UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Management

2526-1-FDS02Q001-FDS02Q00101

Aims

At the end of the module students will be able to define and implement a complete data pipeline from data acquisition to data storage (relational or not) based on their application needs

Students will be able to evaluate for each phase the appropriate tool to use

Contents

Introduction to data management in big data context

data life cycle
data acquisition techniques
data modelling and storage
data preparation, cleansing, quality and explorative data analysis
advance data management concepts

Detailed program

1. **Data life cycle**
2. **data acquisition**
3. web scraping
4. rest api
5. real time data acquisition
6. use of LLM

7. **data storage and modelling**
8. Introduction to NoSQL models
9. key value and columnar models
10. Document based system
11. Graph db

6.**data preparation, cleansing, quality and explorative data analysis**

12. Data integration
13. Data quality
14. **Advanced data management concepts**
15. data warehouse
16. data lake

Prerequisites

knowledge of the relational data model, SQL language, at least one programming language

Teaching form

the course hours are 46 organized as follows
28 hours of lessons
18 hours of laboratory

all hours will be delivered in presence and in interactive mode

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 with open questions and discussion on an integration and analysis project of at least two different data sources acquired with at least one of the data acquisition techniques seen in the lessons. the project can be carried out in groups of 2-3 people

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

Office hours

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

Sustainable Development Goals

QUALITY EDUCATION | INDUSTRY, INNOVATION AND INFRASTRUCTURE