

CORSO: Data Management for Big Data Analysis

Docente responsabile: Danilo Croce – croce@info.uniroma2.it

Teaching Assistant: Andriy Shcherbakov – shcherbakov@revealsrl.it

DESCRIZIONE DEL CORSO

Questo corso mira a fornire una solida base di conoscenza nei formalismi, nelle metodologie e nelle tecnologie utilizzate nella gestione dei dati, con un focus specifico sugli standard industriali legati ai moderni sistemi di gestione dei database nell'era dei Big Data. Durante il corso, verranno esplorate le pratiche industriali chiave relative ai linguaggi di progettazione (come ER) e definizione dei dati (compresi ER e SQL). Sarà data particolare attenzione alle attuali tendenze nella progettazione dei database, con un'enfasi speciale sulle soluzioni non relazionali (noSQL) utilizzate nei contesti di analisi dei Big Data.

OBIETTIVI DI APPRENDIMENTO

- ✓ Introduzione ai formalismi, metodologie e tecnologie per la gestione dei dati.
- ✓ Approfondimento degli standard industriali dei Database Management Systems nell'era dei Big Data.
- ✓ Insegnamento delle prassi principali nei linguaggi di progettazione e definizione dei dati (ER, SQL).
- ✓ Introduzione al linguaggio e database SQL.
- ✓ Esplorazione delle tendenze attuali nella progettazione delle basi di dati.
- ✓ Enfasi sulle soluzioni non relazionali (noSQL) per Big Data Analytics.

METODOLOGIA

- Lezioni teoriche.
- Esercitazioni pratiche.
- Laboratori con il computer.

VALUTAZIONE

Esame scritto, peso: 70%; Lavoro di gruppo, peso 30%

PROGRAMMA

1. Introduzione ai Sistemi di Gestione delle Basi di Dati
2. Il modello relazionale
 - a. Il modello Entità-Relazione
 - b. Dall'algebra relazionale al linguaggio SQL
3. La progettazione logica: dallo schema ER al relazionale
4. Introduzione allo scenario dei Big Data
 - a. Introduzione alla Ottimizzazione delle Interrogazioni in scenari Big Data
 - b. La gestione dei Dati nei Big Data
5. Introduzione a NoSQL

Laboratori

1. Applicazioni e Basi di Dati: tradizionali
2. Laboratorio di MySQL/MariaDB
3. Laboratorio di MongoDB

Libri di Testo

- Materiale Didattico fornito a lezione.
- "Database Management Systems", 3rd Edition, Raghu Ramakrishnan e Johannes Gehrke, McGraw Hill, 2000 (ISBN 0-07-232206-3)
 - Versione Italiana. "Sistemi di Basi di Dati", di Raghu Ramakrishnan e Johannes Gehrke, Edizione Italiana, McGraw Hill, 2004