



# Data Management and Visualization

INTRODUCTION TO THE COURSE

Academic Year 2024-2025

Daniele Apiletti

# Teaching staff

## Daniele Apiletti



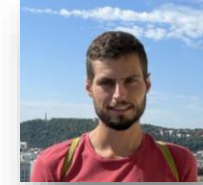
- Data Warehousing
- OLAP, extended SQL
- NoSQL data management
- Lectures and... any other business

## Diego Monti



- Data Visualization
- Lectures and practices

## Simone Monaco



- practice of NoSQL (MongoDB)
- practice on Data Warehousing

You can contact us

- during lectures and practices
- on **Piazza**

<https://piazza.com/polito.it/fall2024/01txasm>

# Piazza Q&A

We are using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from both classmates and teachers. Rather than emailing questions to the teaching staff, please post your questions on Piazza, even **anonymously** or **privately**.

We might use Piazza for **announcements** in case of **failure** of either the Polito teaching portal or the Virtual Classroom services (or both).

The screenshot displays the Piazza Q&A interface. The top navigation bar includes the Piazza logo, the course identifier '01TXASM', and menu items for 'Q & A', 'Resources', 'Statistics', and 'Manage Class'. The user 'Daniele Apiletti' is logged in. Below the navigation bar, there are tabs for 'LIVE Q&A', 'Drafts', and a list of folders including 'exam', 'datawarehousing', 'nosql', 'dataviz', and various 'lab' folders. A secondary navigation bar shows filters for 'Unread', 'Updated', 'Unresolved', and 'Following'. The main content area is titled 'private note @4' and shows a draft post with the following text:

**Introduce Piazza to your students**

**Post a Welcome Note!**

In your first post on Piazza, welcome your students to their new class:

Students,

Welcome to Piazza! We'll be conducting all class-related discussion here this term. The quicker you begin asking questions on Piazza (rather than via emails), the quicker you'll benefit from the collective knowledge of your classmates and instructors. We encourage you to ask questions when you're struggling to understand a concept—you can even do so anonymously.

-Daniele Apiletti

An 'Add Post' button is visible below the text area. On the left sidebar, a list of posts is shown, with the selected post 'Introduce Piazza to your stu...' highlighted. The bottom of the page contains a section titled 'Include this blurb in your syllabus' with a paragraph of text and a link to the class page: <https://piazza.com/polito.it/fall2022/01txasm/home>.

# Material

- Announcements on the official «teaching portal» private page <https://didattica.polito.it/> or through institutional email



POLITECNICO DI TORINO

myPoli u-GOV

News Guide Materiali Forum Students Elaborati Lectu  
Materials Homework

**01TXASM - Data management and visualization (8 cfu)**

Teaching staff: Daniele Apiletti

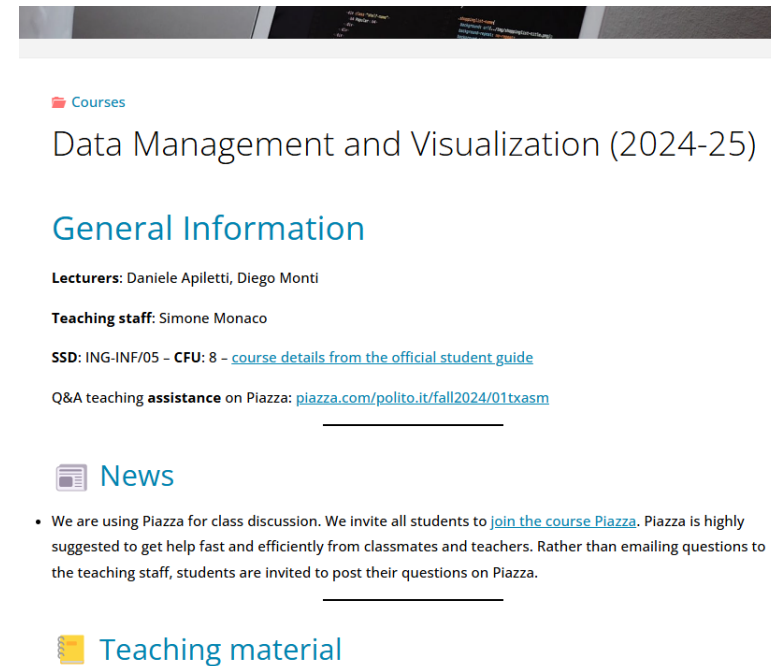
Semester: 1-1

Abilitazione gestione Pagina Corso a terzi ?

Accedi ad Exercise

Avvisi del corso

- Slides, texts of the practices, and all other materials are available on the **public page of the course** <https://dbdmg.polito.it/>



Courses

Data Management and Visualization (2024-25)

General Information

Lecturers: Daniele Apiletti, Diego Monti

Teaching staff: Simone Monaco

SSD: ING-INF/05 - CFU: 8 - [course details from the official student guide](#)

Q&A teaching assistance on Piazza: [piazza.com/polito.it/fall2024/01txasm](https://piazza.com/polito.it/fall2024/01txasm)

News

- We are using Piazza for class discussion. We invite all students to [join the course Piazza](#). Piazza is highly suggested to get help fast and efficiently from classmates and teachers. Rather than emailing questions to the teaching staff, students are invited to post their questions on Piazza.

Teaching material

# Pre-requisites: relational model + SQL

- «Introduction to databases» **videlectures** on the **portal** <https://didattica.polito.it/>

Materials

- Introduction to the course (2 slides per page)
- Introduction to the databases (2 slides per page, 6 slides per page)
- Relational data model (2 slides per page, 6 slides per page)
- Relational algebra (2 slides per page, 6 slides per page)
- SQL language:
  - Basics (2 slides per page, 6 slides per page)
  - The SELECT statement: basics (2 slides per page, 6 slides per page)
  - Nested queries (2 slides per page, 6 slides per page)
  - Set operators (2 slides per page, 6 slides per page)
  - Update commands (2 slides per page, 6 slides per page)
  - Managing tables (2 slides per page, 6 slides per page)
- SQL language: other definitions
  - Management of views (2 slides per page, 6 slides per page)
  - Transactions (2 slides per page, 6 slides per page)
  - SQL for applications (2 slides per page, 6 slides per page)
  - Access control (2 slides per page, 6 slides per page)
  - Index management (2 slides per page, 6 slides per page)
- Database design
  - Design techniques and models (1 slide per page)
  - Conceptual design (1 slide per page)
    - Time representation (1 slide per page)
  - Logical design (1 slide per page)
  - Normalization (1 slide per page)

- «Introduction to databases» **slides** on the public web page <https://dbdmg.polito.it/wordpress/teaching/databases/>  
[https://dbdmg.polito.it/dbdmg\\_web/2022/introduction-data-bases/](https://dbdmg.polito.it/dbdmg_web/2022/introduction-data-bases/)

## Material

- **Introduction to the course** ([pdf](#))
- **Introduction to databases** ([pdf](#), [pdf6x](#))
- **Relational data model** ([pdf](#), [pdf6x](#)) [Last update: 10/03/2024]
- **Database design**
  - Entity-Relation Model ([pdf](#), [pdf6x](#)) [Last update: 13/03/2024]
  - Time representation ([pdf](#), [pdf6x](#)) [Last update: 04/04/2024]
  - Logical design – Translation ([pdf](#), [pdf6x](#)) [Last update: 19/03/2024]
  - Conceptual design example ([pdf](#), [pdf6x](#))
  - Logical design example ([pdf](#), [pdf6x](#))
  - Normalization ([pdf](#), [pdf6x](#)) [Last update: 04/04/2024]
- **Relational algebra** ([pdf](#), [pdf6x](#))
- **SQL language**
  - Fundamentals ([pdf](#), [pdf6x](#)) [Last update: 15/04/2024]
  - Nested queries ([pdf](#), [pdf6x](#)) [Last update: 16/04/2024]
  - Set operators ([pdf](#), [pdf6x](#)) [Last update: 20/04/2024]
  - Advanced queries ([pdf](#), [pdf6x](#))
  - Update commands ([pdf](#), [pdf6x](#))
  - Table management ([pdf](#), [pdf6x](#))
  - Advanced SQL ([pdf](#), [pdf6x](#))
- **SQL for applications**
  - Web applications ([pdf](#))
  - Streamlit ([pdf](#))
  - Live coding ([pdf](#))
- **Trigger**
  - Trigger ([pdf](#), [pdf6x](#)) [Last update: 24/5/2024]

Pre-requisites: relational model + SQL



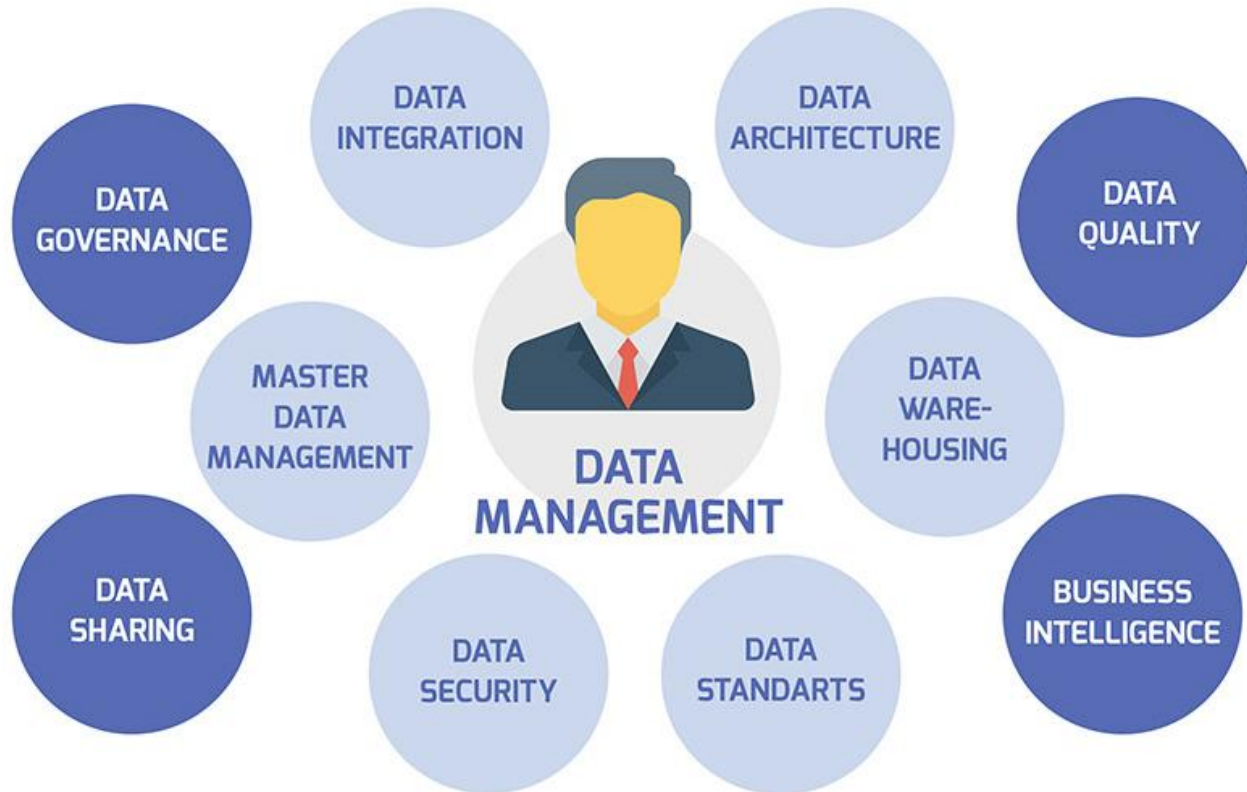
Please join the poll at

<https://www.menti.com/al3fu6qams14>

Enter code **1433 1103**



# What is data management? (1)



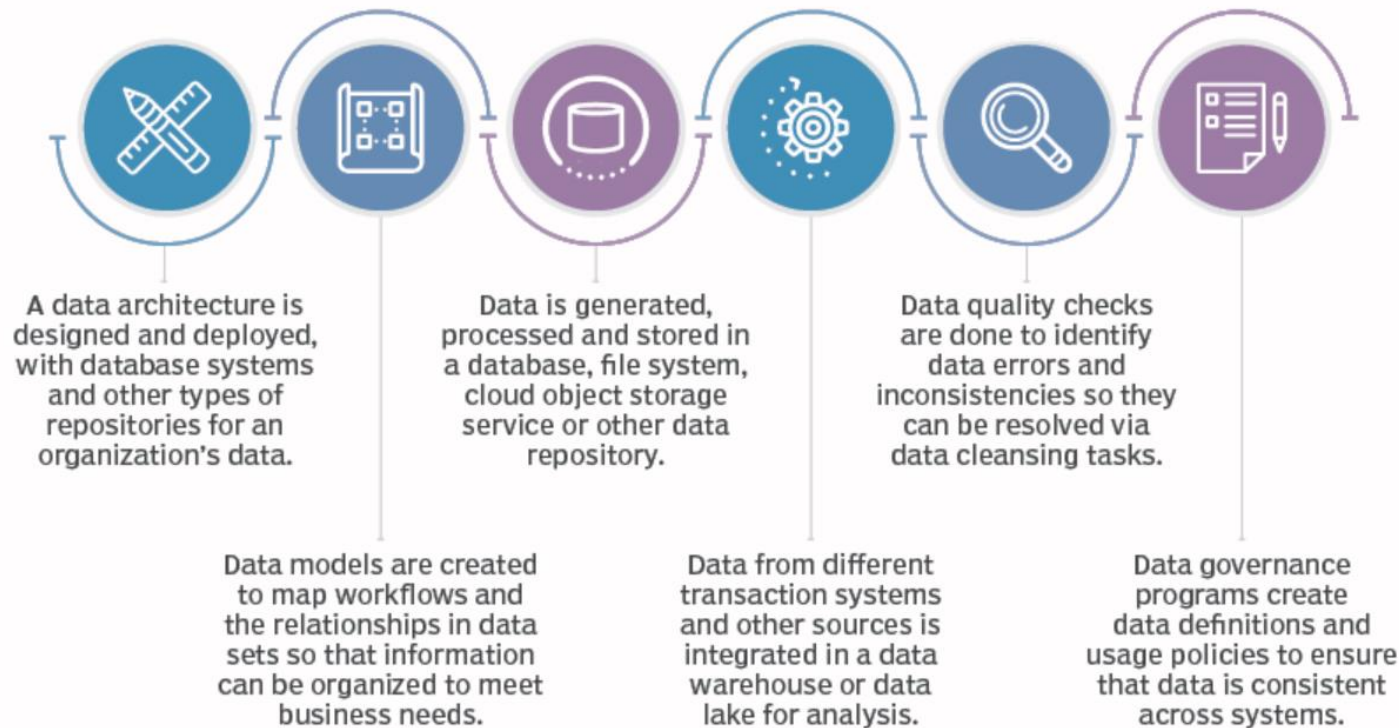
Data management is

- a **business practice**
- used in **organizing** and maintaining **data processes**
- that meet ongoing **information lifecycle** needs
- within **every company**.

A global need for data management began with the electronics era or digital age of data processing [...]

- [...] **acquiring, storing, protecting,** and in-depth **processing** required data
- to ensure the required **accessibility, reliability,** and **timeliness** of all data for its users

# What is data management? (2)

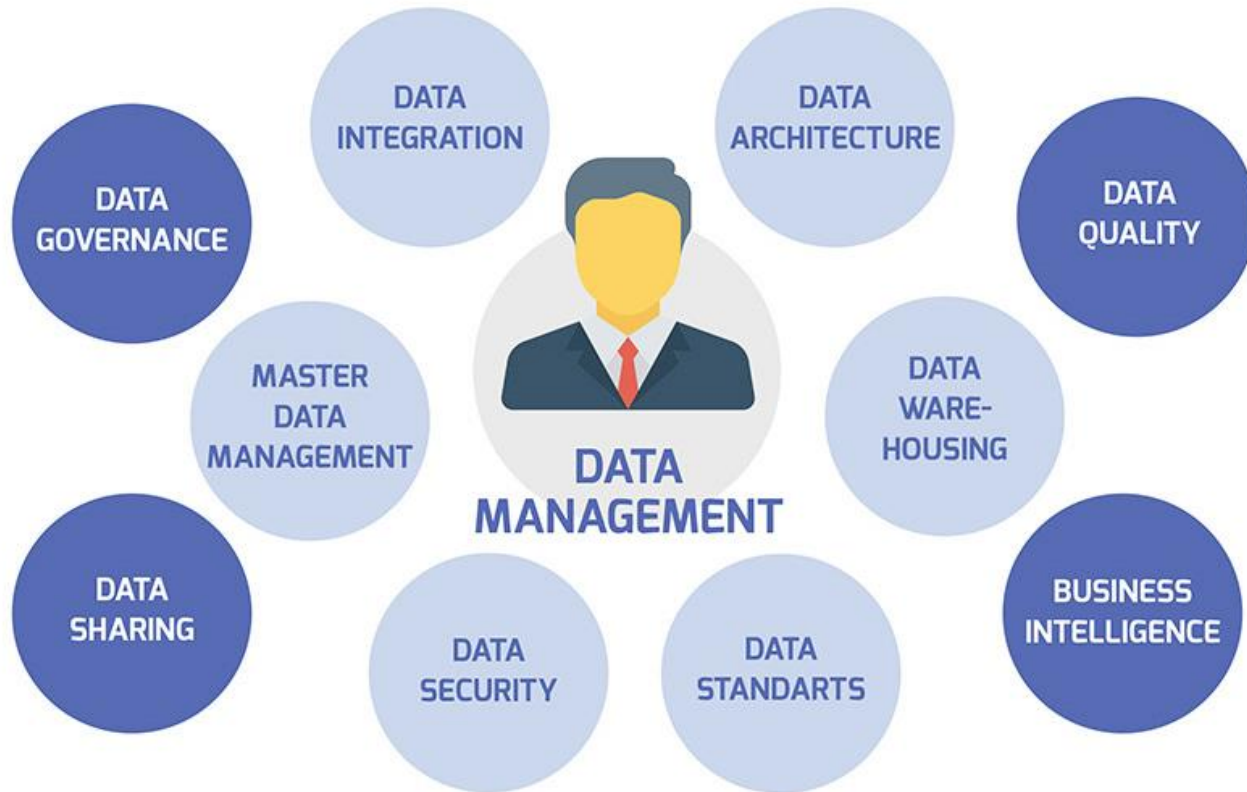


Data management is

- the process of **ingesting, storing, organizing** and **maintaining** the data created and collected by an organization.
- [...] deploying the **IT systems** that run business applications and provide **analytical information** to help drive operational **decision-making** and strategic planning by corporate executives, business managers and other end users.
- make sure that the data in corporate systems is **accurate, available** and **accessible**.

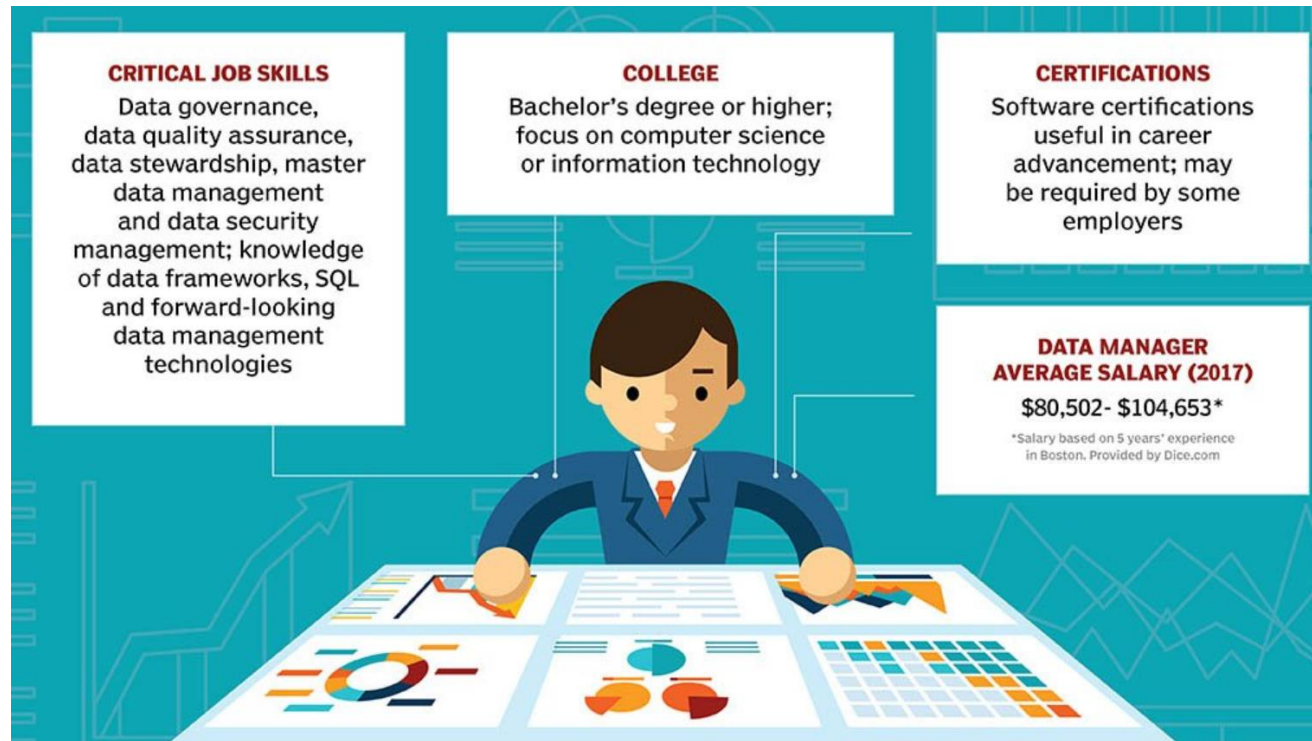


# Why data management? (1)



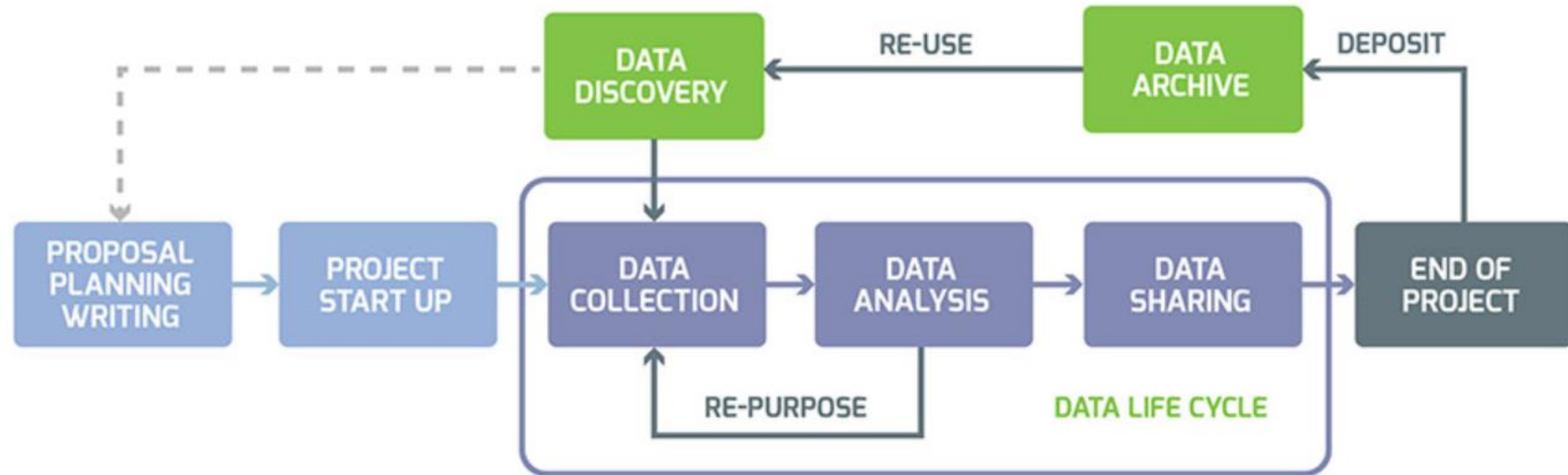
- [...] in the digital age, **data is king**. That is why it is seen as
- **one of the most important assets** of an organization;
- it is the foundation of information and the basis on which people **make decisions**.
- hence it would follow that if the data are **accurate, complete, organized and consistent**,
- it will contribute to the **growth** of the organization.

# Why data management? (2)



- Data are increasingly seen as a **corporate asset**
- used to make more-informed **business decisions**, [...], **optimize** business operations and reduce costs, all with the goal of **increasing revenue and profits**.
- a lack of proper data management can saddle organizations with incompatible **data silos**, **inconsistent** data sets and data **quality** problems [...] or, worse, lead to **faulty findings**.
- grown in importance as businesses are subjected to an increasing number of **regulatory compliance** requirements, e.g., data privacy and protection laws (GDPR)

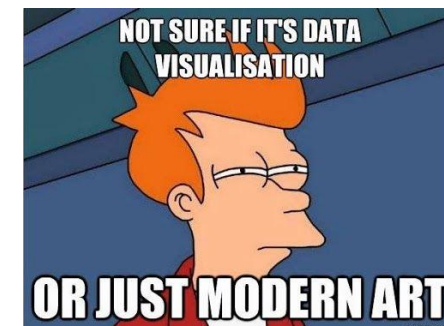
# A sample data-management process



# Data Visualization



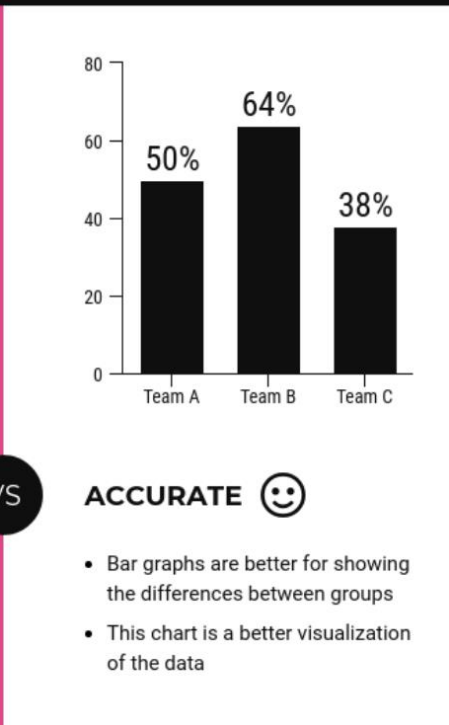
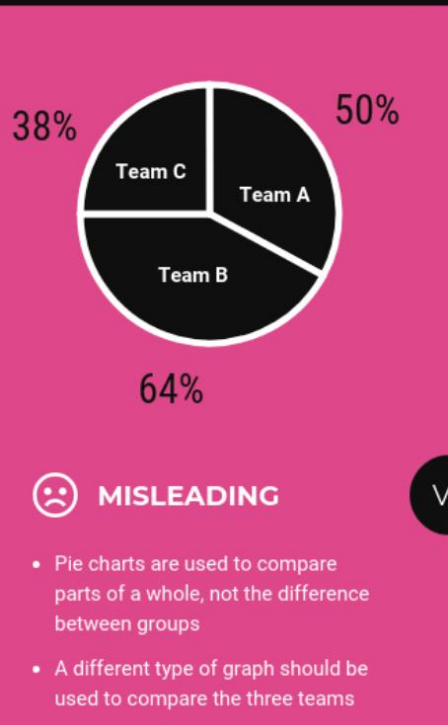
- Data visualization is the **visual presentation of data** or information.
- The goal of data visualization is to **communicate data** or information **clearly** and **effectively** to readers.
- It combines both **art** and **data science**: it should be creative, **pleasing** to look at, and **functional** in its visual communication of the data.





# USING THE WRONG GRAPH

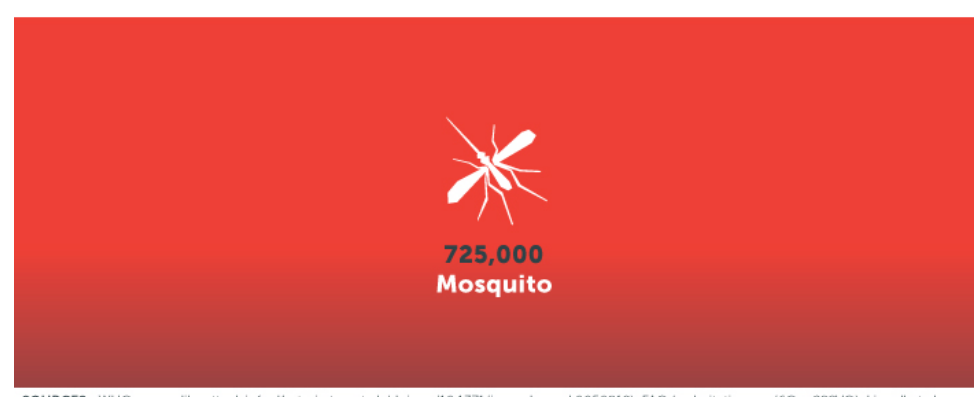
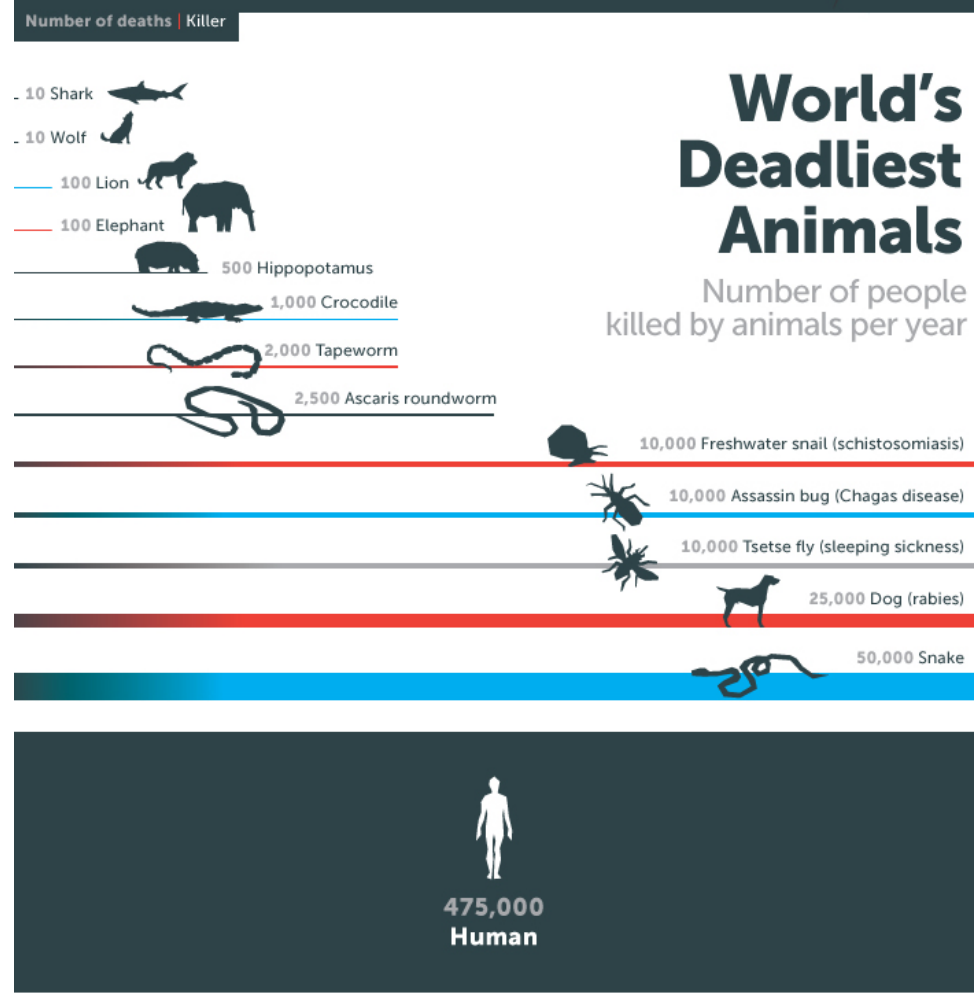
The type of graph you use should depend on the type of data you want to visualize. Using the wrong type of graph can skew the data. Writers will sometimes use the wrong type of graph on purpose.



VS

Visuals are processed **60,000 X** faster than text.

2



SOURCES: WHO; crocodile-attack.info; Kasturiratne et al. (doi.org/10.1371/journal.pmed.0050218); FAO (webcitation.org/6OgpS8SVO); Linnell et al. (webcitation.org/6ORL7DBUO); Packer et al. (doi.org/10.1038%2F436927a); Alessandro De Maddalena. All calculations have wide error margins.



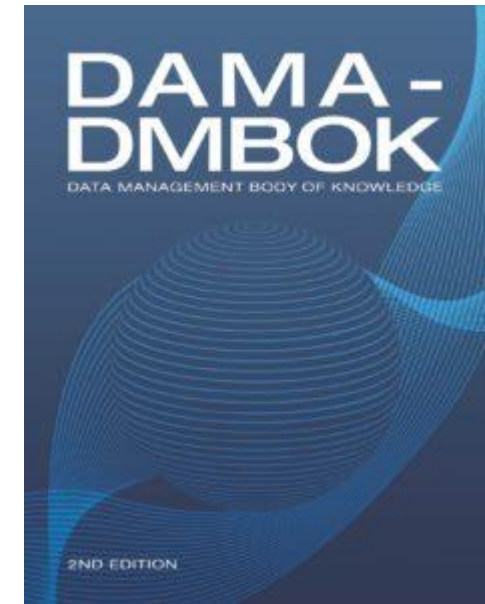
# Diving much deeper...

- **DAMA** International, the Data Governance Professionals Organization work to advance understanding of data management disciplines.
- They published DMBOK in **2009**, a reference book that attempts to define a standard view of data management functions and methods.
- The Data Management Body of Knowledge 2nd Edition, **2017**, (DMBOK2 for short)
- “Provides a functional framework for the implementation of enterprise data management practices; including widely adopted practices, methods and techniques, functions, roles, deliverables and metrics.”



<https://dama.org/content/body-knowledge>

- Data Handling Ethics
- Data Governance
- **Data Architecture**
- **Data Modeling and Design**
- **Data Storage and Operations**
- Data Security
- **Data Integration & Interoperability**
- **Document and Content Management**
- Reference and Master Data
- **Data Warehousing** and Business Intelligence
- Metadata Management
- **Data Quality** Management
- **Big Data** and **Data Science**
- Data Management Maturity Assessment
- Data Management Organization and Role Expectations
- Data Management and Organizational Change Management



# Course contents at a glance

## **Data Management**

- OLAP (Online Analytical Processing), multi-dimensional analytical queries
- Data Warehousing
- Data retrieval (querying and indexing)
  - extended SQL
  - specific NoSQL language
- NoSQL data management
- Data modeling
- Distributed data management

## **Data Visualization**

- Motivation and history of Data Visualization
- Visual perception and reasoning
- Graph construction principles
- Data quality

# Schedule

## Lectures

- **Monday**  
13:00-16:00  
classroom 1P + online
- **Thursday**  
16:00-19:00  
classroom 1P + online

Schedule changes announced on the teaching portal.

## Lab practice

- **Tuesday**  
11:30-13:00 team A  
13:00-14:30 team B  
Laboratory Laib3B (no online...)
- Starting on ...  
(tomorrow exception: recovery lecture)
- Till the end of the course
- Each student will be assigned to a single Team (**either A or B**)
  - based on their surname
  - changing Team is allowed upon request
- Simone Monaco, Diego Monti

# Exam

- See exam policy on the [official course web page](#) on the teaching portal
- Exam + Lockdown browser
- Bring your own notebook + test everything in advance
  - WiFi, power plug, updates...
- Students can fail the exam
  - scholarship...
- **OPEN BOOK!**

**Exam:** Computer-based written test in class using POLITO platform;

The exam lasts 90 minutes and consists of theoretical questions and written exercises, as described in the following:

- [max 6 points] 3-6 multiple-choice questions on theoretical topics of the course, such as conceptual, logical, and physical data warehouse design, extended SQL language, technological characteristics of NoSQL databases and their usage, data management issues in distributed (non-relational) databases, data visualization techniques
- [max 12 points] exercises on data warehousing, including 2-4 open and/or multiple-choice questions on data warehouse design, and 2-3 queries for data access in extended SQL (open questions with answers to be provided in a text box)
- [max 9 points] 1-2 exercises on NoSQL database design and 1-2 queries for data access (open questions with answers to be provided in a text box)
- [max 5 points] 1 exercise on visualization analysis and design with open questions (answers to be provided in a text box)

Students are not allowed to use textbooks, notes, or additional electronic devices during the exam, besides their own notebook with Lockdown/Respondus.

Exercises are evaluated according to the correctness of the proposed solution and to the appropriateness of the adopted resolution methodologies.

Specific points for each exercise are indicated in the exam text.

Multiple-choice questions have a penalty for wrong answers, whereas no-penalty no-points in case no answer is provided.

Learning objectives assessment.

The exam will assess:

- the knowledge of data warehouse architectures and of their design methodologies (conceptual, logical, and physical)
- the ability to design a data warehouse in a provided use case
- the ability to write extended SQL queries to extract data of interest from a data warehouse
- the knowledge of the main technological characteristics of NoSQL databases
- the ability to design NoSQL databases and to query NoSQL databases
- the ability to design dashboards and KPIs
- the knowledge of the basic principles of cognitive and perceptive aspects related to visualization, and of the main visualization techniques

Questions?





# Data Management and Visualization

INTRODUCTION TO THE COURSE

Daniele Apiletti