Technologies for Information Systems

Cinzia Cappiello
A.A. 2023-2024

These slides are based on Prof. Tanca slides. Thanks!

1

Teachers' info

Technical Lectures: Prof. Cinzia Cappiello
Communication part: Prof. Nicoletta Di Blas
Exercises: Ing. Bavaro Marcello

emails: name.surname@polimi.it

2

Course Overview

For the technical part:

- Analyze the data sources and design a **Data Integration** system

- Design a **Data Warehouse**

From the point of view of Communication, the students will learn how to clearly explain a technological or methodological issue at the correct level of abstraction, considering the common ground with the interlocutor

3

Course Material page - WeBeep

Technologies for Information Systems 2023-2024

Calendar September 2023

Lectures - Cinzia Cappiello

Date	Topic	Recordings' link
14/09/23	Introduction (slides)	

Exercise sessions - Marcello Bavaro

4

Further Information

Prerequisites: Databases I and (possibly) Databases II (the latter can also be attended in the same semester as TIS). For the students who have not attended DB1 there is a set of recorded lectures (Crash Course on Databases) that supplies the main fundamental topics.

The course is completely offered in English.

5

Exam

Theoretical questions	Exercises	➕	Communication Assignment

- The whole exam IS WRITTEN and consists of:
    - design exercises and questions on theoretical topics

    - a Communication Assignment (that must be passed in order to pass the exam), in which the student has to demonstrate the acquired communication skills.

6

Exam details

The exam consists of two main parts:
1.Theoretical Questions
    Format: 2 open-ended questions.
    Total Points: 10 (5 points each question).
    Minimum Required Points: 5.
2.Exercises
    Total Points: 22.
    Minimum Required Points: 13.

    The two parts can be taken in different exam calls, and that result will be retained (if sufficient) for future exam sessions (the entire academic year).

Please note: This is a closed-book examination.

7

Communication assignment

Only for the Communication part of the exam, the mark is a pass/no pass.

Once a student successfully passes the Communication component of the exam, that result will be retained for future exam sessions (the entire academic year). Even if the student does not pass the other exam sections, there is no need to retake the Communication part in subsequent exam calls.

8

Exam enrollment

In the enrollment phase of the exam the student will have the possibility to select the part or the parts that she/he wants to take.

9

WHAT'S NEW IN INFORMATION TECHNOLOGIES?

10

The Challenges For Modern Information Systems

✓ Pre-history: focused on challenges that occur *within enterprises*

✓ The Web era:
  • scaling to a much larger number of sources
  • Handling sources with less structure.

✓ Nowadays:
  • large scientific experiments rely on data management for progress.
  • People continuously create data fragments by interacting with services on the Web.
  • Massive use of social media and smart devices
  • User-generated content merges with the Internet of Things
  • Users *as sensors and actuators*
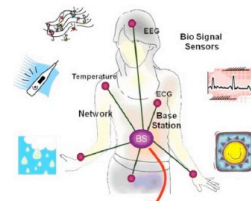
How do we make sense of this mass of data?

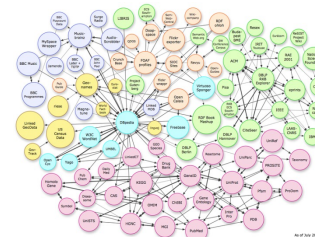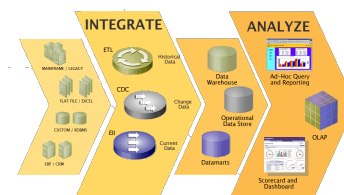11

The four V's of Big Data

Volume
Velocity
Variety
Veracity

We should be able to govern:
  – data abundance
  – data and user dynamicity and mobility
  – heterogeneity, data semantics
  – incompleteness/uncertainty
  – interaction with the real-world

Making sense of all this data:
→ Extract useful knowledge

12

The new challenges in the Data Management Area

From the Seattle Report on Database Research (*) - what has changed in the last few years:

- **Technological breakthroughs in machine learning (ML) and artificial intelligence** (AI): rise of data science as a discipline that combines elements of data cleaning and transformation, statistical analysis, data visualization, and ML techniques
- **Cloud computing has become mainstream**. The industry now offers on-demand resources that provide on-demand, elastic storage and computation services.
- For cloud-based systems, the industry has converged on **data lakes** as novel systems for data integration
- **Modern data warehousing** query engines
- Society has become more concerned about the state of **data governance**, concentrated in data usage, e.g. quality-aware, privacy-aware, ethical and fair use of data

(*) Communications of the ACM, August 2022, Vol. 65 No. 8, Pages 72-79

13

The new challenges in the Data Management Area (2)

Extraction of (synthetic and useful) knowledge: build environments that mimic the progressive inspecting, observing, surveying activity with which users make decisions.

Massive data analysis and processing: A process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

Massive data integration: People and enterprises need to integrate data and the systems that handle those data: Relational DBMSs and their extensions, legacy data and legacy DBMSs, sensors and user-generated content produce structured or unstructured data

Data warehousing: A single, complete and consistent store of data obtained from a variety of different sources made available to end users, so that they can understand and use it in a business context.[Barry Devlin]
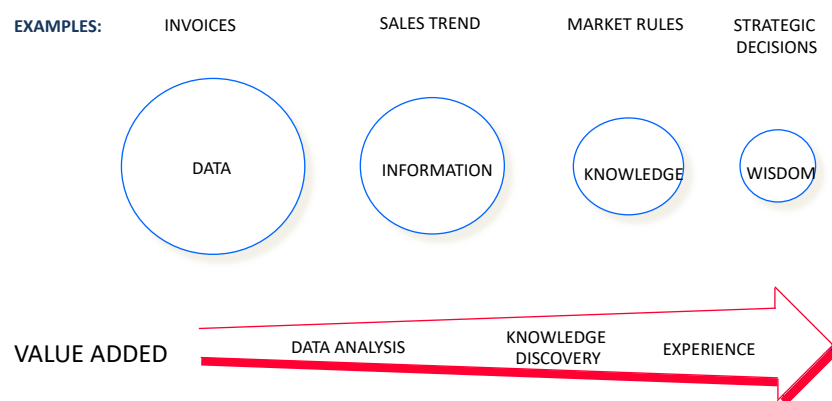
14

The new challenges in the Data Management Area (3)

Extraction of (synthetic and useful) knowledge:

- *Knowledge representation and reasoning*: using conceptual models and ontologies: formal specifications allowing offering a common vocabulary for automatic knowledge sharing; using reasoning services, which allow some forms of deduction and inference.

- *Personalization and context-awareness*: can eliminate "information noise" reducing the available data only to the part that is appropriate for the current user and context

15

Knowledge Distillation

EXAMPLES:	INVOICES	SALES TREND	MARKET RULES	STRATEGIC DECISIONS
	DATA	INFORMATION	KNOWLEDGE	WISDOM

VALUE ADDED — DATA ANALYSIS — KNOWLEDGE DISCOVERY — EXPERIENCE

(courtesy of prof. Fabio A. Schreiber)

16