

Basi di Dati e di Conoscenza

Corso di Laurea: Corso di Laurea in Informatica;

Crediti: 12

Durata: Annuale

Titolare del corso: Prof.ssa Paola Vocca

Presentazione

Il Corso di Basi di Dati e di Conoscenza ha lo scopo di far acquisire allo studente una buona conoscenza dei principi che governano la progettazione di basi di dati e di conoscenza sia tradizionali sia più avanzati come database distribuiti e NoSQL ed interazione con applicativi. Il Corso propone i concetti basilari dell'algebra e del calcolo relazionale e li declina nello studio del modello Entità-Relazione e, più in generale, delle tematiche relative alla progettazione di basi di dati relazionali. Inoltre, obiettivo formativo del Corso è fornire allo studente una conoscenza nel dettaglio sul funzionamento dei sistemi di gestione delle basi di dati fondati sul linguaggio SQL (Standard Query Language) e l'impiego di quest'ultimo come linguaggio di definizione e manipolazione della base di dati.

Obiettivi Formativi

Il Corso di Basi di Dati ha i seguenti obiettivi formativi:

- Illustrare il modello relazionale per la rappresentazione dei dati.
- Illustrare le nozioni base di Algebra e Calcolo Relazionale.
- illustrare le tecniche di progettazione di basi di dati fondate sul modello Entità-Relazione.
- Illustrare i sistemi di gestione delle basi di dati fondati sul linguaggio SQL.
- Illustrare le problematiche relative alle basi di dati relazionali.
- Illustrare l'interazione con basi di dati relazionali tramite l'uso di SQL come Data Definition Language (DDL) e Data Management Language (DML).
- Illustrare le tecniche di programmazione avanzata di base di dati
- Illustrare le tecniche di ottimizzazione.
- Illustrare le tecniche di gestione delle transazioni e il controllo della concorrenza
- Illustrare i concetti di basi di dati distribuiti
- Illustrare i Database NoSQL
- Illustrare il trattamento dei big data con MapReduce e Spark

Conoscenza e capacità di comprensione

Lo studente al termine del Corso avrà conoscenza delle problematiche relative alle basi di dati relazionali, e ai relativi sistemi di gestione fondati sul linguaggio SQL (Standard Query Language) e alla progettazione di basi di dati basati sul metodo Entità-Relazione. Inoltre, tramite le esercitazioni e il laboratorio gli studenti acquisiranno la capacità di implementare i propri progetti, basati su casi di studio, in appositi software per la gestione di un DBMS (DataBase Management System) e delle Basi di dati NoSQL. Base di dati distribuite.

Conoscenze e capacità di comprensione applicate:

Lo studente sarà in grado di utilizzare strumenti, ambienti di sviluppo e metodi per affrontare problemi di progettazione di basi di dati; sarà inoltre in grado di utilizzare le funzionalità di base dei sistemi di gestione di basi di dati, come la definizione della base di dati, l'aggiornamento della base di dati, e le interrogazioni sulla base di dati, nell'ambito dello sviluppo e dell'esercizio di sistemi informativi

Autonomia di giudizio

Lo studente sarà in grado di effettuare una analisi di un sistema di basi di dati complesso e sarà in grado di progettarlo e di realizzarlo in un apposito software per la gestione di una base di dati (DBMS).

Abilità comunicative

Lo studente sarà in grado di descrivere e sostenere conversazioni su tematiche relative alla realizzazione di basi di dati, adoperando una terminologia adeguata alla descrizione dei processi di analisi e di progettazione di basi di dati, con soluzioni software ad hoc a seconda del tipo di sistema informativo da realizzare.

Capacità di apprendere

Lo studente al termine del Corso avrà conoscenza delle problematiche nella progettazione e gestione di una base di dati relazionale. Tutto ciò gli consentirà di proseguire gli studi ingegneristici con maggiore maturità e gli fornirà le basi per poter mettere a frutto le conoscenze acquisite anche in contesti in cui si utilizzano modelli di dati e sistemi di gestione diversi dai modelli e dai sistemi relazionali.

Frequenza

Per questo insegnamento la frequenza non è obbligatoria, e non saranno fatte distinzioni in sede di esame tra studenti frequentanti e non frequentanti. Tuttavia, gli studenti che non frequentano devono prevedere di dedicare del tempo individuale per le indispensabili esercitazioni pratiche con un proprio calcolatore, o nelle aule informatiche del Dipartimento. Tutto il materiale presentato al corso viene reso disponibile di volta in volta sulla pagina web del corso.

Programma dettagliato

Introduzione alle basi di dati e modello relazionale dove sono affrontati i seguenti argomenti: Introduzione al corso; Basi di Dati e Sistemi di Gestione di Basi di Dati; Modelli dei Dati e Linguaggi delle Basi di Dati; Modello Relazionale dei Dati.

Progettazione concettuale e logica dove sono affrontati i seguenti argomenti: progettazione concettuale di una base dati; progettazione logica di una base dati

Esercitazione: Progettazione di un modello E-R a partire da un caso di studio

Algebra relazionale dove sono affrontati i seguenti argomenti: l'Algebra Relazionale come Linguaggio per la scrittura di Interrogazioni.

Calcolo relazionale dove sono affrontati i seguenti argomenti: Calcolo Relazionale come linguaggio d'interrogazione.

Esercitazione: Progettazione concettuale e logica di una base dati a partire dal modello E-R derivante da un caso di studio, interrogazioni in algebra relazionale.

Linguaggio SQL ed intro alla progettazione dove sono affrontati i seguenti argomenti: introduzione al linguaggio SQL; caratteristiche evolute del SQL; SQL per le applicazioni; metodologie e modelli per la progettazione di una base dati.

Progettazione fisica e normalizzazione dove sono affrontati i seguenti argomenti: progettazione fisica di una base dati, normalizzazione.

Esercitazione: Progettazione fisica e normalizzazione di una base dati a partire dal progetto concettuale e logico derivante da un caso di studio.

Interrogazioni e transazioni dove sono affrontati i seguenti argomenti: organizzazione fisica e gestione delle interrogazioni; gestione delle transazioni.

Esercitazione sul linguaggio SQL: installazione del software DBMS, creazione e popolamento di un database, interrogazioni semplici ed annidate.

Sviluppo applicazioni dove sono affrontati i seguenti argomenti: metodi e librerie per lo sviluppo di applicazioni fondate su basi di dati.

Ottimizzazione delle prestazioni: Il processo di ottimizzazione delle query SQL, Tecniche di accesso ai dati dei DBMS relazionali, Tecniche di ottimizzazione, Interpretazione dei piani di esecuzione. Calcolo del costo di esecuzione per query SQL.

Basi di dati distribuite: Concetti di basi di dati distribuite, controllo della concorrenza e recovery delle base di dati distribuite. Architetture delle basi di dati distribuite.

Basi di dati NoSQL. Modelli dati non relazionali (Document-based, Graph-based, Column oriented). Architetture per DBMS non relazionali. Cluster, Sharding e persistenza. Modellazione dei dati in DBMS NoSQL. Esercitazioni di laboratorio.

Gestione dei Big Data. Trattamento di grandi quantità di dati. Map Reduce e Hadoop. Hadoop Distributed File System. Spark SQL.

Prerequisiti

Lo studente deve aver sostenuto gli esami di

1. Matematica discreta.
2. Programmazione dei calcolatori con laboratorio.

Modalità d'esame

La prova d'esame consiste:

1. Realizzazione di un progetto (40%)
2. Prova scritta (40%)
3. Prova orale sul programma del corso. (20%)

Si potrà accedere alla prova orale solo se la valutazione del progetto e dello scritto avranno riportato un esito positivo.

I voti verranno assegnati in base a una valutazione complessiva delle conoscenze, competenze, capacità di presentazione e discussione delle tematiche trattate. I range di voti corrispondono possono essere descritti come segue:

18-23: preparazione e capacità di analisi sufficienti ma relative ad un numero limitato di argomenti affrontati nel corso, utilizzo di un linguaggio complessivamente corretto;

24-27: preparazione tecnicamente adeguata ma con alcuni limiti rispetto agli argomenti trattati, capacità di analisi buone, anche se non particolarmente articolate, espresse in un linguaggio corretto;

28-30: ottima conoscenza di un ampio numero di temi affrontati nel corso, buone capacità di analisi e di critica, padronanza della terminologia specifica;

30L: conoscenza eccellente e molto approfondita ed esaustiva dei temi affrontati nel corso, capacità di analisi critica e di collegamento, padronanza della terminologia specifica.

Testi di riferimento

- Atzeni, Ceri, Fraternali, P araboschi, Torlone Basi di dati -Modelli e Linguaggi di interrogazione- ed. McGraw-Hill 4nd edition
- Elmasri, Navathe Sistemi di basi di dati, Fondamenti e complementi Pearson 7a edizione italiana

Materiale del corso on-line

Il materiale del corso è disponibile alla pagina del corso

Per informazioni o domande, è possibile scrivere per posta elettronica alla Prof.ssa Paola Vocca paola.vocca@uniroma2.it