

Principles Of Database Query Processing For Advanced Applications The Morgan Kaufmann Series In Data Management Systems

This text covers basic database concepts to provide a conceptual understanding of data and databases necessary for database design and development.

Principles of Database Query Processing for Advanced Applications Morgan Kaufmann

Adaptive Query Processing surveys the fundamental issues, techniques, costs, and benefits of adaptive query processing. It begins with a broad overview of the field, identifying the dimensions of adaptive techniques. It then looks at the spectrum of approaches available to adapt query execution at runtime - primarily in a non-streaming context. The emphasis is on simplifying and abstracting the key concepts of each technique, rather than reproducing the full details available in the papers. The authors identify the strengths and limitations of the different techniques, demonstrate when they are most useful, and suggest possible avenues of future research. **Adaptive Query Processing** serves as a valuable reference for students of databases, providing a thorough survey of the area. Database researchers will benefit from a more complete point of view, including a number of approaches which they may not have focused on within the scope of their own research.

With the growth of Java and the rise of database-powered Web applications, the need to use Java with SQL is clear. Until now, authoritative coverage of the techniques available to meet these challenges and reap their benefits-both programming and career benefits-didn't exist. **Understanding SQL and Java Together** examines all the standards for combining SQL and Java. It shows you exactly how to use their features to write efficient and effective code supporting Java access to SQL data in a variety of ways. You'll gain a thorough understanding of the relationship between SQL and Java, which will allow you to write static and dynamic SQL programs in Java, merge Java code with SQL databases and SQL code, and use other data management techniques wherever appropriate. * Covers all the technologies for using SQL and Java together, including JDBC, Java Blend, and SQLJ Parts 0, 1, and 2 * Explains how to embed SQL code in Java and take advantage of Java's ability to compile that code for a specific DBMS * Explains how to store and invoke Java routines in an SQL database-and how to store Java objects in an SQL database for seamless interchange among application layers * Covers dynamic SQL access techniques using JDBC and advantageous ways to combine static and dynamic SQL * Comes with a CD-ROM containing Oracle's JDeveloper , Sybase's Adaptive Server Anywhere, Informix's Cloudscape, the complete database schema, and the complete text of most of the examples

Principles of Distributed Database Systems

Advanced SQL:1999

Semantics in Databases

Designing Data-Intensive Web Applications

A Complete Guide to DB2 Universal Database

4th International Conference on Extending Database Technology, Cambridge, United Kingdom, March 28 - 31, 1994. Proceedings

*Mining the Web: Discovering Knowledge from Hypertext Data is the first book devoted entirely to techniques for producing knowledge from the vast body of unstructured Web data. Building on an initial survey of infrastructural issues—including Web crawling and indexing—Chakrabarti examines low-level machine learning techniques as they relate specifically to the challenges of Web mining. He then devotes the final part of the book to applications that unite infrastructure and analysis to bring machine learning to bear on systematically acquired and stored data. Here the focus is on results: the strengths and weaknesses of these applications, along with their potential as foundations for further progress. From Chakrabarti's work—painstaking, critical, and forward-looking—readers will gain the theoretical and practical understanding they need to contribute to the Web mining effort. * A comprehensive, critical exploration of statistics-based attempts to make sense of Web Mining. * Details the special challenges associated with analyzing unstructured and semi-structured data. * Looks at how classical Information Retrieval techniques have been modified for use with Web data. * Focuses on today's dominant learning methods: clustering and classification, hyperlink analysis, and supervised and semi-supervised learning. * Analyzes current applications for resource discovery and social network analysis. * An excellent way to introduce students to especially vital applications of data mining and machine learning technology.*

Linked Data Management presents techniques for querying and managing Linked Data that is available on today's Web. The book shows how the abundance of Linked Data can serve as fertile ground for research and commercial applications. The text focuses on aspects of managing large-scale collections of Linked Data. It offers a detailed introduction to Linked Data and related standards, including the main principles distinguishing Linked Data from standard database technology. Chapters also describe how to generate links between datasets and explain the overall architecture of data integration systems based on Linked Data. A large part of the text is devoted to query processing in different setups. After presenting methods to publish relational data as Linked Data and efficient centralized processing, the book explores lookup-based, distributed, and parallel solutions. It then addresses advanced topics, such as reasoning, and discusses work related to read-write Linked Data for system interoperation. Despite the publication of many papers since Tim Berners-Lee developed the Linked Data principles in 2006, the field lacks a comprehensive, unified overview of the state of the art. Suitable for both researchers and practitioners, this book provides a thorough, consolidated account of the new data publishing and data

integration paradigm. While the book covers query processing extensively, the Linked Data abstraction furnishes more than a mechanism for collecting, integrating, and querying data from the open Web—the Linked Data technology stack also allows for controlled, sophisticated applications deployed in an enterprise environment.

Introductory, theory-practice balanced text teaching the fundamentals of databases to advanced undergraduates or graduate students in information systems or computer science.

An Overview of Multidatabase Systems: Past and Present / Athman Bouguettaya, Boualem Benatallah, Ahmed Elmagarmid / - Local Autonomy and Its Effects on Multidatabase Systems / Ahmed Elmagarmid, Weimin Du, Rafi Ahmed / - Semantic Similarities Between Objects in Multiple Databases / Vipul Kashyap, Amit Sheth / - Resolution of Representational Diversity in Multidatabase Systems / Joachim Hammer, Dennis McLeod / - Schema Integration: Past, Present, and Future / Sudha Ram, V. Ramesh / - Schema and Language Translation / Bogdan Czejdo, Le Gruenwald / - Multidatabase Languages / Paolo Missier, Marek Rusinkiewicz, W. Jin / - Interdependent Database Systems / George Karabatis, Marek Rusinkiewicz, Amit Sheth / - Correctness Criteria and Concurrency Control / Panos K. Chrysanthis, Krithi Ramamritham / - Transaction Management in Multidatabase Systems: Current Technologies and Formalisms / Ken Barker, Ahmed Elmagarmid / - Transaction-Based Recovery / Jari Veijalainen. ...

Practical Machine Learning Tools and Techniques with Java Implementations

Query Processing in Database Systems

A Guide to SQLJ, JDBC, and Related Technologies

Concepts in Practice

The Interpolation-based Approach to Query Reformulation

Encyclopedia of Database Systems

SQL: 1999 is the best way to make the leap from SQL-92 to SQL:1999, but it is much more than just a simple bridge between the two. The latest from celebrated SQL experts Jim Melton and AI is a comprehensive, eminently practical account of SQL's latest incarnation and a potent distillation of the details required to put it to work. Written to accommodate both novice and experienced SQL:1999 focuses on the language's capabilities, from the basic to the advanced, and the ways that real applications take advantage of them. Throughout, the authors illustrate features and tech often entertaining references to their own custom database. Gives authoritative coverage from an expert team that includes the editor of the SQL-92 and SQL:1999 standards. Provides a general that helps you understand its constituent parts, history, and place in the realm of computer languages. Explains SQL:1999's more sophisticated features, including advanced value expressions, pre SQL query expressions, and support for active databases. Explores key issues for programmers linking applications to SQL databases. Provides guidance on troubleshooting, internationalization, and anticipated in the next version of SQL. Contains appendices devoted to database design, a complete SQL:1999 example, the standardization process, and more.

Applications that require a high degree of distribution and loosely-coupled connectivity are ubiquitous in various domains, including scientific databases, bioinformatics, and multimedia retrieval. In applications, data is typically voluminous and multidimensional, and support for advanced query operators is required for effective querying and efficient processing. To address this challenge, we P2P architecture and propose novel indexing and query processing algorithms. We present a scalable framework that relies on data summaries that are distributed and maintained as multidimens indices. Different types of data summaries enable efficient processing of a variety of advanced query operators.

This book presents a coherent survey on exciting developments in database semantics. The origins of the volume date back to a workshop held in Prague, Czech Republic, in 1995. The nine revised surveys presented were carefully reviewed for inclusion in the book. They address more traditional aspects like dealing with integrity constraints and conceptual modeling as well as new areas of orientation, incomplete information, database transformations and other issues are investigated by applying formal semantics, e.g. the evolving algebra semantics.

This book is an anthology of the results of research and development in database query processing during the past decade. The relational model of data provided tremendous impetus for research processing. Since a relational query does not specify access paths to the stored data, the database management system (DBMS) must provide an intelligent query-processing subsystem which w of potentially efficient strategies for processing the query and select the one that optimizes a given performance measure. The degree of sophistication of this subsystem, often called the optimi performance of the DBMS. Research into query processing thus started has taken off in several directions during the past decade. The emergence of research into distributed databases has enor the tasks of the optimizer. In a distributed environment, the database may be partitioned into horizontal or vertical fragments of relations. Replicas of the fragments may be stored in different si even migrate to other sites. The measure of performance of a query in a distributed system must include the communication cost between sites. To minimize communication costs for-queries inv relations across multiple sites, optimizers may also have to consider semi-join techniques.

Linked Data Management

Joe Celko's Thinking in Sets: Auxiliary, Temporal, and Virtual Tables in SQL

Joe Celko's SQL for Smarties

Understanding Relational Language Components

Peer-to-Peer Query Processing over Multidimensional Data

The fourth international conference on Extending Data Base Technology was held in Cambridge, UK, in March 1994. The biannual EDBT has established itself as the premier European database conference. It provides an international forum for the presentation of new extensions to database technology through research, development, and application. This volume contains the scientific papers of the conference. Following invited papers by C.M. Stone and A. Herbert, it contains 31 papers grouped into sections on object views, intelligent user interface, distributed information servers, transaction management, information systems design and evolution, semantics of extended data models, accessing new media, join algorithms, query optimization, and multimedia databases.

This third edition of a classic textbook can be used to teach at the senior undergraduate and graduate levels. The material concentrates on fundamental theories as well as techniques and algorithms. The advent of the Internet and the World Wide Web, and, more recently, the emergence of cloud computing and streaming data applications, has forced

a renewal of interest in distributed and parallel data management, while, at the same time, requiring a rethinking of some of the traditional techniques. This book covers the breadth and depth of this re-emerging field. The coverage consists of two parts. The first part discusses the fundamental principles of distributed data management and includes distribution design, data integration, distributed query processing and optimization, distributed transaction management, and replication. The second part focuses on more advanced topics and includes discussion of parallel database systems, distributed object management, peer-to-peer data management, web data management, data stream systems, and cloud computing. New in this Edition: • New chapters, covering database replication, database integration, multidatabase query processing, peer-to-peer data management, and web data management. • Coverage of emerging topics such as data streams and cloud computing • Extensive revisions and updates based on years of class testing and feedback Ancillary teaching materials are available.

An industry consultant shares his most useful tips and tricks for advanced SQL programming to help the working programmer gain performance and work around system deficiencies.

The latest edition of a popular text and reference on database research, with substantial new material and revision; covers classical literature and recent hot topics. Lessons from database research have been applied in academic fields ranging from bioinformatics to next-generation Internet architecture and in industrial uses including Web-based e-commerce and search engines. The core ideas in the field have become increasingly influential. This text provides both students and professionals with a grounding in database research and a technical context for understanding recent innovations in the field. The readings included treat the most important issues in the database area--the basic material for any DBMS professional. This fourth edition has been substantially updated and revised, with 21 of the 48 papers new to the edition, four of them published for the first time. Many of the sections have been newly organized, and each section includes a new or substantially revised introduction that discusses the context, motivation, and controversies in a particular area, placing it in the broader perspective of database research. Two introductory articles, never before published, provide an organized, current introduction to basic knowledge of the field; one discusses the history of data models and query languages and the other offers an architectural overview of a database system. The remaining articles range from the classical literature on database research to treatments of current hot topics, including a paper on search engine architecture and a paper on application servers, both written expressly for this edition. The result is a collection of papers that are seminal and also accessible to a reader who has a basic familiarity with database systems.

Advanced SQL Programming Second Edition

SQL: 1999

Principles of Database Query Processing for Advanced Applications

Database and Expert Systems Applications

Mining the Web

Readings in Database Systems

This text surveys research from the fields of data mining and information visualisation and presents a case for techniques by which information visualisation can be used to uncover real knowledge hidden away in large databases.

First uniform treatment of moving objects databases, the technology that supports GPS and RFID data analysis.

The topic of using views to answer queries has been popular for a few decades now, as it cuts across domains such as query optimization, information integration, data warehousing, website design and, recently, database-as-a-service and data placement in cloud systems. This book assembles foundational work on answering queries using views in a self-contained manner, with an effort to choose material that constitutes the backbone of the research. It presents efficient algorithms and covers the following problems: query containment; rewriting queries using views in various logical languages; equivalent rewritings and maximally contained rewritings; and computing certain answers in the data-integration and data-exchange settings. Query languages that are considered are fragments of SQL, in particular select-project-join queries, also called conjunctive queries (with or without arithmetic comparisons or negation), and aggregate SQL queries. This second edition includes two new chapters that refer to tree-like data and respective query languages. Chapter 8 presents the data model for XML documents and the XPath query language, and Chapter 9 provides a theoretical presentation of tree-like data model and query language where the tuples of a relation share a tree-structured schema for that relation and the query language is a dialect of SQL with evaluation techniques appropriately modified to fit the richer schema.

This book constitutes the refereed proceedings of the 20th International Conference on Database and Expert Systems Applications, DEXA 2009, held in Linz, Austria, in August/September 2009. The 35 revised full papers and 35 short papers presented were carefully reviewed and selected from 202 submissions. The papers are organized in topical sections on XML and databases; Web, semantics and ontologies; temporal, spatial, and high dimensional databases; database and information system architecture, performance and security; query processing and optimisation; data and information integration and quality; data and information streams; data mining algorithms; data and information modelling; information retrieval and database systems; and database and information system architecture and performance.

Understanding Object-Relational and Other Advanced Features

Database Modeling and Design

Generating Plans from Proofs

With Application to GIS

Answering Queries Using Views

Second Edition

Data Model Patterns: A Metadata Map not only presents a conceptual model of a metadata repository but also demonstrates a true enterprise

data model of the information technology industry itself. It provides a step-by-step description of the model and is organized so that different readers can benefit from different parts. It offers a view of the world being addressed by all the techniques, methods, and tools of the information processing industry (for example, object-oriented design, CASE, business process re-engineering, etc.) and presents several concepts that need to be addressed by such tools. This book is pertinent, with companies and government agencies realizing that the data they use represent a significant corporate resource recognize the need to integrate data that has traditionally only been available from disparate sources. An important component of this integration is management of the "metadata" that describe, catalogue, and provide access to the various forms of underlying business data. The "metadata repository" is essential to keep track of the various physical components of these systems and their semantics. The book is ideal for data management professionals, data modeling and design professionals, and data warehouse and database repository designers. A comprehensive work based on the Zachman Framework for information architecture—encompassing the Business Owner's, Architect's, and Designer's views, for all columns (data, activities, locations, people, timing, and motivation) Provides a step-by-step description of model and is organized so that different readers can benefit from different parts Provides a view of the world being addressed by all the techniques, methods and tools of the information processing industry (for example, object-oriented design, CASE, business process re-engineering, etc.) Presents many concepts that are not currently being addressed by such tools – and should be

The latest techniques and principles of parallel and grid database processing The growth in grid databases, coupled with the utility of parallel query processing, presents an important opportunity to understand and utilize high-performance parallel database processing within a major database management system (DBMS). This important new book provides readers with a fundamental understanding of parallelism in data-intensive applications, and demonstrates how to develop faster capabilities to support them. It presents a balanced treatment of the theoretical and practical aspects of high-performance databases to demonstrate how parallel query is executed in a DBMS, including concepts, algorithms, analytical models, and grid transactions. High-Performance Parallel Database Processing and Grid Databases serves as a valuable resource for researchers working in parallel databases and for practitioners interested in building a high-performance database. It is also a much-needed, self-contained textbook for database courses at the advanced undergraduate and graduate levels.

The topic of using views to answer queries has been popular for a few decades now, as it cuts across domains such as query optimization, information integration, data warehousing, website design, and, recently, database-as-a-service and data placement in cloud systems. This book assembles foundational work on answering queries using views in a self-contained manner, with an effort to choose material that constitutes the backbone of the research. It presents efficient algorithms and covers the following problems: query containment; rewriting queries using views in various logical languages; equivalent rewritings and maximally contained rewritings; and computing certain answers in the data-integration and data-exchange settings. Query languages that are considered are fragments of SQL, in particular, select-project-join queries, also called conjunctive queries (with or without arithmetic comparisons or negation), and aggregate SQL queries.

This book describes the theory, algorithms, and practical implementation techniques behind transaction processing in information technology systems.

New Concepts for Parallel Object-Relational Query Processing

Moving Objects Databases

Probabilistic Databases

Principles of Database Management

Joe Celko's Data and Databases

Data Management and Query Processing in Semantic Web Databases

This is a guide designed to familiarize users with the DB2 standard while helping to optimize their use of the technology.

Query reformulation refers to a process of translating a source query—a request for information in some high-level logic-based language—into a target plan that abides by certain interface restrictions. Many practical problems in data management can be seen as instances of the reformulation problem. For example, the problem of translating an SQL query written over a set of base tables into another query written over a set of views; the problem of implementing a query via translating to a program calling a set of database APIs; the problem of implementing a query using a collection of web services. In this book we approach query reformulation in a very general setting that encompasses all the problems above, by relating it to a line of research within mathematical logic. For many decades logicians have looked at the problem of converting "implicit definitions" into "explicit definitions," using an approach known as interpolation. We will review the theory of interpolation, and explain its close connection with query reformulation. We will give a detailed look at how the interpolation-based approach is used to generate translations between logic-based queries over different vocabularies, and also how it can be used to go from logic-based queries to programs.

The Semantic Web, which is intended to establish a machine-understandable Web, is currently changing from being an emerging trend to a technology used in complex real-world applications. A number of standards and techniques have been developed by the World Wide Web Consortium (W3C), e.g., the Resource Description Framework (RDF), which provides a general method for conceptual

descriptions for Web resources, and SPARQL, an RDF querying language. Recent examples of large RDF data with billions of facts include the UniProt comprehensive catalog of protein sequence, function and annotation data, the RDF data extracted from Wikipedia, and Princeton University's WordNet. Clearly, querying performance has become a key issue for Semantic Web applications. In his book, Groppe details various aspects of high-performance Semantic Web data management and query processing. His presentation fills the gap between Semantic Web and database books, which either fail to take into account the performance issues of large-scale data management or fail to exploit the special properties of Semantic Web data models and queries. After a general introduction to the relevant Semantic Web standards, he presents specialized indexing and sorting algorithms, adapted approaches for logical and physical query optimization, optimization possibilities when using the parallel database technologies of today's multicore processors, and visual and embedded query languages. Groppe primarily targets researchers, students, and developers of large-scale Semantic Web applications. On the complementary book webpage readers will find additional material, such as an online demonstration of a query engine, and exercises, and their solutions, that challenge their comprehension of the topics presented.

This guide documents SQL: 1999Us advanced features in the same practical, "programmercentric" way that the first volume documented the language's basic features. This is no mere representation of the standard, but rather authoritative guidance on making an application conform to it, both formally and effectively.

Adaptive Query Processing

Understanding SQL and Java Together

20th International Conference, DEXA 2009, Linz, Austria, August 31 - September 4, 2009, Proceedings

High-Performance Parallel Database Processing and Grid Databases

High Performance Computing and Communications

Management of Heterogeneous and Autonomous Database Systems

The most prominent Web applications in use today are data-intensive. Scores of database management systems across the Internet access and maintain large amounts of structured data for e-commerce, on-line trading, banking, digital libraries, and other high-volume sites. Developing and maintaining these data-intensive applications is an especially complex, multi-disciplinary activity, requiring all the tools and techniques that software engineering can provide. This book represents a breakthrough for Web application developers. Using hundreds of illustrations and an elegant intuitive modeling language, the authors—all internationally-known database researchers—present a methodology that fully exploits the conceptual modeling approach of software engineering, from idea to application. Readers will learn not only how to harness the design technologies of relational databases for use on the Web, but also how to transform their conceptual designs of data-intensive Web applications into effective software components. * A fully self-contained introduction and practitioner's guide suitable for both technical and non-technical members of staff, as well as students. * A methodology, development process, and notation (WebML) based on common practice but optimized for the unique challenges of high-volume Web applications. * Completely platform- and product-independent; even the use of WebML is optional. * Based on well-known industry standards such as UML and the Entity Relationship Model. * Enhanced by its own Web site (<http://www.webml.org>), containing additional examples, papers, teaching materials, developers' resources, and exercises with solutions.

The authors explore and explain current techniques for handling the specialised data that describes geographical phenomena in a study that will be of great value to computer scientists and geographers working with spatial databases.

During the last few years, parallel object-relational database management systems have emerged as the leading data management technology on the market. These systems are extensible by user-defined data types and user-defined functionality for the data. This work focuses on the efficient parallel execution of user-defined functionality. The main contributions describe techniques to support data parallelism for user-defined scalar and aggregate functions and intra-function parallelism for the execution of a scalar function on a large object, and a new technology to provide extensibility with regard to new set-oriented database operations that can efficiently implement user-defined functionality in parallel object-relational database management systems.

A thorough presentation of query processing techniques in a broad range of database systems for advanced applications. Provides the most effective query processing techniques and ways to optimize the information retrieval process. Intended for database systems designers creating advanced applications.

Data Model Patterns: A Metadata Map

Database Modeling with Microsoft® Visio for Enterprise Architects

Handbook of Research on Innovative Database Query Processing Techniques

The Practical Guide to Storing, Managing and Analyzing Big and Small Data

Spatial Databases

Transactional Information Systems

Component Database Systems is a collection of invited chapters by the researchers making the most influential contributions in the database industry's trend toward componentization. This book represents the sometimes-divergent, sometimes-convergent approaches taken by leading database vendors as they seek to establish commercially viable componentization strategies. Together, these contributions form the first book devoted entirely to the technical and architectural design of component-based database systems. In addition to detailing the current state of their research, the authors also take up many of the issues affecting the likely future directions of component databases. If you have a stake in the evolution of any of today's leading database systems, this book will make fascinating reading. It will also help prepare you for the technology that is likely to become widely available over the next several years. * Is comprised of contributions from the field's most highly respected researchers, including key figures at IBM, Oracle, Informix, Microsoft, and POET. * Represents the entire spectrum of approaches taken by leading software companies working on DBMS componentization strategies. * Covers component-focused architectures, methods for hooking components into an overall system, and support for component development. * Examines the component technologies that are most valuable to Web-based and multimedia databases. * Presents a thorough classification and overview of component database systems.

There is no other manual for the over 200,000 database administrators using Visio.

Perfectly intelligent programmers often struggle when forced to work with SQL. Why? Joe Celko believes the problem lies with their procedural programming mindset, which keeps them from taking full advantage of the power of declarative languages. The result is overly complex and inefficient code, not to mention lost productivity. This book will change the way you think about the problems you solve with SQL programs.. Focusing on three key table-based techniques, Celko reveals their power through detailed examples and clear explanations. As you master these techniques, you ' ll find you are able to conceptualize problems as rooted in sets and solvable through declarative programming. Before long, you ' ll be coding more quickly, writing more efficient code, and applying the full power of SQL • Filled with the insights of one of the world ' s leading SQL authorities - noted for his knowledge and his ability to teach what he knows. • Focuses on auxiliary tables (for computing functions and other values by joins), temporal tables (for temporal queries, historical data, and audit information), and virtual tables (for improved performance). • Presents clear guidance for selecting and correctly applying the right table technique.

This work has been revised and updated to provide a comprehensive treatment of database design for commercial database products and their applications. The book covers the basic foundation of design as well as more advanced techniques, and also incorporates coverage of data warehousing and OLAP (On-Line Analytical Processing), data mining, object-relational, multimedia, and temporal/spatial design.

Discovering Knowledge from Hypertext Data

Theory, Algorithms, and the Practice of Concurrency Control and Recovery

Advances in Database Technology - EDBT '94

Component Database Systems

Information Visualization in Data Mining and Knowledge Discovery

First International Conference, HPCC 2005, Sorrento, Italy, September, 21-23, 2005, Proceedings

This book constitutes the refereed proceedings of the First International Conference on High-Performance Computing and Communications, HPCC 2005, held in Sorrento, Italy in September 2005. The 76 revised full papers and 44 revised short papers presented were carefully reviewed and selected from 273 submissions. The papers are organized in topical sections on network protocols, routing, and algorithms; languages and compilers for HPC; parallel and distributed system architectures; embedded systems; parallel and distributed algorithms, wireless and mobile computing, Web services and Internet computing; peer-to-peer computing, grid and cluster computing, reliability, fault-tolerance, and security; performance evaluation and measurement; tools and environments for software development; distributed systems and applications; high performance scientific and engineering computing; database applications and data mining; HPSRF; pervasive computing and communications; and LMS.

This book offers a thorough grounding in machine learning concepts combined with practical advice on applying machine learning tools and techniques in real-world data mining situations. Clearly written and effectively illustrated, this book is ideal for anyone involved at any level in the work of extracting usable knowledge from large collections of data. Complementing the book's instruction is fully functional machine learning software.

Research and development surrounding the use of data queries is receiving increased attention from computer scientists and data specialists alike. Through the use of query technology, large volumes of data in databases can be retrieved, and information systems built based on databases can support problem solving and decision making across industries. The Handbook of Research on Innovative Database Query Processing Techniques focuses on the growing topic of database query processing methods, technologies, and applications. Aimed at providing an all-inclusive reference source of technologies and practices in advanced database query systems, this book investigates various techniques, including database and XML queries, spatiotemporal data queries, big data queries, metadata queries, and applications of database query systems. This comprehensive handbook is a necessary resource for students, IT professionals, data analysts, and academicians interested in uncovering the latest methods for using queries as a means to extract information from databases. This all-inclusive handbook includes the latest research on topics pertaining to information retrieval, data extraction, data management, design and development of database queries, and database and XM queries.

Probabilistic databases are databases where the value of some attributes or the presence of some records are uncertain and known only with some probability. Applications in many areas such as information extraction, RFID and scientific data management, data cleaning, data integration, and financial risk assessment produce large volumes of uncertain data, which are best modeled and processed by a probabilistic database. This book presents the state of the art in representation formalisms and query processing techniques for probabilistic data. It starts by discussing the basic principles for representing large probabilistic databases, by decomposing them into tuple-independent tables, block-independent-disjoint tables, or U-databases. Then it discusses two classes of techniques for query evaluation on probabilistic databases. In extensional query evaluation, the entire probabilistic inference can be pushed into the database engine and, therefore, processed as effectively as the evaluation of standard SQL queries. The relational queries that can be evaluated this way are called safe queries. In intensional query evaluation, the probabilistic inference is performed over a propositional formula called lineage expression: every relational query can be evaluated this way, but the data complexity dramatically depends on the query being evaluated, and can be #P-hard. The book also discusses some advanced topics in probabilistic data management such as top-k query processing, sequential probabilistic databases, indexing and materialized views, and Monte Carlo databases. Table of Contents: Overview / Data and Query Model / The Query Evaluation Problem / Extensional Query Evaluation / Intensional Query Evaluation / Advanced Techniques

Data Mining