UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Large Scale Data Management

2526-2-F1801Q147

Obiettivi

Lo studente acquisirà competenze relative alle attività di (i) analisi e riconciliazione di sorgenti operazionali eterogenee, (ii) progettazione concettuale relativa alla realizzazione di schemi di fatto, (iii) utilizzo architetture di big data per la gestione e l'analisi dei dati

Contenuti sintetici

Il termine data warehouse si riferisce ai metodi, alle tecniche e strumenti in grado di aggregare ed elaborare i dati di una organizzazione per condurre analisi di tipo strategico. Nella prima parte del corso, erogato in lingua italiana, verranno presentare le metodologie di progettazione e sviluppo di un data warehouse. Nella seconda parte del corso si approfondiranno le architetture di big data e la loro relazione con le soluzioni di datawarehouse esistenti.

Programma esteso

1 "Analisi e riconciliazione di sorgenti operazionali eterogenee:

a. tecniche per la comparazione e l’allineamento di schemi concettuali;

b. tecniche per integrazione e ristrutturazione di schemi concettuali.

2 Progettazione concettuale relativa alla realizzazione di schemi di fatto:

a. il Dimensional Fact Model come modello concettuale grafico per la definizione di

schemi di fatto;

b. metodologia per la definizione di schemi di fatto a partire da schemi E-R.

3 Big data

a. architetture data lake, data lakehouse

b. framework di analisi dati: Apache Spark

d. Architetture per dati in near real time

e. modelli per big data: apache iceberg, arrow, parquet

f. Data Lakehouse

Prerequisiti

Conoscenze di base sulla progettazione di basi di dati. Conoscenza del modello entità-relazione (E-R)

Modalità didattica

Lezioni frontali ed esercitazioni in lingua italiana erogata in modalità interatitiva

Materiale didattico

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

Periodo di erogazione dell'insegnamento

primo semestre

Modalità di verifica del profitto e valutazione

sviluppo di un progetto di datawarehousing o di big data analisys e relativa discussione

Orario di ricevimento

per appuntamento

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ | IMPRESE, INNOVAZIONE E INFRASTRUTTURE