# TECNOLOGIES FOR INFORMATION SYSTEMS

## INTRODUCTION

Prof. Letizia Tanca
http://tanca.faculty.polimi.it/

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

# COURSE OVERVIEW

The course gives an overview of some of the most advanced methodologies adopted to solve the conceptual and technological problems encountered in modern information system design and operation, spanning from the currently consolidated methodologies to the most advanced ones.

**ORGANIZATION:**

- 30 lecture hours (MYSELF)
- 20 exercise hours (DR. PANIGATI)

## IMPORTANT:

- Sometimes the course schedule may change because of problems or unenvisaged, undeferrable engagements
- Please always watch the course site for news !!!

# Presentation Overview

- Motivation
- Description of the course scope and contents
- Practical Information

# THE CHALLENGES FOR MODERN INFORMATION SYSTEMS

- Organizations grow and generate more information: they capture billions of bytes of information about their customers, suppliers and operations
- The pervasiveness of digital technologies has changed the way individuals interact with the external world (sensor technologies) and with one another (social media), generating a huge mass of content

→ Data has become as a torrent that flows through all possible digital channels.

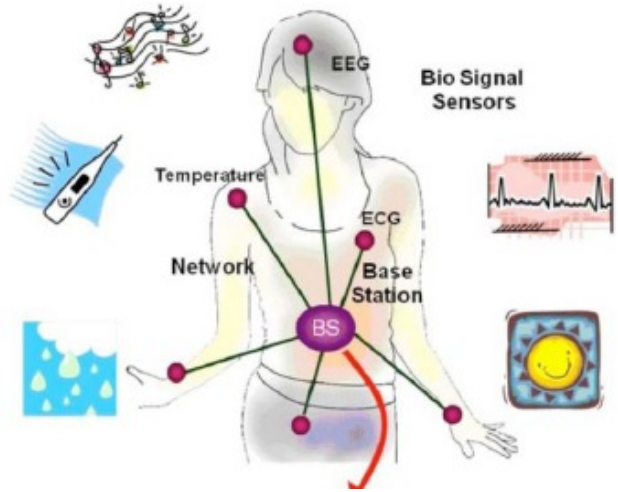
# THE CHALLENGES FOR MODERN INFORMATION SYSTEMS


## Challenges

- Internet of Things
  - Autonomic data-sources
  - Data and Users Mobility
  - Embedded systems Databases
  - Uncertainty and Lineage Management
- Multimodality
- Ambient and body intelligence
- Information noise

## Problems

- Past problems are not completely solved.
- Missing or expensive infrastructures.
- Sound and complete data integration in open world is unpractical.
- 85% of interesting data is unstructured!



# THE CHALLENGES FOR MODERN INFORMATION SYSTEMS (II)

- 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?

# *Information overload*

- The term was used by Alvin Toffler in his book *Future Shock*, already back in 1970
- It refers to the difficulty of *understanding and making decisions* when too much information is available
- This is the main challenge presented by “Big Data”

# Information Systems in the Era of Big data

**Extraction of (synthetic and useful) knowledge:**

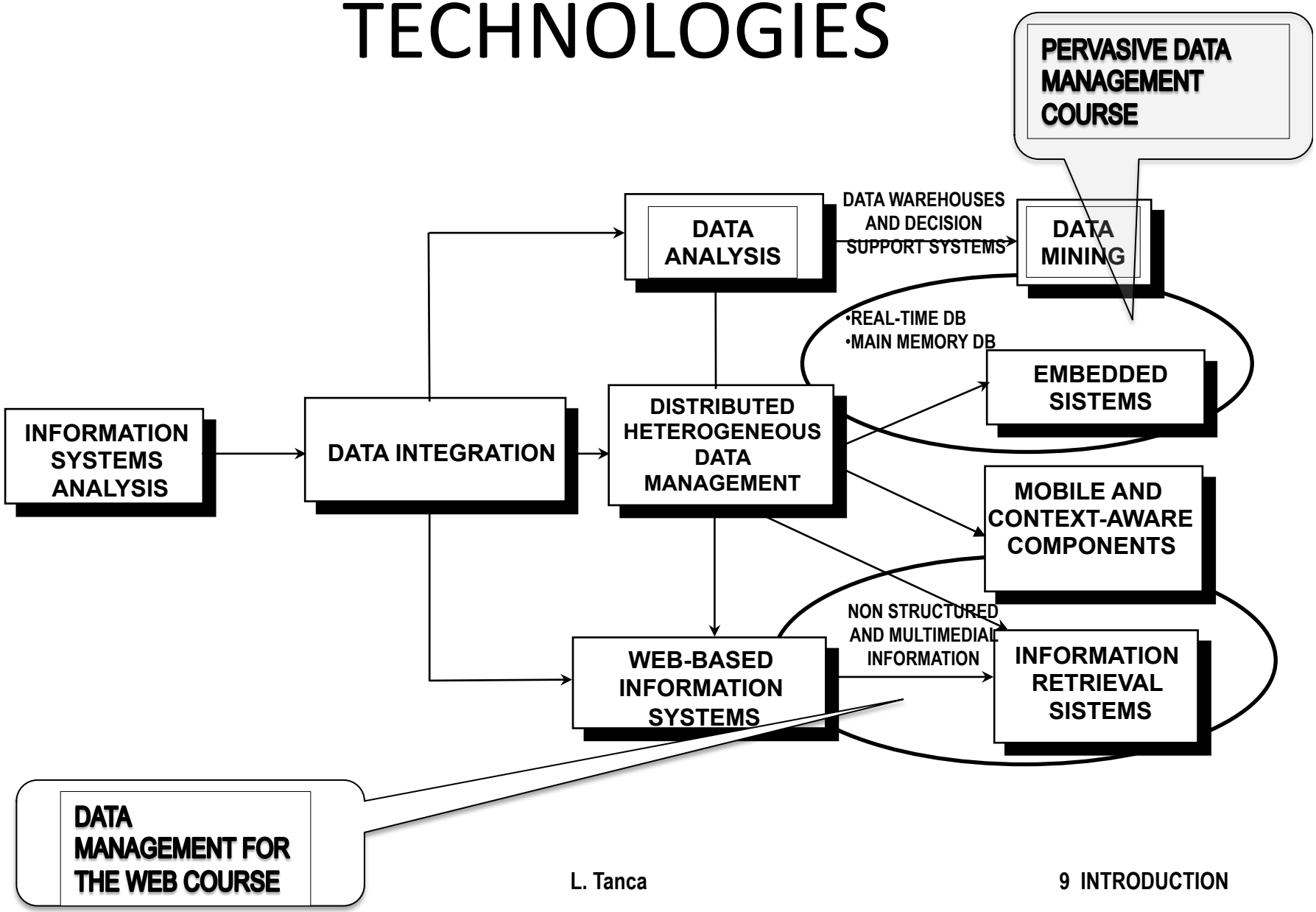
- ***Massive data integration:*** People and enterprises need to integrate data and the systems that handle those data: Relational DBMSs and their extensions, legacy data and legacy DBMSs, structured or unstructured data
- ***Massive data analysis and processing:*** data analysis and data mining research focuses on studying algorithms and techniques to find interesting patterns representing implicit knowledge stored in massive data repositories, useful to generate concise models of the analyzed 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]

# Information Systems in the Era of Big Data (continued)

**Extraction of (synthetic and useful) knowledge:**

- ***Knowledge representation and reasoning:*** using *conceptual models and ontologies*, formal specifications allows for use of 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
- ***Build environments which mimick the progressive inspecting, observing, surveying activity with which users take decisions***

# INFORMATION MANAGEMENT TECHNOLOGIES



# LECTURES AND EXERCISES

**Information System Architectures and Heterogeneous Data Integration: structured and non-structured data (12 hrs lectures, 10 hrs exercises):**

Introduction to the architectures of modern information systems

Data heterogeneity: model heterogeneity, semantic heterogeneity at the schema level, heterogeneity at the data level

dynamic data integration, wrappers and mediators

integration based on meta-models

**Data Warehousing and Analysis (10 hrs lectures, 10 hrs exercises):**

Data Warehouse Architecture and Design

Data Mining and its Applications, introduction to data exploration

**Time Representation and Management in Information Systems (4 hrs lectures):**

Time Ontology

Temporal Databases

**Advanced topics (4 hrs lectures):**

Big Data analysis techniques, introduction to data exploration, intensional data representation, personalization and context-awareness

# FURTHER INFORMATION

- Prerequisites: Data bases I and Data bases II
- The course is completely offered in English
- The exam IS WRITTEN and consists of design exercises and questions on theoretical topics
- Support materials available on the course web site, reachable from my personal web page

# Personal info

- Prof. Letizia Tanca
  - Receiving time: watch PoliSelf
    - currently Wednesdays 14:30-16:30, email: letizia.tanca@polimi.it
  - tel: 02-2399-3531, fax: 02-2399-3411
  - Personal web site: http://tanca.faculty.polimi.it/ containing:
    - The course page
    - Information on my group's research interests, for students who are interested in projects (for example within the course Pervasive Data Management) and theses