UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Management

2223-1-FDS01Q001-FDS01Q001M

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**
4. Introduction to NoSQL models
5. Cap Theorem
6. key value and columnar models
7. Document based system
8. Graph db
9. Data integration
10. Data quality
11. **Volume**
12. Data distribution
13. Replication
14. hadoop architecture
15. Data lake
16. **Velocity**
17. Lambda and Kappa architecture
18. 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.

Someelf-assessment tests, not considered for the final evaluation will be provided

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

The exam is divided into two parts

Data Management (50% of the final evaluation): Written exam and a project related to the topic of the module

Data visualization(50% of the final evaluation): test and a project related to the topic of the module

Office hours

Please send an e-mail to teachers to arrange an appointment

Sustainable Development Goals

INDUSTRY, INNOVATION AND INFRASTRUCTURE