UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Large Scale Data Management

2627-2-F1802Q130

Aims

The student will acquire competences related to the following activities: (i) analysis and reconciliation of heterogeneous data sources, (ii) definition of schemas of fact, (iii) big data techniques

Contents

Data warehouse refers to the set of methods, techniques, and tools able to integrate and manage data of a given organization for supporting strategic decision. In the first part of the course, provided in Italian, it will be shown design and develop methodolodology of data warehouse. In the second part of the course the big data architecture and its relationship with data warehouse will be discuss. A number of real cases presented by companies will complete the course

Detailed program

1 Analysis and reconciliation of heterogeneous operational data:

a. techniques for the comparison and for the alignment of conceptual schemas;

b. techniques for the integration and for the restructuring of conceptual schemas.

2 Conceptual design related to the definition of fact schemas:

a. the Dimensional Fact Model as a graphical conceptual model for the definition of fact schemas;

b. a methodology for the definition of fact schemas starting from E-R schemas.

3 Big data,

Big data

a. data lake, data lakehouse architectures

b. data analysis frameworks: Apache Spark

d. near real time data architecture

e. big data models: apache iceberg, arrow, parquet

f. Data Lakehouse

4. Introduction to data visualization

a. Introduction to key components of data visualization and data drive storytelling

b. Good and worst practice in the design and developement of visual interface

5. LLM e multi agent architecture

a. Usage of Generative AI for data management

b. Multi agent architecture for large scale data management

Prerequisites

Basic knowledge on database design. Knowledge of the Entity-Relationship (E-R) model.

Teaching form

lessons and exercises in Italian

Textbook and teaching resource

Matteo Golfarelli e Stefano Rizzi. Data Warehouse – Teoria e pratica della progettazione (seconda edizione). McGraw-Hill.

Semester

first semester

Assessment method

design and development of a small datawarehouse or a big data analysis project

Office hours

on appoitment

Sustainable Development Goals

QUALITY EDUCATION | INDUSTRY, INNOVATION AND INFRASTRUCTURE