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**.

https://searchdatamanagement.techtarget.com/definition/data-management
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)

https://searchdatamanagement.techtarget.com/definition/data-management
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.

https://venngage.com/blog/data-visualization/
**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.

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

**Misleading VS Accurate**

- Pie charts are used to compare parts of a whole, not the difference between groups.
- A different type of graph should be used to compare the three teams.

**Bar graphs are better for showing the differences between groups.**

This chart is a better visualization of the data.

**World’s Deadliest Animals**

Number of people killed by animals per year:

- 1,000 Freshwater snail (schistosomiais)
- 10,000 Assassin bug (Chagas disease)
- 10,000 Tsetse fly (sleeping sickness)
- 25,000 Dog (rabies)
- 50,000 Snake

**Human**

725,000

**Mosquito**

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.
- “Provides a functional framework for the implementation of enterprise data management practices; including widely adopted practices, methods and techniques, functions, roles, deliverables and metrics.”

- 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

https://dama.org/content/body-knowledge
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
Teaching staff

• **Silvia Chiusano**
  - Relational data management
  - OLAP, Data Warehousing
  - Lectures

• **Diego Monti**
  - Data Visualization
  - Lectures and practices

• **Daniele Apiletti**
  - NoSQL data management
  - Data Warehousing exercises
  - Lectures and... everything (any other business)

• **Eliana Pastor**
  - practice of OLAP, Data Warehousing

• **Alessandro Fiori**
  - practice of NoSQL (MongoDB)

• **Simone Monaco**
  - assistance on Data Warehousing

For private issues, you can contact us via email at: name.surname@polito.it

or better ask general course-related questions on Piazza:

https://piazza.com/polito.it/fall2021/01txasm
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.

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Subject</th>
<th>Instructor/Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lunedì 08/11/2021</td>
<td>8:00</td>
<td>Data science lab: process a...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>9:00</td>
<td>Numerical optimization for ...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>10:00</td>
<td>Computational linear algebra...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>11:00</td>
<td>Decision making and optimiz...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td>Martedì 09/11/2021</td>
<td>8:00</td>
<td>Information Theory and Appl...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>9:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>10:00</td>
<td>Statistical methods in data...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>11:00</td>
<td>Numerical optimization for ...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td>Mercoledì 10/11/2021</td>
<td>8:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>9:00</td>
<td>Decision making and optimiz...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>10:00</td>
<td>Information Theory and Appl...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>11:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td>Giovedì 11/11/2021</td>
<td>8:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>9:00</td>
<td>Decision making and optimiz...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>10:00</td>
<td>Information Theory and Appl...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>11:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td>Venerdì 12/11/2021</td>
<td>8:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>9:00</td>
<td>Decision making and optimiz...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>10:00</td>
<td>Information Theory and Appl...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
<tr>
<td></td>
<td>11:00</td>
<td>Data management and visual...</td>
<td>A. A. Z. Z. - 0</td>
</tr>
</tbody>
</table>

*Notes:* Numbers in brackets indicate room or location details.
Schedule

Lectures

- **Monday**
  13:00-14:30
  14:30-16:00
  classroom R1 + online

- **Thursday**
  11:30-13:00
  classroom 3M + online

- **Friday**
  16:00-17:30
  online only

We will host some seminars toward the end of the course. Schedule changes announced on the teaching portal.

Lab practice

- **Tuesday**
  13:00-14:30 team A (Lab practice)
  14:30-16:00 team B (Lab practice)
  Laboratory LAIB4 + online

- Starting on **Tuesday, October 19, 2021**

- 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
Schedule

Go to www.menti.com

Enter the code 5670 3828
Material

- Announcements on the official «teaching portal» private page https://didattica.polito.it/
- Slides, texts of the practices, and all other materials are available on the public page of the course https://dbdmg.polito.it/
Pre-requisites: relational model + SQL

• «Introduction to databases» videolectures on the portal: https://didattica.polito.it/

• «Introduction to databases» slides on the public web page: https://dbdmg.polito.it/wordpress/teaching/databases/
Pre-requisites: **relational model + SQL**

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

- «Introduction to databases» **slides** on the public web page: https://dbdmg.polito.it/wordpress/teaching/databases/

Go to www.menti.com and use the code 26517433

Please join the poll here: https://www.menti.com/2zeyu73xj
Exam

• See exam policy on the official course web page on the teaching portal

• For students of the last academic year, mind the changelog
  • due to the Covid experience and the in-presence exam
  • Some improved topics: + NoSQL queries + extended NoSQL design patterns

Data management and visualization

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

- (max 5 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 electronic devices during the exam.

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?
INTRODUCTION TO THE COURSE
Academic Year 2021-2022
Daniele Apiletti