



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Data Management and Visualization

2021-1-F9101Q037

---

#### Aims

At the end of the module students will be able to select, design and query a database (relational or not) according to their application needs

Students will be able to use a NoSql database management system to acquire, memorize and query semi structured data

At the end of the course students will have acquired skills in analysis, evaluation and, to a lesser extent, development of complex and interactive infographics.

#### Contents

Introduction to data management in big data context

data lifecycle

Variety: nosql models and architecture

Volume: data distribution and replication, hadoop architecture

Velocity: data architecture for capturing and elaborating near real time data

## Detailed program

1. **Introduction to big data (variety, volume and velocity )**
2. **Data life cycle**
3. **Variety**
  1. Introduction to NoSQL models
  2. Cap Theorem
    1. key value and columnar models
    2. Document based system
  3. Graph db
  3. Data integration
  4. Data quality
4. **Volume**
  1. Data distribution
  2. Replication
  3. hadoop architecture
  4. Data lake
5. **Velocity**
  1. Lambda and Kappa architecture
  2. ELK architecture

### Data visualization

- Introduction to the Human Data Interaction (Definitions, main concepts and methodologies)
- Data Transformation into sources of knowledge through visual representation.
- Requirements and heuristics for high-quality visualizations: dos and donts.
- Charts and standard views: relevance and appropriateness.
- Advanced and innovative tools for data visualization and advanced quantitative analysis.
- The evaluation of the quality of visualizations and infographics.
  - o Qualitative assessment: expert and heuristic;
  - o Quantitative assessment: user tasks; inferential statistical techniques.
  - o Validated psychometric questionnaires and their analysis and understanding.
- Elements of visual semiotics and social semiotics.

## Prerequisites

knowledge of relational model

## Teaching form

Lectures and exercises in the classroom and on virtual lab

Lectures with the support of slideware, discussion of practical cases through the forum, discussion of practical home-work projects.

During the COVID-19 emergency the course will be delivered remotely asynchronously. Some videoconferencing events will also be organized as self-assessment tests, not considered for the final evaluation

## Textbook and teaching resource

G. Harrison Next Generation Databases, Apress, 2015

A. Rezzani Big data analytics Apogeo 2017

Yau, N. (2011). *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons.

Ware, C. (2012). *Information visualization: perception for design*. Elsevier.

Scientific articles and class pack provided by the lecturers.

## Semester

first semester

## Assessment method

written exam related to syllabus topics (40% of the final score) and project work realized alone or in group involving at least 2 of 3 V (60% of the final score)

## Office hours

Please send an e-mail to the teacher to arrange an appointment

---