



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Management

2425-1-FDS01Q001-FDS01Q001M

Obiettivi

Alla fine del modulo gli studenti saranno in grado di definire e realizzare una pipeline di dati completa dall'acquisizione dei dati all'archiviazione dei dati (relazionale o meno) in base alle loro esigenze applicative

Gli studenti saranno in grado di valutare per ogni fase lo strumento appropriato da utilizzare

Contenuti sintetici

Introduzione alla gestione dei dati nel contesto dei big data

ciclo di vita dei dati
tecniche di acquisizione dati
modellazione e archiviazione dei dati
preparazione dei dati, pulizia, qualità e analisi esplorativa dei dati
concetti avanzati di gestione dei dati

Programma esteso

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

Prerequisiti

conoscenza del modello relazionale dei dati, linguaggio SQL, almeno un linguaggio di programmazione

Modalità didattica

le ore del corso sono 46 così organizzate
28 ore di lezione
18 ore di laboratorio

tutte le ore saranno erogate in presenza in modalità interattiva

Materiale didattico

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.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

L'esame è diviso in due parti

Data management (50% della valutazione finale): Esame scritto con domande aperte e discussione su un progetto di integrazione e analisi di almeno due sorgenti dati diverse acquisite con almeno una delle tecniche di acquisizione dati viste a lezioni. il progetto può esse svolto in gruppi di 2-3 persone

Visualizzazione dei dati (50% della valutazione finale): test e un progetto correlato all'argomento del modulo

Orario di ricevimento

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

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ | IMPRESE, INNOVAZIONE E INFRASTRUTTURE
