

Prefazione

Questo testo presenta i concetti fondamentali sulle basi di dati, sui linguaggi di interrogazione e di gestione, e sulle tecniche e sui metodi di progettazione. Esso nasce da una lunga esperienza di insegnamento in corsi riguardanti le basi di dati, in ambito sia universitario sia industriale e applicativo, e pertanto si rivolge al pubblico degli studenti (in particolare di Ingegneria e di Scienze dell'informazione o Informatica) e a quello dei professionisti (utenti e progettisti di applicazioni).

Contenuti

Il libro si articola in tre parti:

Basi di dati relazionali, modello e linguaggi. Vengono presentate le caratteristiche fondamentali delle basi di dati che risultano di interesse per gli utenti e i programmatori. In particolare, si illustrano il modello relazionale e i relativi linguaggi, in modo preciso e concreto, con riferimento sia alle definizioni formali (del modello, dell'algebra e del calcolo) sia ai sistemi esistenti (con riferimento soprattutto al linguaggio SQL).

Progettazione di basi di dati. Viene illustrato ed esemplificato il processo di progettazione concettuale, logica e fisica delle basi di dati relazionali, che permette, partendo dai requisiti di utente, di arrivare a produrre strutture di basi di dati di buona qualità. Vengono poi discusse le tecniche principali per l'utilizzo delle basi di dati nelle applicazioni.

Tecnologia delle basi di dati. Vengono descritte le caratteristiche interne dei sistemi di basi di dati in rapporto all'architettura hardware e software del sistema informativo, in modo da comprenderne il funzionamento e sfruttarne appieno le potenzialità.

Ciascun capitolo è corredato di numerosi esempi ed esercizi, nonché di una nota bibliografica che indica le fonti per possibili approfondimenti, elencate poi globalmente alla fine del volume.

Utilizzo didattico

Nell'esperienza degli Autori, gli argomenti trattati in questo volume vengono svolti in modo completo in un tipico primo corso di basi di dati, corrispondente cioè a un modulo da 5-6 crediti (circa 30 ore di lezione e 20 di esercitazione) oppure ad un modulo più esteso, da 8-9 crediti; in particolare, l'attuale revisione tiene conto delle

esigenze espresse da molti docenti italiani che insegnano un primo corso di basi di dati con queste caratteristiche. A essi è opportuno associare un'ampia attività pratica, in particolare lo svolgimento di un progetto di un piccolo sistema informativo che includa una base di dati. Informazioni utili circa l'organizzazione di alcuni DBMS relazionali sono presenti nei siti in appoggio al presente volume, descritti al termine della prefazione.

Il presente testo ha una naturale continuazione nel testo "Basi di dati: architetture e linee di evoluzione" [6], che può essere utilizzato per un secondo modulo più avanzato oppure, insieme a questo, per un corso di 10-12 crediti. Tale testo, cui nel seguito faremo spesso riferimento come "il secondo volume", copre i seguenti argomenti.

Evoluzione dei modelli e dei linguaggi per basi di dati. Vengono illustrate le moderne varianti rispetto al modello e al linguaggio relazionale, focalizzandosi sulle basi di dati a oggetti, sulla gestione di dati XML, sulle basi di dati attive e sulla gestione di stream di dati.

Architetture evolute per basi di dati. Vengono illustrate le principali architetture dei sistemi informativi moderni, focalizzandosi sulla distribuzione dei dati, sulla integrazione con il World Wide Web e sui sistemi per l'analisi dei dati e il supporto alle decisioni.

Esperienze e ringraziamenti

L'organizzazione di questo testo e i suoi contenuti riflettono l'esperienza didattica degli Autori, che hanno tenuto per molti anni il corso universitario di Basi di dati e hanno svolto in altri contesti corsi sugli stessi temi. In particolare, Paolo Atzeni ha svolto in passato il corso di Basi di dati presso la facoltà di Ingegneria dell'Università di Roma "La Sapienza" e prima ancora presso l'Università di Toronto, tiene ora i corsi di Basi di dati II e di Sistemi informativi presso il Dipartimento di Ingegneria dell'Università Roma Tre, dove ha anche tenuto per molti anni il corso di Basi di dati I. Stefano Ceri tiene i corsi di Basi di dati 1, Basi di dati 2 e "Search Computing" presso la Scuola di Ingegneria dell'Informazione del Politecnico di Milano. Ha inoltre tenuto varie edizioni del corso "Principles of Distributed Databases" presso l'Università di Stanford. Piero Fraternali tiene i corsi di Dasi di dati e Web e Advanced Web Technologies presso la Scuola di Ingegneria dell'Informazione del Politecnico di Milano. Stefano Paraboschi tiene corsi di Basi di dati e Sistemi informativi presso il Dipartimento di Ingegneria dell'Università di Bergamo e la Scuola di Ingegneria dell'Informazione del Politecnico di Milano. Riccardo Torlone tiene corsi di Basi di dati presso il Dipartimento di Ingegneria dell'Università Roma Tre.

Alla concezione e alla revisione di questo testo hanno contribuito, direttamente o indirettamente, anche attraverso discussioni sui contenuti didattici dei corsi o suggerimenti di vario tipo, numerosi colleghi, collaboratori e lettori. Citiamo, fra gli altri, Maristella Agosti, Giorgio Ausiello, Elena Baralis, Giovanni Barone, Carlo Battini, Giampio Bracchi, Daniele Braga, Francesca Bugiotti, Luca Cabibbo, Alessandro Campi, Paolo De Nictolis, Giuseppe Di Battista, Angelo Foglietta, Maurizio Lenzerini, Gianni Mecca, Paolo Merialdo, Barbara Pernici, Silvio Salza, Pierangela Samarati,

Fabio Schreiber, Giuseppe Sindoni, Elena Tabet e Letizia Tanca. A ciascuno di essi, nonché a coloro che abbiamo dimenticato, va il nostro più sincero ringraziamento.

Nota di edizione

Questa nuova edizione esce a quasi vent'anni di distanza dalla pubblicazione del volume: "Basi di dati: concetti, linguaggi e architetture", che ha visto la sua seconda edizione e un'edizione internazionale nel 1999. Da allora, abbiamo suddiviso il materiale in due volumi, le cui prime edizioni sono uscite nel 2003 e 2004, e le seconde edizioni nel 2006 e 2007, e una terza edizione del primo volume nel 2009.

Il costante lavoro di aggiornamento e integrazione di queste opere tiene conto del continuo progresso dei linguaggi e della tecnologia per la gestione dei dati (che, per esempio, nell'ultimo decennio ha visto un'esplosione nelle applicazioni su Internet), e di commenti sull'uso del testo e di suggerimenti sui nuovi requisiti didattici espressi da numerosi colleghi, docenti di corsi di Basi di dati offerti dalle sedi accademiche sia di Ingegneria sia di Scienze dell'informazione o Informatica, raccolti da McGraw-Hill e analizzati con attenzione dagli autori.

Nell'evoluzione dalla prima alla terza edizione avevamo già operato numerose modifiche, aggiornando i capitoli relativi a SQL, tenendo conto dell'evoluzione dello standard SQL-3, collegando la progettazione concettuale alla progettazione UML, estendendo il capitolo sulla normalizzazione, e aggiungendo un capitolo sulla progettazione fisica che già conteneva alcuni cenni sulle strutture dati.

Rispetto alla precedente, questa quarta edizione presenta le seguenti novità.

- È stato profondamente rivisto il capitolo relativo allo *Sviluppo di applicazioni per basi di dati*, che ha esteso i contenuti relativi all'uso di SQL nelle applicazioni tenendo conto dell'evoluzione della tecnologia, ed in particolare della possibilità di integrare le basi di dati nel paradigma ad oggetti tramite mappatura (il cosiddetto *Object Relational Mapping*); in tal modo viene facilitato il compito di trasformare tuple relazionali in oggetti del linguaggio e di costruire interrogazioni all'interno di un programma. Il capitolo è collocato a valle della progettazione concettuale e logica, in modo da utilizzare i concetti del modello Entità-Relazioni; così facendo la parte sulla progettazione si arricchisce di un passo relativo al progetto delle applicazioni, a valle del progetto logico.
- È stato rivisto il capitolo relativo alle caratteristiche evolute di SQL, includendo un paragrafo sui trigger che comprende la descrizione sia degli standard sia di alcune sue versioni supportate da DB2 e Oracle.
- È stato esteso il paragrafo relativo alla progettazione fisica, che ora si avvale di una descrizione più estesa dell'organizzazione fisica dei dati.
- È stata aggiunta la parte di *Tecnologie delle basi di dati*, che descrive l'organizzazione fisica dei dati nella memoria secondaria, la gestione e ottimizzazione delle interrogazioni, e la gestione delle transazioni, in particolare il controllo di affidabilità e di concorrenza.
- Le appendici, relative ai sistemi *Postgres*, *Microsoft Access* e *DB2 Universal Database* sono state aggiornate e spostate dal testo e sono disponibili sul sito web come ebook.

Inoltre, tutto il materiale è stato rivisto e aggiornato, e al termine di ciascun capitolo sono stati aggiunti numerosi esercizi.

Materiale aggiuntivo

Materiale didattico di supporto al libro è disponibile ai siti:

<http://www.ateneonline.it/atzeni>
<http://www.dia.uniroma3.it/librobd>

In particolare, per gli studenti sono disponibili le soluzioni di tutti gli esercizi presenti nel testo, alcuni esercizi aggiuntivi, l'Appendice A dell'edizione del 1996 (relativa al modello reticolare e omessa nelle edizioni successive) e le Appendici A, B e C della terza edizione, relative ai sistemi *Microsoft Access*, *DB2 Universal Database* e *Postgres*.

Per i docenti che utilizzano il testo sono disponibili, oltre ai materiali presenti nell'area studenti, i lucidi che coprono in modo completo i dodici capitoli della presente edizione.

Per eventuali segnalazioni di errori e altri suggerimenti, gli Autori possono essere contattati attraverso i siti stessi.

Autori

Paolo Atzeni è professore ordinario di basi di dati all'Università Roma Tre, dove è attualmente Direttore del Dipartimento di Ingegneria. Si è laureato in ingegneria elettronica nel 1980 all'Università di Roma "La Sapienza" ed è stato ricercatore allo IASI-CNR di Roma e poi professore associato a all'Università di Napoli e professore ordinario all'Università di Roma "La Sapienza". Ha svolto attività didattica e scientifica presso la University of California, Los Angeles (UCLA), la University of Toronto e la Microsoft Research. Ha svolto ricerche su vari temi nel settore della basi di dati, fra cui la teoria relazionale, i modelli concettuali e gli strumenti di sviluppo, le basi di dati nel mondo Web, la gestione e la traduzione di schemi eterogenei. Ha fondato e dirige il gruppo di basi di dati di Roma Tre che comprende cinque docenti e numerosi assegnisti e dottorandi.

Il gruppo collabora con numerosi gruppi qualificati in Italia e all'estero e partecipa a progetti nazionali e internazionali. Paolo Atzeni è vicepresidente del VLDB Endowment ed ex-Presidente e membro del Consiglio dell'Associazione EDBT.

Stefano Ceri è professore ordinario di basi di dati al Politecnico di Milano, dove è attualmente Direttore dell'Alta Scuola Politecnica. Si è laureato in ingegneria elettronica nel 1978 al Politecnico di Milano, è stato ricercatore al Politecnico fino al 1985, poi professore ordinario all'Università di Modena, e infine professore Ordinario al Politecnico di Milano dal 1990. È stato professore visitatore del Dipartimento di Computer Science alla Stanford University, ove dal 1983 al 1992 ha insegnato il corso "Distributed Databases: Principles and Systems". Ha svolto ricerche in vari aspetti

delle basi di dati e del Web, inclusi database distribuiti, attivi, deduttivi e a oggetti e metodi e modelli per la progettazione di applicazioni sul Web. Ha ricevuto un IDEAS Advanced Grant ERC su “Search Computing” (2008-2013) ed è coordinatore nazionale del progetto PRIN su “Data-Centric Genomic Computing” (2013-2016). È socio fondatore della spin-off WebRatio.

Piero Fraternali è professore ordinario di Tecnologie Web presso il Dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano. Si occupa delle tecniche di specifica e progettazione del software, dell’integrazione tra Sistemi informativi e World Wide Web, e dell’uso delle reti sociali e dei giochi nella computazione. Dal 1996 si occupa di metodologie strumenti per lo sviluppo assistito da calcolatore di applicazioni per il Web. È coautore del linguaggio di specifica concettuale di applicazioni Web WebML (WWW Modelling Language, <http://www.webml.org>), brevettato internazionalmente. Nel 2001, assieme al Politecnico di Milano, Piero Fraternali ha fondato la società WebRatio, che sviluppa e commercializza un nuovo software per la costruzione di applicazioni Web di grandi dimensioni (<http://www.webratio.com>). Nel dicembre 2002 pubblica presso l’editore Morgan Kaufman, insieme ad altri ricercatori del Politecnico di Milano, il testo “Designing Data-Intensive Web Applications”, che descrive una metodologia di sviluppo per applicazioni Web basata sulla specifica concettuale dei requisiti. Ha diretto diversi progetti scientifici finanziati dalla Comunità Europea.

Stefano Paraboschi è professore ordinario dal 2002 all’Università degli Studi di Bergamo, dove presiede il corso di laurea in Ingegneria Informatica. Si è laureato in Ingegneria elettronica nel 1990 al Politecnico di Milano, dove ha conseguito il titolo di dottore di ricerca e ha poi ricoperto i ruoli di ricercatore e professore associato. Ha svolto periodi di ricerca presso la Stanford University, l’IBM Almaden Research Center e la George Mason University. Nella sua ricerca ha esplorato vari temi nell’area delle basi di dati, tra cui il progetto e l’uso di regole attive, il disegno di sistemi multidimensionali, l’integrazione tra basi di dati e Web e la gestione di dati XML. Di recente i suoi interessi si sono spostati sullo studio di problemi all’intersezione tra la sicurezza informatica e la gestione dei dati. Di recente ha svolto il ruolo di coordinatore nazionale per un progetto PRIN su basi di dati crittografate e un altro progetto PRIN su Privacy e protezione di dati personali. È responsabile per la partecipazione dell’Università di Bergamo a due progetti europei di tipo IP, finanziati dal 7th Framework Programme.

Riccardo Torlone è professore ordinario di basi di dati all’Università Roma Tre, dove è attualmente il coordinatore dei corsi di studio di Ingegneria informatica. Si è laureato in ingegneria elettronica all’Università di Roma “La Sapienza”, è stato ricercatore allo IASI-CNR di Roma e ha svolto attività di ricerca presso la University of California, Los Angeles (UCLA). La sua attività di ricerca ha riguardato vari argomenti nel settore delle basi di dati, tra cui: la teoria delle basi di dati, le basi di dati attive e deduttive, i data warehouse, i sistemi informativi su Web, l’integrazione e la trasformazione di dati, i sistemi adattativi e personalizzati. Fa parte del gruppo di basi di dati di Roma Tre e collabora con numerosi gruppi qualificati in Italia e all’estero.