

Questa nuova edizione italiana di *Fondamenti di Sistemi di Basi di Dati* dell'opera *Fundamentals of Database Systems* di Ramez Elmasri e Shamkant Navathe – da anni best seller mondiale e testo di riferimento nel settore della basi di dati – è un'edizione rivista e riorganizzata del precedente volume, che si adegua alla settima edizione dell'opera americana.

Questo volume, rispetto alla precedente edizione, si adatta bene sia ai contenuti tipici di un corso triennale, offrendo una trattazione ampia e dettagliata delle nozioni e dei principi fondamentali per la progettazione, l'implementazione e la gestione delle basi di dati, sia agli approfondimenti necessari per la gestione di un corso magistrale. Il testo fornisce infatti un'aggiornata panoramica sullo stato dell'arte delle più recenti tecnologie, con riferimento ai sistemi di gestione delle basi di dati relazionali, ma anche con particolare attenzione ai nuovi sistemi NOSQL che si stanno sempre di più affermando per poter coprire l'esigenza della gestione di grandi volume di dati (Big Data).

In tutti gli ambiti si è voluto fornire una serie di esercizi con alcune soluzioni guidate per quelli particolarmente significativi.

Si è scelto inoltre, in questa edizione italiana rispetto a quella americana, di non considerare i Sistemi di Basi di Dati Orientati agli Oggetti (OODBMS) nelle linee evolutive delle basi di dati, poiché si ritiene che allo stato attuale abbiano uno scarso riscontro pratico e non abbiano incontrato il favore delle aziende o del gran numero di utenti che si occupano di basi di dati. Sono stati comunque mantenuti, ove comprensibili, i riferimenti a questa realtà e alle sue tematiche.

Sono stati aggiunti invece, verso la fine del testo, contenuti evoluti quali i concetti e i modelli di datawarehousing, di data mining, e appunto i modelli non relazionali (NOSQL) nelle loro varie forme, nonché un'approfondita analisi delle problematiche e dei protocolli alla base della gestione transazionale, della concorrenza dei dati e del loro recupero. Il tutto fornisce una visione completa ed esaustiva per un approccio scientifico approfondito a tali problematiche. Infine una nuova parte dedicata all'ottimizzazione permette di coprire un aspetto spesso sottovalutato e che riguarda l'efficienza delle basi di dati.

Il testo, suddiviso in parti e corredato delle sue parti online, ben si adatta quindi a chi vuole approfondire particolari tematiche, a chi intende sviluppare tesi specifiche nella materia, o a chi semplicemente vuole avere una visione completa della gestione dei dati in tutti i suoi aspetti e sotto tutti i punti di vista.

Il testo è suddiviso in 11 Parti più 2 Appendici con i seguenti contenuti:

Parte 1 (Capitoli 1 e 2) : descrive i concetti introduttivi di base necessari per una buona comprensione dei modelli, dei sistemi e dei linguaggi delle basi di dati. Nei capitoli 1 e 2 si introducono le basi di dati, i concetti di DBMS, la terminologia e l'architettura, nonché una discussione sull'evoluzione delle

tecnologie delle basi di dati nel tempo e una breve storia dei modelli di dati, comprensivi dei nuovi sistemi NOSQL.

- Parte 2 (Capitoli 3 e 4) : comprende la presentazione della modellazione Entità-Relazione e della progettazione delle basi di dati; è importante sottolineare che gli insegnanti possono illustrare i capitoli del modello relazionale (dal Capitolo 5 al Capitolo 8) prima dei capitoli 3 e 4 se questo è il loro ordine preferito di presentazione del materiale del corso. Nel Capitolo 3 sono presentati i concetti del modello Entità-Relazione (ER) e gli schemi ER che sono usati per illustrare la progettazione concettuale delle basi di dati. Il Capitolo 4 illustra come il modello base ER può essere esteso per incorporare concetti di progettazione aggiuntiva portando al modello di dati ER esteso (EER) e agli schemi EER. Viene mostrata infine una panoramica delle basi di dati spaziali.
- Parte 3 (Capitoli 5, 6, 7, 8 e 9) : comprende una presentazione dettagliata delle basi di dati relazionali e dell' SQL con alcuni nuovi costrutti aggiunti rispetto all'edizione precedente. Il Capitolo 5 descrive il modello relazionale di base, con i suoi vincoli di integrità e le operazioni di aggiornamento. Il Capitolo 6 descrive le operazioni formali dell' algebra relazionale e introduce il calcolo relazionale. Se l'insegnante lo desidera questo capitolo può essere posposto ai capitoli 7 e 8 sull'SQL. Il Capitolo 7 descrive alcuni degli elementi di base dell'SQL standard per basi di dati relazionali, incluse la definizione dei dati, le operazioni di modifica dei dati e semplici interrogazioni SQL. Il Capitolo 8 presenta interrogazioni SQL più complesse, ma anche i concetti di trigger, asserzioni, viste e modifiche dello schema. Il Capitolo 9 infine comprende il mapping dai modelli ER e EER a quelli relazionali, con l'utilizzo di alcuni algoritmi per la progettazione.
- Parte 4 (Capitolo 10) : il Capitolo 10 (insieme al successivo Capitolo 13) riguarda la teoria della progettazione relazionale e la normalizzazione. Gli aspetti più formali degli algoritmi di progettazione sono stati posti online nella quinta parte. Il Capitolo 10 definisce le dipendenze funzionali e le forme normali basate su queste e sviluppa anche un approccio intuitivo alla normalizzazione e definisce le dipendenze multivalore e di join.
- Parte 5 (Capitoli 11, 12 e 13) : riguarda capitoli sulle tecniche generali di programmazione sulle basi di dati. Questi capitoli possono essere assegnati come materiale in lettura ed incrementati con il materiale del particolare linguaggio usato nel corso di Programmazione. Il Capitolo 11 introduce la programmazione alle basi di dati sul Web, utilizzando negli esempi il linguaggio di scripting PHP. Il Capitolo 12 descrive, oltre ad una parte sulle basi di dati attive (trigger), argomenti tradizionali della programmazione SQL, come embedded SQL, SQL dinamico, ODBC, SQLJ, JDBC e SQL/CLI, includendo nuovo materiale sulla tecnologia Java per la

programmazione delle basi di dati su Web. Il Capitolo 13 (online) torna invece ad occuparsi della normalizzazione e dei vari algoritmi, teorie e formalismi sviluppati per la progettazione delle basi di dati relazionali tramite la normalizzazione, anche con algoritmi di decomposizione e sintesi relazionale.

Parte 6 (Capitoli 14 e 15) : contiene capitoli sull'organizzazione dei file su disco (Capitolo 14) e sull'indicizzazione dei file di basi di dati (Capitolo 15). Il Capitolo 14 (online) descrive i metodi principali di organizzazione dei file su disco, come i file ordinati (sorted), non ordinati (heap) e hashed, con trattazione delle tecniche di hashing. Il Capitolo 14 è stato aggiornato per comprendere materiale sulle strategie di gestione dei buffer nei DBMS e offre anche una panoramica dei nuovi device di memorizzazione standard. Il Capitolo 15 descrive le tecniche di indicizzazione per i file, incluse strutture dati quali Btree, B⁺tree e file griglia, ed è stato aggiornato con nuovi esempi ed un'approfondita discussione sugli indici, su come sceglierli e quando crearli.

Parte 7 (Capitoli 16 e 17) : comprende i capitoli sugli algoritmi di elaborazione delle interrogazioni (Capitolo 16) e sulle tecniche di ottimizzazione (Capitolo 17). Il Capitolo 16 (online) descrive algoritmi per la ricerca di record su file di disco, per collegare record tra due file o tabelle, e descrive nuovi argomenti, quali le operazioni di semi-join e anti-join, nonché tecniche per stimare la selettività. Il Capitolo 17 (online) comprende tecniche per l'ottimizzazione delle interrogazioni che utilizzano la stima sui costi e regole euristiche; comprende inoltre nuovo materiale sull'ottimizzazione delle sotto-interrogazioni nidificate, l'uso degli istogrammi, ottimizzazione fisica e metodi di ordinamento dei join e ottimizzazione delle interrogazioni tipiche nei data warehouse.

Parte 8 (Capitoli 18, 19 e 20) : descrive i concetti di elaborazione delle transazioni, di controllo della concorrenza e del recupero delle basi di dati dai guasti, aggiornati con le tecniche più nuove dei DBMS commerciali ed Open Source. Il Capitolo 18 introduce i metodi necessari per i sistemi di elaborazione delle transazioni e definisce i concetti di recuperabilità e serializzabilità di schedule; ha sezioni nuove sulle politiche di sostituzione nei buffer e sull'isolamento di snapshot. Il Capitolo 19 fornisce una panoramica dei vari tipi di protocolli per il controllo della concorrenza, focalizzandosi sul locking a due fasi, sul timestamp, nonché su metodi ottimistici, sul locking a granularità multipla e sui nuovi metodi di isolamento degli snapshot. Infine il Capitolo 20 si focalizza sui protocolli di recupero delle basi di dati e fornisce una panoramica delle tecniche usate a questo scopo.

Parte 9 (Capitoli 21, 22 e 23) : comprende il Capitolo sulle basi di dati distribuite (Capitolo 21) più due nuovi capitoli sui sistemi di memorizzazione NOSQL per big data (Capitolo 22) e tecnologie di big data basate su Hadoop e MapReduce (Capitolo 23). Il Capitolo 21 introduce i concetti di basi di dati distribuite, includendo la disponibilità e la scalabilità, la replicazione e la frammentazione dei dati, e molto altro. Nel Capitolo 22 i sistemi NOSQL sono classificati in quattro categorie generali, con un sistema di esempio per ciascuna categoria, e con modelli di dati, operazioni, strategie di replicazione/distribuzione/scalabilità per ciascun tipo di sistema NOSQL, confrontati fra loro. Nel Capitolo 23 è introdotto il modello di programmazione MapReduce per l'elaborazione distribuita dei big data, e viene data una presentazione del sistema Hadoop e dell'HDFS (File system distribuito Hadoop), nonché delle interfacce di alto livello Pig e Hive e dell'architettura Yarn.

Parte 10 (Capitoli 24 e 25) : è intitolato modelli avanzati di sistemi di basi di dati e loro applicazioni. Il Capitolo 24 è una descrizione dei data warehouse che comprende argomenti come modelli e operazioni di data warehouse, e il procedimento di costruzione di un data warehouse. Il Capitolo 25 è un'introduzione al data mining che comprende la descrizione di vari metodi di data mining : il mining delle regole associate, il clustering, la classificazione e la scoperta di pattern sequenziali.

Parte 11 (Capitolo 26) : il Capitolo 26 (online) descrive la sicurezza delle basi di dati e comprende una discussione sui comandi SQL per il controllo di accesso discrezionale (Grant, Revoke) e per ruoli. Il capitolo tratta anche delle minacce, come gli attacchi da iniezioni SQL e di altre tecniche e metodi relativi alla sicurezza e alla privacy dei dati.

Appendice A : infine l'Appendice A propone un certo numero di notazioni diagrammatiche alternative per visualizzare uno schema concettuale ER o EER, che possono sostituire la notazione usata, se l'insegnante le preferisce.

Appendice B : l'Appendice B fornisce alcuni importanti parametri fisici per i dischi.

Un particolare ringraziamento al dott. Alessandro Valli per l'ottima traduzione delle nuove parti, a tutto il personale di Pearson Italia coinvolto nella realizzazione di questo libro e soprattutto alla Dott.ssa Chiara Tartara per la sua preziosa collaborazione e per la sua cortese presenza, disponibilità e professionalità. Un ringraziamento infine a tutti gli studenti dei miei corsi di questi anni che, con le loro osservazioni e i loro errori, mi hanno permesso di potermi soffermare e di poter valutare meglio ogni aspetto di questa disciplina.

*Dott.ssa Loredana Vigliano – Dipartimento di Ingegneria dell'Impresa –
Università degli Studi di Roma "Tor Vergata"*