UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Architecture

2526-1-F1802Q100

Aims

The course, delivered in Italian, aims to provide knowledge and skills in the field of modern data architectures. Both the theoretical and research aspects that are still open and how these technologies can be used to solve specific application contexts will be presented. Particular attention will be paid to the study and analysis of how the architectures and solutions presented allow the execution of writing, even concurrent in a distributed environment, and reading in a distributed environment and in the presence of possible malfunctions on the network. The use of generative AI models in data management will be also discussed
At the end of the course the student will be able to choose the most appropriate architectural solution for data management

Contents

Through the presentation of different use cases, the theoretical, methodological and application aspects of the course will be addressed. Each use case intends to present problems for which new solutions are foreseen in architectural terms compared to what has been seen up to that point

1. distributed relational systems
2. non-relational and polyglot systems
3. data centric ai (data management for machine learning)
4. Generative AI for data management

Detailed program

User cases 1

references to centralized relational architectures: transactions, query optimization
distributed dbms architectures: two phase commit protocol, distributed deadlock
Use case 2
non-relational models
distributed architectures of non-relational systems
polyglot systems
Use case 3
data management for machine learning
data understading
data validation
data preparation (integration, quality, fusion)
feature engineering
MLOps
Use case 4
Generative Ai models
Fine tuning
RAG architectures

Prerequisites

Knowledge on data models, relational model and Entity Relationship model, and database design methodologies is useful, but non mandatory

Teaching form

A total of 48 hours are scheduled for the course; so organised
32 hours of in-person classroom lessons
16 hours of in-person classroom exercises

both the hours of lessons and exercises will be interactive. The teacher will ask questions to develop the students' critical reasoning skills

Textbook and teaching resource

slides will be published on the Elearning platform

Text book

Next Generation Databases: NoSQL, NewSQL, and Big Data

by Guy Harrison Publisher: Apress

Release Date: January 2016

ISBN 9781484213292

Semester

second semester

Assessment method

Written exam with open-ended questions on the course contents, numerical exercises possibly followed by an oral exam at the teacher's request

Alternatively, project relating to one of the use cases presented during the course. the project can be extended to a thesis and can also be done with groups of students. the topic of the project must be agreed with the teacher

Office hours

Please send an email in advance indicating the name of the course. the reception can also be carried out online or at the end of the lessons

Sustainable Development Goals

QUALITY EDUCATION