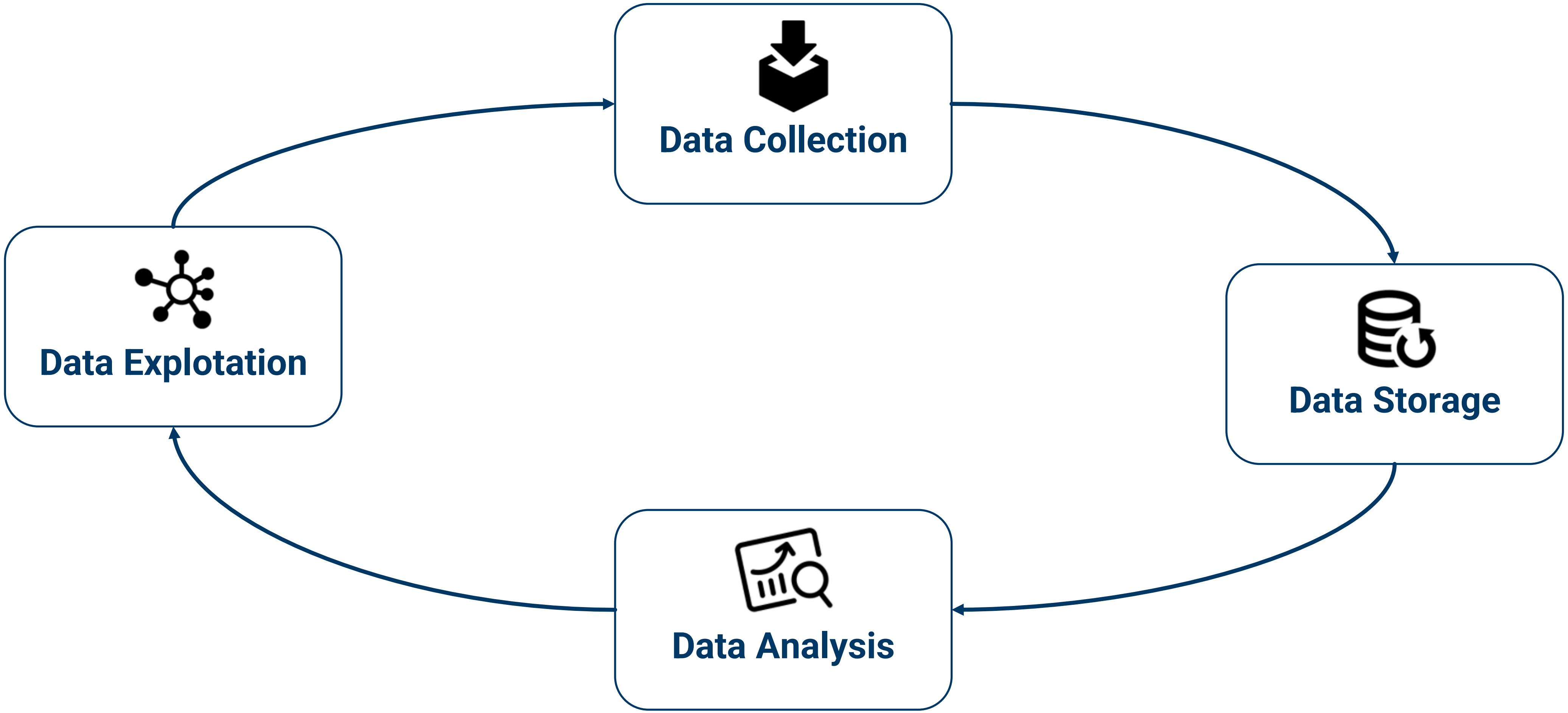
Terabyte (TB)

Exabyte (EB)

Yottabyte (YB)

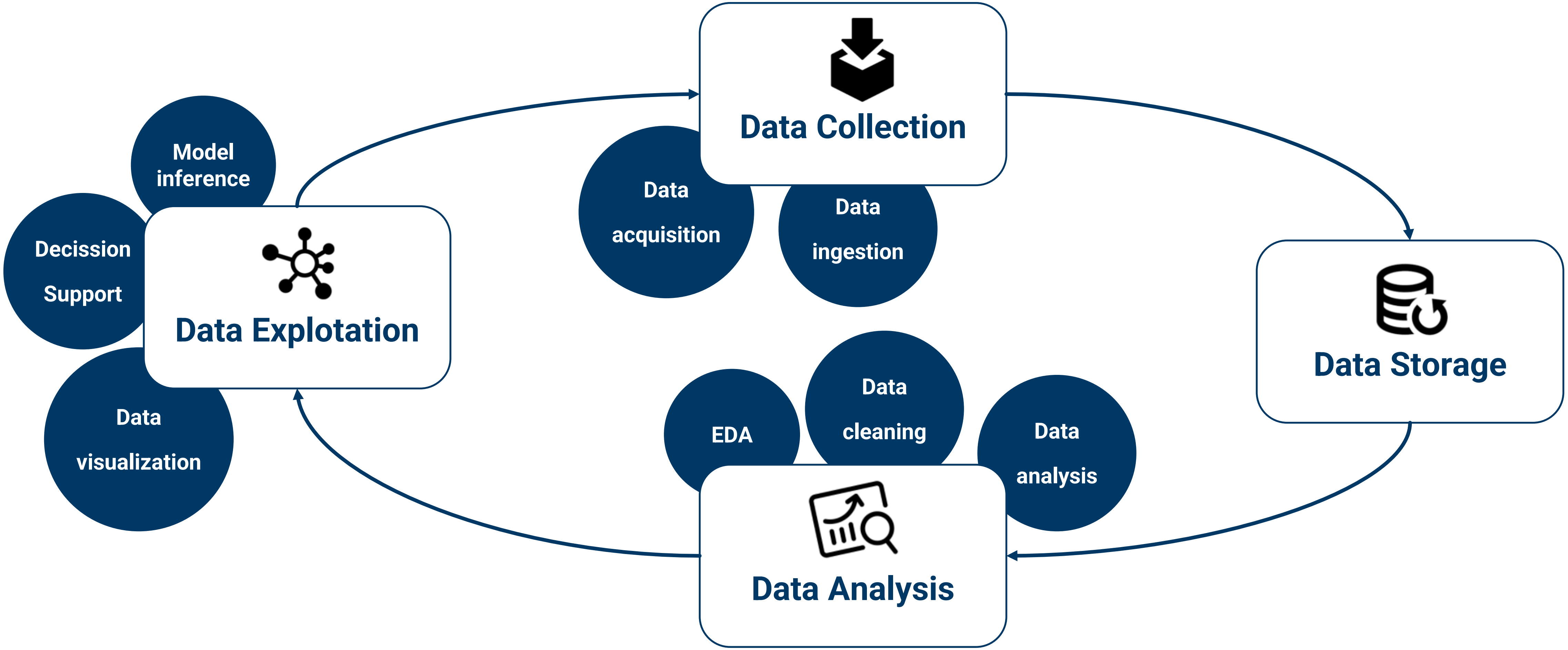
Our digital universe today = 250 trillion of DVDs.

## The big data lifecycle



UFV

## The big data lifecycle



## The big data lifecycle – Data collection

The Data Collection phase involves gathering data from a variety of sources, such as social media, sensors, transactional systems, logs, and more. This phase includes two key subtasks:

- Data Acquisition is the process of identifying and Integrating the different data sources such as APIs, IoT devices, web scraping, or internal systems, and establish connections to these sources to begin collecting data.

**Web scrapers or web crawlers**

**APIs**

## The big data lifecycle – Data collection

The Data Collection phase involves gathering data from a variety of sources, such as social media, sensors, transactional systems, logs, and more. This phase includes two key subtasks:

- Data Ingestion is the process of transferring and loading data into a storage system or processing pipeline, making it available for further analysis and use
  - Real-Time (Streaming) processing which implements mechanisms to collect data as it is generated, allowing for real-time processing and analysis. This is crucial for time-sensitive applications like monitoring, alerts, and real-time analytics.
  - Batch processing which schedules regular intervals to collect and load large volumes of data, suitable for processing in bulk. This approach is ideal for handling historical data or performing end-of-day analysis.

**Message queues**

**Streaming tools**

## The big data lifecycle – Data storage

The Data collection phase involves storing the collected and ingested data in a way that supports easy access, retrieval, and future processing. There are different systems to manage and store large volumes of data

- Data Lakes which are used for storing vast amounts of raw, unstructured, and semi-structured data. Data lakes allow for flexible data storage without a predefined schema, making them ideal for storing data in its original format until it's needed for processing.
- Data Warehouses which is a structured repositories optimized for query performance and analysis. Data warehouses store structured data in organized tables, with a focus on relational data that is ready for immediate analysis.
- NoSQL Databases which are used for storing semi-structured or unstructured data, such as JSON documents, key-value pairs, or graph data. NoSQL databases are highly scalable and can handle large volumes of diverse data types.

**This phase is essential for managing large volumes of data effectively, ensuring that the data is available, secure, and organized for subsequent analysis.**

## The big data lifecycle – Data Analysis

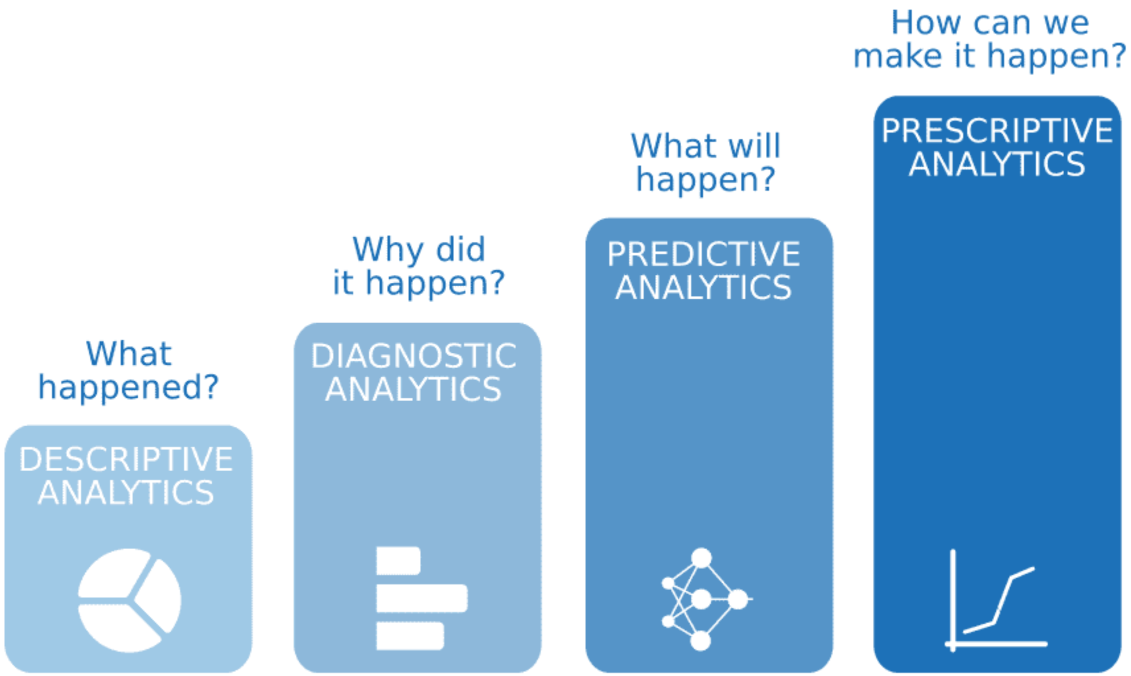
The Data Analysis phase involves examining the processed and stored data to uncover insights, patterns, and trends that can inform decision-making.

- Data preparation involves initially **exploring the data** to understand its structure and key characteristics, cleaning it by handling missing values and correcting errors, and enhancing it through feature engineering to improve the accuracy and reliability of analytical models.
  - Exploratory Data Analysis (EDA) is the process to explore the data and uncover underlying patterns, detect anomalies, test hypotheses, and check assumptions through a combination of statistical and graphical techniques.
  - Data cleaning is the process to identify and correct or remove errors, inconsistencies, and inaccuracies within a dataset to ensure its quality and reliability.

## The big data lifecycle – Data Analysis

The Data Analysis phase involves examining the processed and stored data to uncover insights, patterns, and trends that can inform decision-making.

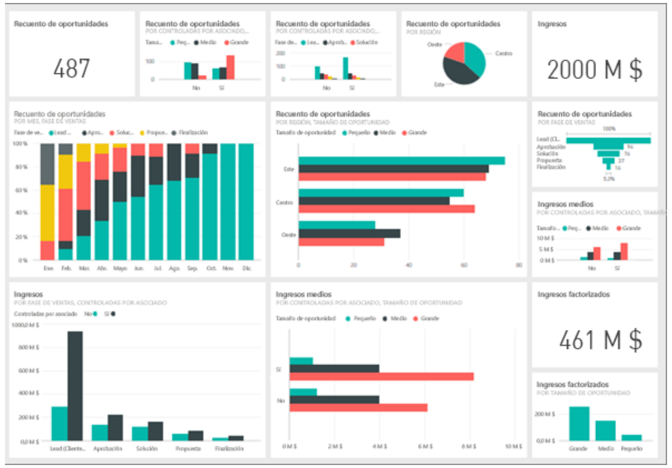
- Data analysis involves applying different techniques, mostly of them related to Artificial Intelligence, to extract meaningful insights from the data
  - Descriptive Analytics
  - Diagnostic Analytics
  - Predictive Analytics
  - Prescriptive Analytics

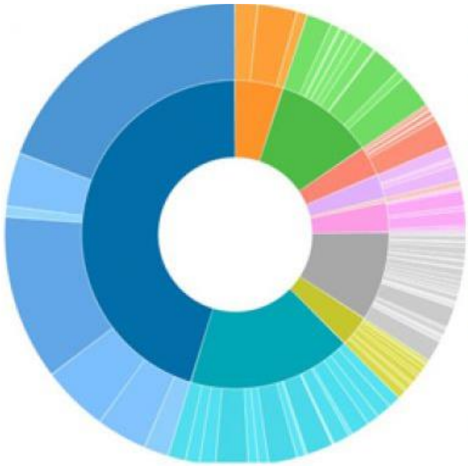


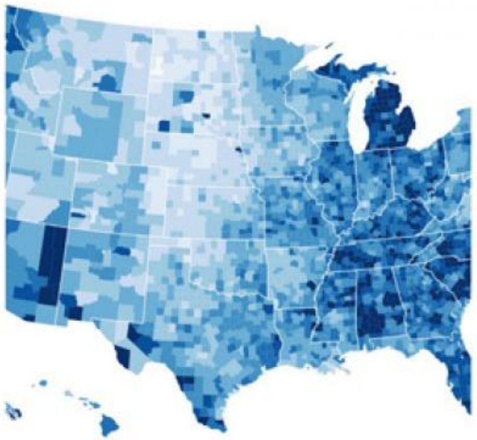
## The big data lifecycle – Data exploitation

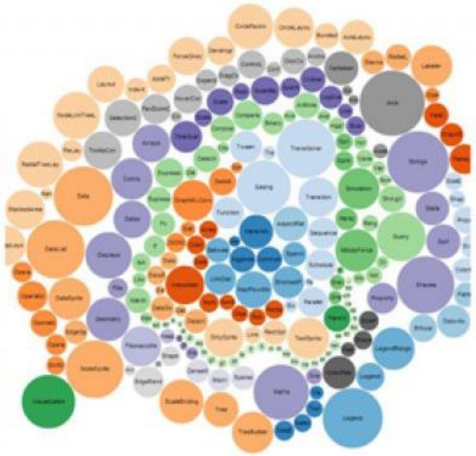
The Data Exploitation phase involves generating actionable insights and drive decision-making using the processed and analysed data.

- Data Visualization consists of translating the analytical results into visual formats like graphs, charts, heatmaps, and dashboards.







## The big data lifecycle – Data exploitation

The Data Exploitation phase involves generating actionable insights and drive decision-making using the processed and analysed data.

- Decision Support consists of using the insights derived from the analysis to support business decisions. These insights are extracted from interactive dashboards that allow users to explore the data and analysis results dynamically.
- Model inference consists of utilizing predictive models based on Machine Learning (ML) and Deep Learning (DL) to make informed predictions or classifications based on new data.

# Laws and ethics

03

# Why do you believe laws are necessary in computer engineering?

OFF-EXAM CONTENT

## Evolution of the data laws in Spain

**1978:** Constitución Española. Art. 18.4 donde se garantiza el derecho de las personas al honor y la intimidad personal y familiar.

**1992:** LORTAD. Es la Ley Orgánica de Protección del Tratamiento Automatizado de los Datos de Carácter Personal (no vigente en la actualidad).

**1994:** Reglamento que desarrolla determinados aspectos de la LORTAD (este Reglamento sigue vigente a pesar de la derogación de la LORTAD).

**1995:** Directiva comunitaria relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. La LOPD española se deriva de esta Directiva.

**1999:** Reglamento de Medidas de Seguridad (RMS). Especifica las medidas de seguridad técnicas y organizativas que se deben adoptar para los ficheros que contengan datos de carácter personal (11 de junio de 1999).

**1999:** **LOPD - Ley Orgánica de Protección de Datos de Carácter Personal** (adaptación de la antigua LORTAD a la Directiva Comunitaria de 1995). La ley sólo se aplica a los datos personales de las personas físicas, no de las personas jurídicas (empresas) (13 de diciembre de 1999).

OFF-EXAM CONTENT

## Ley Orgánica de Protección de datos (LOPD)

The Organic Law 15/1999 of December 13 on Protection of Personal Data (Ley Orgánica de Protección de Datos de Carácter Personal, LOPD) was a Spanish organic law that aimed to guarantee and protect the processing of personal data, safeguarding public liberties and fundamental human rights, with a particular focus on personal and family honor and privacy.

- This law regulated the treatment of personal data and files, regardless of the medium in which they were processed.
- This law outlined the rights of citizens concerning their personal data, ensuring they had control over how their data was used. It also defined the obligations of those who create or process this data, emphasizing the importance of maintaining data security and privacy.

OFF-EXAM CONTENT

## Ley Orgánica de Protección de datos (LOPD)

The GDPR (General Data Protection Regulation) is a regulation in European Union (EU) law that governs data protection and privacy within the EU and the European Economic Area (EEA). This regulation has been mandatory in all EU member countries since May 25, 2018.

The LOPD-GDD (Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales) is the national law in Spain that aligns with the GDPR, integrating its principles and extending its regulations. This law unifies all European data protection regulations under a single legal framework while also providing additional protections specific to digital rights in Spain.

# How to implement the new LOPD-GDD in a company?

OFF-EXAM CONTENT

## Ley Orgánica de Protección de datos (LOPD)

Any type of company or business that deals with sensitive data of third parties, must comply with each and every one of the requirements established in the new regulations of the Law on Protection of Personal Data and Guarantees of Digital Rights. The LOPD-GDD will be applied when the following treatments are given:

- Data processing of individual entrepreneurs and liberal professionals.
- Commercial operations.
- Use of surveillance systems.
- Advertising exclusion systems.
- Communication channels and complaints.
- Credit information systems.

OFF-EXAM CONTENT

## Ley Orgánica de Protección de datos (LOPD)

The most important point of the new law:

- Protection of minors: The consent of a minor will only be valid when he is over fourteen years of age, being necessary the authorization of the father, mother or guardian if it is not.
- Control of personal data: To avoid the use of personal data for commercial use without prior consent, the LOPD-GDD establishes that the control of personal data falls directly on the user,always requiring their consent to use them.
- Employee privacy: It is forbidden to take recordings in the areas intended for the rest of the workers, toilets and other places intended for leisure.

OFF-EXAM CONTENT

## Ley Orgánica de Protección de datos (LOPD)

The most important point of the new law:

- Right to be forgotten: it establishes the right to delete data on social networks and other equivalent services.
- Data of deceased persons: In the event of death, any family member linked to the deceased person may request access, rectification or deletion of the shared data.
- Clear information about the use of data: Companies must inform users in a clear, simple and concise way about the possible use of the personal data they have been given. Companies could be fined up to EUR 20 million.

# Why is ethics important in computer engineering?

# Laws and ethics

UFV

OFF-EXAM CONTENT

Reuters  World  Business  Markets  Sustainability  Legal  Breakingviews  Technology  Investigations

World

## Insight - Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

October 11, 2018 2:50 AM GMT+2 · Updated 6 years ago

Aa


SAN FRANCISCO (Reuters) - Amazon.com Inc's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

## Watch as Tesla Model X SUV 'breaks multiple road laws during Full Self-Driving session'

A driver has posted footage appearing to highlight multiple errors within Tesla's Full Self-Driving V12 software - but not all viewers are convinced.

Stanford MEDICINE | News Center

Email  Tweet

## Researchers say use of artificial intelligence in medicine raises ethical questions

In a perspective piece, Stanford researchers discuss the ethical implications of using machine-learning tools in making health care decisions for patients.

## *Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms*

The company has suspended Gemini's ability to generate human images while it vowed to fix the issue.

Share full article  89


## After Uber, Tesla incidents, can artificial **intelligence** be trusted?

Apr 9, 2018 | News Stories

The reliability of self-driving cars and other forms of artificial intelligence is one of several factors that affect humans' trust in AI, **machine learning** and other technological advances, write two Missouri University of Science and Technology researchers in a recent journal article. "Trust is the cornerstone of ...

OFF-EXAM CONTENT

## To differentiate between what is right and what is wrong.

# Laws and ethics

OFF-EXAM CONTENT

Ethics, also called moral philosophy, is the discipline concerned with what is morally good and bad and morally right and wrong.

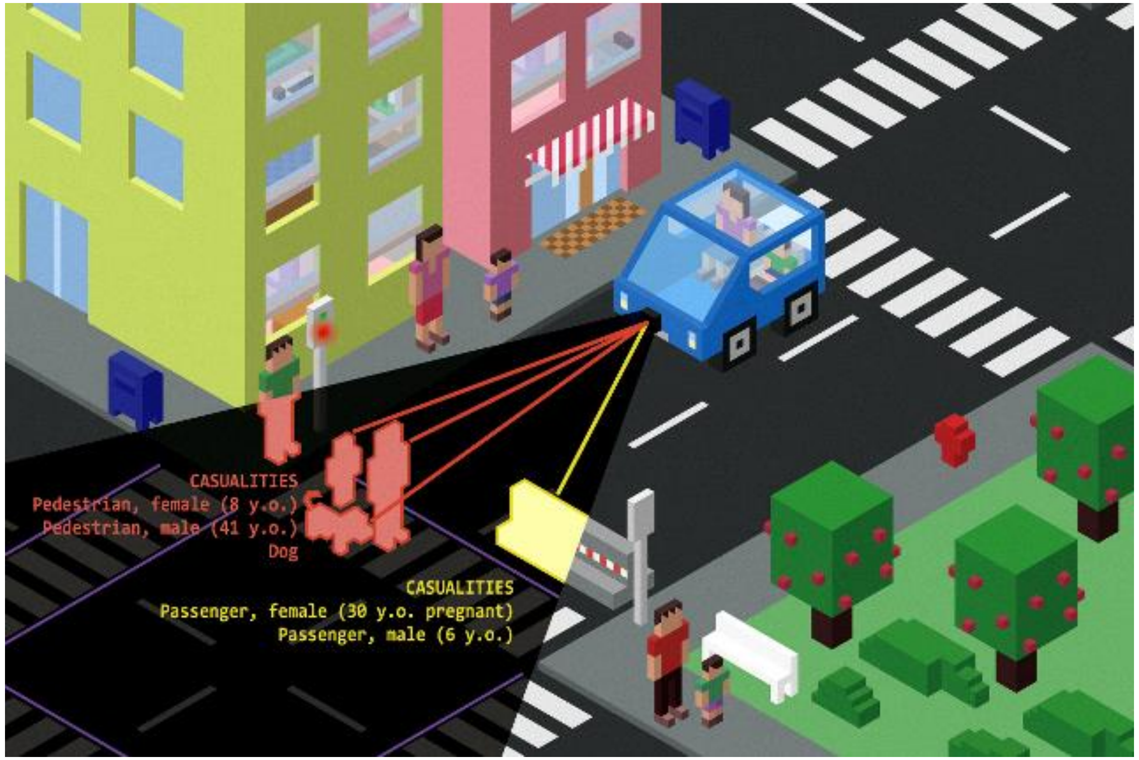
Why do we need ethics when we are creating software?

- Human have biases that we include in the information we create and analyze.
- During the design and development of software and data we contribute unconscious knowledge, we must ask ourselves if we have taken into account enough examples.
- Ideologies, expressions, validated technical information, natural/artificial light, atypical cases, etc.

# Laws and ethics

OFF-EXAM CONTENT

https://www.moralmachine.net/




**Option 1**

In this case, the self-driving car with a sudden brake failure will swerve and drive through a crosswalk in the other lane. This will result in the death of an elderly woman, two athletes and a child.

Please note that affected pedestrians are complying with the law when crossing with the green signal

**Option 2**

In this case, the self-driving car with sudden brake failure will continue forward and crash into a concrete barrier. This will result in the death of a child, a pregnant woman (passengers) and a dog.

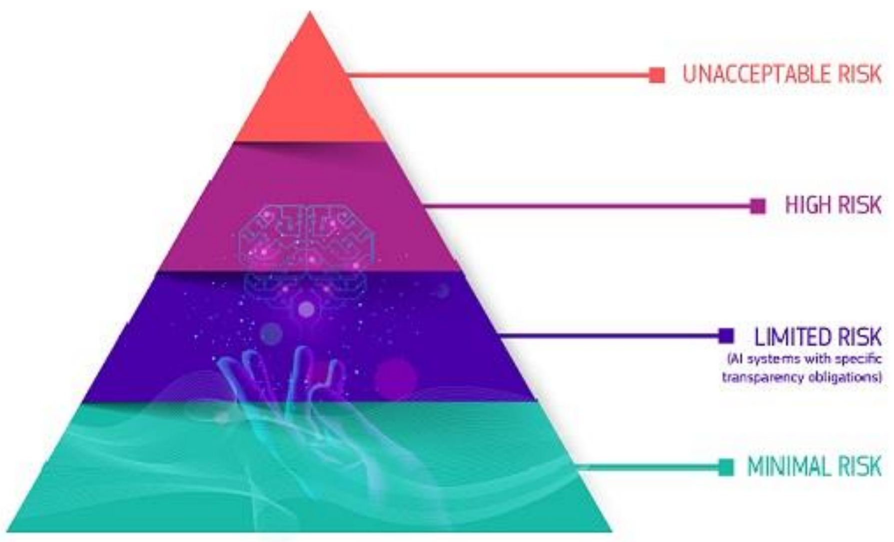
https://www.eeworldonline.com/this-mit-game-lets-you-choose-who-lives-and-dies-in-a-self-driving-car-wreck/

# How are we ensuring that software and data are created ethically?

OFF-EXAM CONTENT

## New regulatory framework on AI (European Union).

- Unacceptable risk: A very limited set of particularly harmful uses of AI that contravene EU values because they violate fundamental rights.
  - Social scoring for governments (This is happening in China).
- High Risk: A limited number of AI systems defined in the proposal, creating an adverse impact on people's safety or their fundamental rights (as protected by the EU Charter of Fundamental Rights) are considered to be high-risk.
  - Infrastructure.
  - Education.
  - Security.
  - Public services.
  - Immigration or border line controls.
- Limited risk: For certain AI systems specific transparency requirements are imposed, for example where there is a clear risk of manipulation (e.g. via the use of chatbots). Users should be aware that they are interacting with a machine.
- Minimal risk: All other AI systems can be developed and used subject to the existing legislation without additional legal obligations. The vast majority of AI systems currently used in the EU fall into this category.



# Laws and ethics

OFF-EXAM CONTENT

## Regulatory frameworks on other countries.

- Algorithmic Accountability Act 2019: Requires companies to provide an assessment of the risks posed by the automated decision system to the privacy or security and the risks that contribute to inaccurate, unfair, biased, or discriminatory decisions impacting consumers.
- California Consumer Privacy Act: Requires companies to rethink their approach to capturing, storing, and sharing personal data to align with the new requirements by January 1, 2020.
- Washington Bill 1655: Establishes guidelines for the use of automated decision systems to protect consumers, improve transparency, and create more market predictability.
- Massachusetts Bill H.2701: Establishes a commission on automated decision-making, transparency, fairness, and individual rights.
- Illinois House Bill 3415: States predictive data analytics determining creditworthiness or hiring decisions may not include information that correlates with the applicant race or zip code.