Data Warehouse Design Solutions

Data Warehouse DesignsThe Data Warehouse ToolkitDimensional Modeling: In a Business Intelligence EnvironmentKimball's Data Warehouse Toolkit ClassicsData Warehousing in the Age of Big DataCorporate Information FactoryThe Data Warehouse ToolkitAgile Data Warehousing for the EnterprisePentaho Kettle SolutionsData Warehousing Design and Advanced Engineering Applications: Methods for Complex ConstructionData Warehousing FundamentalsThe Data Warehouse ETL ToolkitData Warehouse SystemsImpossible Data Warehouse SituationsThe Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence InsightsData Warehousing For DummiesMastering Data Warehouse AggregatesAdvanced Data Warehouse DesignData Mapping for Data Warehouse DesignDesigning a Data WarehouseData ArchitecturePentaho SolutionsData Warehouses and OLAPThe Data Warehouse Lifecycle ToolkitDatabase Systems: Design, Implementation, & ManagementBuilding the Data WarehouseStar Schema The Complete ReferenceData Mining: Concepts and TechniquesData Warehouse DesignBuilding a Data WarehouseData Warehouse Design SolutionsDW 2.0: The Architecture for the Next Generation of Data WarehousingAgile Data Warehouse DesignProgramming Collective IntelligenceThe Biml BookData Architecture: A Primer for the Data ScientistManaging Data in MotionBuilding a Scalable Data Warehouse with Data Vault 2.0Mastering Data Warehouse DesignData Warehouse Designs

Data Warehouse Designs

Data warehousing is one of the hottest business topics, and there's more to understanding data warehousing technologies than you might think. Find out the basics of data warehousing and how it facilitates data mining and business intelligence with Data Warehousing For Dummies, 2nd Edition. Data is probably your company's most important asset, so your data warehouse should serve your needs. The fully updated Second Edition of Data Warehousing For Dummies helps you understand, develop, implement, and use data warehouses, and offers a sneak peek into their future. You'll learn to: Analyze top-down and bottom-up data warehouse designs Understand the structure and technologies of data warehouses, operational data stores, and data marts Choose your project team and apply best development practices to your data warehousing projects Implement a data warehouse, step by step, and involve end-users in the process Review and upgrade existing data storage to make it serve your needs Comprehend OLAP, column-wise databases, hardware assisted databases, and middleware Use data mining intelligently and find what you need Make informed choices about consultants and data warehousing products Data Warehousing For Dummies, 2nd Edition also shows you how to involve users in the testing process and gain valuable feedback, what it takes to successfully manage a data warehouse project, and how to tell if your project is on track. You'll find it's the most useful source of data on the topic!

The Data Warehouse Toolkit

A cutting-edge response to Ralph Kimball's challenge to thedata warehouse community that answers some tough questions aboutthe effectiveness of the

relational approach to datawarehousing Written by one of the best-known exponents of the Bill Inmonapproach to data warehousing Addresses head-on the tough issues raised by Kimball andexplains how to choose the best modeling technique for solvingcommon data warehouse design problems Weighs the pros and cons of relational vs. dimensional modelingtechniques Focuses on tough modeling problems, including creating andmaintaining keys and modeling calendars, hierarchies, transactions,and data quality

Dimensional Modeling: In a Business Intelligence Environment

Data Warehousing in the Age of the Big Data will help you and your organization make the most of unstructured data with your existing data warehouse. As Big Data continues to revolutionize how we use data, it doesn't have to create more confusion. Expert author Krish Krishnan helps you make sense of how Big Data fits into the world of data warehousing in clear and concise detail. The book is presented in three distinct parts. Part 1 discusses Big Data, its technologies and use cases from early adopters. Part 2 addresses data warehousing, its shortcomings, and new architecture options, workloads, and integration techniques for Big Data and the data warehouse. Part 3 deals with data governance, data visualization, information life-cycle management, data scientists, and implementing a Big Data–ready data warehouse. Extensive appendixes include case studies from vendor implementations and a special segment on how we can build a healthcare information factory. Ultimately, this book will help you navigate through the complex layers of Big Data and data warehousing while providing you information on how to effectively think about using all these technologies and the architectures to design the next-generation data warehouse. Learn how to leverage Big Data by effectively integrating it into your data warehouse. Includes real-world examples and use cases that clearly demonstrate Hadoop, NoSQL, HBASE, Hive, and other Big Data technologies Understand how to optimize and tune your current data warehouse infrastructure and integrate newer infrastructure matching data processing workloads and requirements

Kimball's Data Warehouse Toolkit Classics

The "father of data warehousing" incorporates the latesttechnologies into his blueprint for integrated decision supportsystems Today's corporate IT and data warehouse managers are required tomake a small army of technologies work together to ensure fast andaccurate information for business managers. Bill Inmon created theCorporate Information Factory to solve the needs ofthese managers. Since the First Edition, the design of the factoryhas grown and changed dramatically. This Second Edition, revisedand expanded by 40% with five new chapters, incorporates thesechanges. This step-by-step guide will enable readers to connecttheir legacy systems with the data warehouse and deal with a hostof new and changing technologies, including Web access mechanisms,e-commerce systems, ERP (Enterprise Resource Planning) systems. Thebook also looks closely at exploration and data mining servers foranalyzing customer behavior and departmental data marts forfinance, sales, and marketing.

Data Warehousing in the Age of Big Data

The Data Vault was invented by Dan Linstedt at the U.S. Department of Defense, and the standard has been successfully applied to data warehousing projects at organizations of different sizes, from small to large-size corporations. Due to its simplified design, which is adapted from nature, the Data Vault 2.0 standard helps prevent typical data warehousing failures. "Building a Scalable Data Warehouse" covers everything one needs to know to create a scalable data warehouse end to end, including a presentation of the Data Vault modeling technique, which provides the foundations to create a technical data warehouse layer. The book discusses how to build the data warehouse incrementally using the agile Data Vault 2.0 methodology. In addition, readers will learn how to create the input layer (the stage layer) and the presentation layer (data mart) of the Data Vault 2.0 architecture including implementation best practices. Drawing upon years of practical experience and using numerous examples and an easy to understand framework, Dan Linstedt and Michael Olschimke discuss: How to load each layer using SQL Server Integration Services (SSIS), including automation of the Data Vault loading processes. Important data warehouse technologies and practices. Data Quality Services (DQS) and Master Data Services (MDS) in the context of the Data Vault architecture. Provides a complete introduction to data warehousing, applications, and the business context so readers can get-up and running fast Explains theoretical concepts and provides hands-on instruction on how to build and implement a data warehouse Demystifies data vault modeling with beginning, intermediate, and advanced techniques Discusses the advantages of the data vault approach over other techniques, also including the latest updates to Data Vault 2.0 and multiple improvements to Data Vault 1.0

Corporate Information Factory

Geared to IT professionals eager to get into the all-importantfield of data warehousing, this book explores all topics needed bythose who design and implement data warehouses. Readers will learnabout planning requirements, architecture, infrastructure, datapreparation, information delivery, implementation, and maintenance.They'll also find a wealth of industry examples garnered from theauthor's 25 years of experience in designing and implementingdatabases and data warehouse applications for majorcorporations. Market: IT Professionals, Consultants.

The Data Warehouse Toolkit

The definitive guide to dimensional design for your data warehouse Learn the best practices of dimensional design. Star Schema: The Complete Reference offers in-depth coverage of design principles and their underlying rationales. Organized around design concepts and illustrated with detailed examples, this is a step-by-step guidebook for beginners and a comprehensive resource for experts. This all-inclusive volume begins with dimensional design fundamentals and shows how they fit into diverse data warehouse architectures, including those of W.H. Inmon and Ralph Kimball. The book progresses through a series of advanced techniques that help you address real-world complexity, maximize performance, and adapt to the requirements of BI and ETL software products. You are furnished with design tasks and deliverables that can be incorporated into any project, regardless of architecture or methodology. Master the fundamentals of star schema design and

slow change processing Identify situations that call for multiple stars or cubes Ensure compatibility across subject areas as your data warehouse grows Accommodate repeating attributes, recursive hierarchies, and poor data quality Support conflicting requirements for historic data Handle variation within a business process and correlation of disparate activities Boost performance using derived schemas and aggregates Learn when it's appropriate to adjust designs for BI and ETL tools

Agile Data Warehousing for the Enterprise

Data mapping in a data warehouse is the process of creating a link between two distinct data models' (source and target) tables/attributes. Data mapping is required at many stages of DW life-cycle to help save processor overhead; every stage has its own unique requirements and challenges. Therefore, many data warehouse professionals want to learn data mapping in order to move from an ETL (extract, transform, and load data between databases) developer to a data modeler role. Data Mapping for Data Warehouse Design provides basic and advanced knowledge about business intelligence and data warehouse concepts including real life scenarios that apply the standard techniques to projects across various domains. After reading this book, readers will understand the importance of data mapping across the data warehouse life cycle. Covers all stages of data warehousing and the role of data mapping in each Includes a data mapping strategy and techniques that can be applied to many situations Based on the author's years of real-world experience designing solutions

Pentaho Kettle Solutions

Develop a custom, agile data warehousing and business intelligence architecture Empower your users and drive better decision making across your enterprise with detailed instructions and best practices from an expert developer and trainer. The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights shows how to plan, design, construct, and administer an integrated end-to-end DW/BI solution. Learn how to choose appropriate components, build an enterprise data model, configure data marts and data warehouses, establish data flow, and mitigate risk. Change management, data governance, and security are also covered in this comprehensive guide. Understand the components of BI and data warehouse systems Establish project goals and implement an effective deployment plan Build accurate logical and physical enterprise data models Gain insight into your company's transactions with data mining Input, cleanse, and normalize data using ETL (Extract, Transform, and Load) techniques Use structured input files to define data requirements Employ top-down, bottom-up, and hybrid design methodologies Handle security and optimize performance using data governance tools Robert Laberge is the founder of several Internet ventures and a principle consultant for the IBM Industry Models and Assets Lab, which has a focus on data warehousing and business intelligence solutions.

Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction

A complete guide to Pentaho Kettle, the Pentaho Data Integration toolset for ETL This practical book is a complete guide to installing, configuring, and managing Pentaho Kettle. If you're a database administrator or developer, you'll first get up to speed on Kettle basics and how to apply Kettle to create ETL solutions—before progressing to specialized concepts such as clustering, extensibility, and data vault models. Learn how to design and build every phase of an ETL solution. Shows developers and database administrators how to use the open-source Pentaho Kettle for enterprise-level ETL processes (Extracting, Transforming, and Loading data) Assumes no prior knowledge of Kettle or ETL, and brings beginners thoroughly up to speed at their own pace Explains how to get Kettle solutions up and running, then follows the 34 ETL subsystems model, as created by the Kimball Group, to explore the entire ETL lifecycle, including all aspects of data warehousing with Kettle Goes beyond routine tasks to explore how to extend Kettle and scale Kettle solutions using a distributed "cloud" Get the most out of Pentaho Kettle and your data warehousing with this detailed guide—from simple single table data migration to complex multisystem clustered data integration tasks.

Data Warehousing Fundamentals

Here is the ideal field guide for data warehousing implementation. This book first teaches you how to build a data warehouse, including defining the architecture, understanding the methodology, gathering the requirements, designing the data models, and creating the databases. Coverage then explains how to populate the data warehouse and explores how to present data to users using reports and multidimensional databases and how to use the data in the data warehouse for business intelligence, customer relationship management, and other purposes. It also details testing and how to administer data warehouse operation.

The Data Warehouse ETL Toolkit

Three books by the bestselling authors on Data Warehousing! The most authoritative guides from the inventor of the technique all for a value price. The Data Warehouse Toolkit, 3rd Edition (9781118530801) Ralph Kimball invented a data warehousing technique called "dimensional modeling" and popularized it in his first Wiley book, The Data Warehouse Toolkit. Since this book was first published in 1996, dimensional modeling has become the most widely accepted technique for data warehouse design. Over the past 10 years, Kimball has improved on his earlier techniques and created many new ones. In this 3rd edition, he will provide a comprehensive collection of all of these techniques, from basic to advanced. The Data Warehouse Lifecycle Toolkit, 2nd Edition (9780470149775) Complete coverage of best practices from data warehouse project inception through on-going program management. Updates industry best practices to be in sync with current recommendations of Kimball Group. Streamlines the lifecycle methodology to be more efficient and user-friendly The Data Warehouse ETL Toolkit (9780764567575) shows data warehouse developers how to effectively manage the ETL (Extract, Transform, Load) phase of the data warehouse development lifecycle. The authors show developers the best methods for extracting data from scattered sources throughout the enterprise, removing obsolete, redundant, and innaccurate data, transforming the remaining data into correctly formatted data structures, and then physically loading them into the data

warehouse. This book provides complete coverage of proven, time-saving ETL techniques. It begins with a quick overview of ETL fundamentals and the role of the ETL development team. It then quickly moves into an overview of the ETL data structures, both relational and dimensional. The authors show how to build useful dimensional stuctures, providing practical examples of beginning through advanced techniques.

Data Warehouse Systems

Your all-in-one resource for using Pentaho with MySQL forBusiness Intelligence and Data Warehousing Open-source Pentaho provides business intelligence (BI) and datawarehousing solutions at a fraction of the cost of proprietarysolutions. Now you can take advantage of Pentaho for your businessneeds with this practical guide written by two major participantsin the Pentaho community. The book covers all components of the Pentaho BI Suite. You'lllearn to install, use, and maintain Pentaho-and find plenty ofbackground discussion that will bring you thoroughly up to speed onBI and Pentaho concepts. Of all available open source BI products, Pentaho offers themost comprehensive toolset and is the fastest growing open sourceproduct suite Explains how to build and load a data warehouse with PentahoKettle for data integration/ETL, manually create JFree (pentahoreporting services) reports using direct SQL queries, and createMondrian (Pentaho analysis services) cubes and attach them to aJPivot cube browser Review deploying reports, cubes and metadata to the Pentahoplatform in order to distribute BI solutions to end-users Shows how to set up scheduling, subscription and automaticdistribution The companion Web site provides complete source code examples,sample data, and links to related resources.

Impossible Data Warehouse Situations

This is the first book to provide in-depth coverage of star schema aggregates used in dimensional modeling-from selection and design, to loading and usage, to specific tasks and deliverables for implementation projects Covers the principles of aggregate schema design and the pros and cons of various types of commercial solutions for navigating and building aggregates Discusses how to include aggregates in data warehouse development projects that focus on incremental development, iterative builds, and early data loads

The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights

Building upon his earlier book that detailed agile data warehousing programming techniques for the Scrum master, Ralph's latest work illustrates the agile interpretations of the remaining software engineering disciplines: Requirements management benefits from streamlined templates that not only define projects quickly, but ensure nothing essential is overlooked. Data engineering receives two new "hyper modeling" techniques, yielding data warehouses that can be easily adapted when requirements change without having to invest in ruinously expensive data-conversion programs. Quality assurance advances with not only a stereoscopic top-down and bottom-up planning method, but also the incorporation

of the latest in automated test engines. Use this step-by-step guide to deepen your own application development skills through self-study, show your teammates the world's fastest and most reliable techniques for creating business intelligence systems, or ensure that the IT department working for you is building your next decision support system the right way. Learn how to quickly define scope and architecture before programming starts Includes techniques of process and data engineering that enable iterative and incremental delivery Demonstrates how to plan and execute quality assurance plans and includes a guide to continuous integration and automated regression testing Presents program management strategies for coordinating multiple agile data mart projects so that over time an enterprise data warehouse emerges Use the provided 120-day road map to establish a robust, agile data warehousing program

Data Warehousing For Dummies

Nine leading data warehouse experts provide answers to problematic situations encountered by a variety of organizations.

Mastering Data Warehouse Aggregates

Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the methods of knowing, preprocessing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, business professionals, and researchers who seek information on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data

Advanced Data Warehouse Design

Cowritten by Ralph Kimball, the world's leading data warehousing authority, whose previous books have sold more than 150,000 copies Delivers real-world solutions for the most time- and labor-intensive portion of data warehousing-data staging, or the extract, transform, load (ETL) process Delineates best practices for extracting data from scattered sources, removing redundant and inaccurate data,

transforming the remaining data into correctly formatted data structures, and then loading the end product into the data warehouse Offers proven time-saving ETL techniques, comprehensive guidance on building dimensional structures, and crucial advice on ensuring data quality

Data Mapping for Data Warehouse Design

A thorough update to the industry standard for designing, developing, and deploying data warehouse and business intelligence systems The world of data warehousing has changed remarkably since the first edition of The Data Warehouse Lifecycle Toolkit was published in 1998. In that time, the data warehouse industry has reached full maturity and acceptance, hardware and software have made staggering advances, and the techniques promoted in the premiere edition of this book have been adopted by nearly all data warehouse vendors and practitioners. In addition, the term "business intelligence" emerged to reflect the mission of the data warehouse: wrangling the data out of source systems, cleaning it, and delivering it to add value to the business. Ralph Kimball and his colleagues have refined the original set of Lifecycle methods and techniques based on their consulting and training experience. The authors understand first-hand that a data warehousing/business intelligence (DW/BI) system needs to change as fast as its surrounding organization evolves. To that end, they walk you through the detailed steps of designing, developing, and deploying a DW/BI system. You'll learn to create adaptable systems that deliver data and analyses to business users so they can make better business decisions.

Designing a Data Warehouse

Readers gain a solid foundation in database design and implementation with the practical and easy-to-understand approach in DATABASE SYSTEMS: DESIGN, IMPLEMENTATION, AND MANAGEMENT, 12E. Filled with diagrams, illustrations, and tables, this market-leading text provides in-depth coverage of database design. Readers learn the key to successful database implementation: proper design of databases to fit within a larger strategic view of the data environment. Renowned for its clear, straightforward writing style, this text provides an outstanding balance of theory and practice. Updates include the latest coverage of cloud data services and a new chapter on Big Data Analytics and NoSQL, including related Hadoop technologies. In addition, new review questions, problem sets, and cases offer multiple opportunities to test understanding and develop useful design skills. Important Notice: Media content referenced within the product description or the product text may not be available in the ebook version.

Data Architecture

Pentaho Solutions

Written by a team of global experts, this book explains how to design next-generation data warehouses using a structured approach inspired by the modern principles of software engineering. All of the design techniques featured in the

book are placed within a practical framework so that solutions can be applied immediately.Data Warehouse Design: Modern Principles and Methodologies explains how to transform databases into data warehouses, allowing organizations to integrate and analyze data across the enterprise in order to make informed business decisions. Readers can access a CASE (computer aided software engineering) tool developed by the authors from www.qbx-tool.com which implements the discussed methodologies in a practical way.

Data Warehouses and OLAP

Market Basket Analysis (MBA) provides the ability to continually monitor the affinities of a business and can help an organization achieve a key competitive advantage. Time Variant data enables data warehouses to directly associate events in the past with the participants in each individual event. In the past however, the use of these powerful tools in tandem led to performance degradation and resulted in unactionable and even damaging information. Data Warehouse Designs: Achieving ROI with Market Basket Analysis and Time Variance presents an innovative, soup-to-nuts approach that successfully combines what was previously incompatible, without degradation, and uses the relational architecture already in place. Built around two main chapters, Market Basket Solution Definition and Time Variant Solution Definition, it provides a tangible how-to design that can be used to facilitate MBA within the context of a data warehouse. Presents a solution for creating home-grown MBA data marts Includes database design solutions in the context of Oracle, DB2, SQL Server, and Teradata relational database management systems (RDBMS) Explains how to extract, transform, and load data used in MBA and Time Variant solutions The book uses standard RDBMS platforms, proven database structures, standard SQL and hardware, and software and practices already accepted and used in the data warehousing community to fill the gaps left by most conceptual discussions of MBA. It employs a form and language intended for a data warehousing audience to explain the practicality of how data is delivered, stored, and viewed. Offering a comprehensive explanation of the applications that provide, store, and use MBA data, Data Warehouse Designs provides you with the language and concepts needed to require and receive information that is relevant and actionable.

The Data Warehouse Lifecycle Toolkit

Agile Data Warehouse Design is a step-by-step guide for capturing data warehousing/business intelligence (DW/BI) requirements and turning them into high performance dimensional models in the most direct way: by modelstorming (data modeling] brainstorming) with BI stakeholders. This book describes BEAM, an agile approach to dimensional modeling, for improving communication between data warehouse designers, BI stakeholders and the whole DW/BI development team. BEAM provides tools and techniques that will encourage DW/BI designers and developers to move away from their keyboards and entity relationship based tools and model interactively with their colleagues. The result is everyone thinks dimensionally from the outset! Developers understand how to efficiently implement dimensional modeling solutions. Business stakeholders feel ownership of the data warehouse they have created, and can already imagine how they will use it to answer their business questions. Within this book, you will learn: Agile

dimensional modeling using Business Event Analysis & Modeling (BEAM) Modelstorming: data modeling that is quicker, more inclusive, more productive, and frankly more fun! Telling dimensional data stories using the 7Ws (who, what, when, where, how many, why and how) Modeling by example not abstraction; using data story themes, not crow's feet, to describe detail Storyboarding the data warehouse to discover conformed dimensions and plan iterative development Visual modeling: sketching timelines, charts and grids to model complex process measurement - simply Agile design documentation: enhancing star schemas with BEAM dimensional shorthand notation Solving difficult DW/BI performance and usability problems with proven dimensional design patterns LawrenceCorr is a data warehouse designer and educator. As Principal of DecisionOne Consulting, he helps clients to review and simplify their data warehouse designs, and advises vendors on visual data modeling techniques. He regularly teaches agile dimensional modeling courses worldwide and has taught dimensional DW/BI skills to thousands of students. Jim Stagnitto is a data warehouse and master data management architect specializing in the healthcare, financial services, and information service industries. He is the founder of the data warehousing and data mining consulting firm Llumino.

Database Systems: Design, Implementation, & Management

"Each chapter is a practice run for the way we all ought to design our data marts and hence our data warehouses."-Ralph Kimball, from the Foreword. Let the experts show you how to customize data warehouse designs for real business needs in Data Warehouse Design Solutions. To effectively design a data warehouse, you have to understand its many business uses. This guidebook shows you how business managers in different corporate functions actually use data warehouses to make decisions. You'll get a rich set of data warehouse designs that flow from realistic business cases. Two top experts show you how to customize your data warehouse designs for real-life business needs including: * Sales and marketing * Production and inventory management * Budgeting and financial reporting * Quality control * Product delivery and fulfillment * Strategic business analysis such as determining market share, rates of return on investment, and other key analytic ratios. CD-ROM includes All sample data warehouse designs with accompanying preformatted reports in HTML for specific business uses such as marketing, sales, and financial analysis.

Building the Data Warehouse

This exceptional work provides readers with an introduction to the state-of-the-art research on data warehouse design, with many references to more detailed sources. It offers a clear and a concise presentation of the major concepts and results in the subject area. Malinowski and Zimányi explain conventional data warehouse design in detail, and additionally address two innovative domains recently introduced to extend the capabilities of data warehouse systems: namely, the management of spatial and temporal information.

Star Schema The Complete Reference

DW 2.0: The Architecture for the Next Generation of Data Warehousing is the first book on the new generation of data warehouse architecture, DW 2.0, by the father of the data warehouse. The book describes the future of data warehousing that is technologically possible today, at both an architectural level and technology level. The perspective of the book is from the top down: looking at the overall architecture and then delving into the issues underlying the components. This allows people who are building or using a data warehouse to see what lies ahead and determine what new technology to buy, how to plan extensions to the data warehouse, what can be salvaged from the current system, and how to justify the expense at the most practical level. This book gives experienced data warehouse professionals everything they need in order to implement the new generation DW 2.0. It is designed for professionals in the IT organization, including data architects, DBAs, systems design and development professionals, as well as data warehouse and knowledge management professionals. * First book on the new generation of data warehouse architecture, DW 2.0. * Written by the "father of the data warehouse", Bill Inmon, a columnist and newsletter editor of The Bill Inmon Channel on the Business Intelligence Network. * Long overdue comprehensive coverage of the implementation of technology and tools that enable the new generation of the DW: metadata, temporal data, ETL, unstructured data, and data quality control.

Data Mining: Concepts and Techniques

Data warehousing and online analysis technologies have shown their effectiveness in managing and analyzing a large amount of disparate data, attracting much attention from numerous research communities. Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction covers the complete process of analyzing data to extract, transform, load, and manage the essential components of a data warehousing system. A defining collection of field discoveries, this advanced title provides significant industry solutions for those involved in this distinct research community.

Data Warehouse Design

Today, the world is trying to create and educate data scientists because of the phenomenon of Big Data. And everyone is looking deeply into this technology. But no one is looking at the larger architectural picture of how Big Data needs to fit within the existing systems (data warehousing systems). Taking a look at the larger picture into which Big Data fits gives the data scientist the necessary context for how pieces of the puzzle should fit together. Most references on Big Data look at only one tiny part of a much larger whole. Until data gathered can be put into an existing framework or architecture it can't be used to its full potential. Data Architecture a Primer for the Data Scientist addresses the larger architectural picture of how Big Data fits with the existing information infrastructure, an essential topic for the data scientist. Drawing upon years of practical experience and using numerous examples and an easy to understand framework. W.H. Inmon, and Daniel Linstedt define the importance of data architecture and how it can be used effectively to harness big data within existing systems. You'll be able to: Turn textual information into a form that can be analyzed by standard tools. Make the connection between analytics and Big Data Understand how Big Data fits within an

existing systems environment Conduct analytics on repetitive and non-repetitive data Discusses the value in Big Data that is often overlooked, non-repetitive data, and why there is significant business value in using it Shows how to turn textual information into a form that can be analyzed by standard tools Explains how Big Data fits within an existing systems environment Presents new opportunities that are afforded by the advent of Big Data Demystifies the murky waters of repetitive and non-repetitive data in Big Data

Building a Data Warehouse

Managing Data in Motion describes techniques that have been developed for significantly reducing the complexity of managing system interfaces and enabling scalable architectures. Author April Reeve brings over two decades of experience to present a vendor-neutral approach to moving data between computing environments and systems. Readers will learn the techniques, technologies, and best practices for managing the passage of data between computer systems and integrating disparate data together in an enterprise environment. The average enterprise's computing environment is comprised of hundreds to thousands computer systems that have been built, purchased, and acquired over time. The data from these various systems needs to be integrated for reporting and analysis, shared for business transaction processing, and converted from one format to another when old systems are replaced and new systems are acquired. The management of the "data in motion" in organizations is rapidly becoming one of the biggest concerns for business and IT management. Data warehousing and conversion, real-time data integration, and cloud and "big data" applications are just a few of the challenges facing organizations and businesses today. Managing Data in Motion tackles these and other topics in a style easily understood by business and IT managers as well as programmers and architects. Presents a vendor-neutral overview of the different technologies and techniques for moving data between computer systems including the emerging solutions for unstructured as well as structured data types Explains, in non-technical terms, the architecture and components required to perform data integration Describes how to reduce the complexity of managing system interfaces and enable a scalable data architecture that can handle the dimensions of "Big Data"

Data Warehouse Design Solutions

Market Basket Analysis (MBA) provides the ability to continually monitor the affinities of a business and can help an organization achieve a key competitive advantage. Time Variant data enables data warehouses to directly associate events in the past with the participants in each individual event. In the past however, the use of these powerful tools in tandem led to performance degradation and resulted in unactionable and even damaging information. Data Warehouse Designs: Achieving ROI with Market Basket Analysis and Time Variance presents an innovative, soup-to-nuts approach that successfully combines what was previously incompatible, without degradation, and uses the relational architecture already in place. Built around two main chapters, Market Basket Solution Definition and Time Variant Solution Definition, it provides a tangible how-to design that can be used to facilitate MBA within the context of a data warehouse. Presents a solution for creating home-grown MBA data marts Includes

database design solutions in the context of Oracle, DB2, SQL Server, and Teradata relational database management systems (RDBMS) Explains how to extract, transform, and load data used in MBA and Time Variant solutions The book uses standard RDBMS platforms, proven database structures, standard SQL and hardware, and software and practices already accepted and used in the data warehousing community to fill the gaps left by most conceptual discussions of MBA. It employs a form and language intended for a data warehousing audience to explain the practicality of how data is delivered, stored, and viewed. Offering a comprehensive explanation of the applications that provide, store, and use MBA data, Data Warehouse Designs provides you with the language and concepts needed to require and receive information that is relevant and actionable.

DW 2.0: The Architecture for the Next Generation of Data Warehousing

Learn Business Intelligence Markup Language (Biml) for automating much of the repetitive, manual labor involved in data integration. We teach you how to build frameworks and use advanced Biml features to get more out of SQL Server Integration Services (SSIS), Transact-SQL (T-SQL), and SQL Server Analysis Services (SSAS) than you ever thought possible. The first part of the book starts with the basics—getting your development environment configured, Biml syntax, and scripting essentials. Whether a beginner or a seasoned Biml expert, the next part of the book guides you through the process of using Biml to build a framework that captures both your design patterns and execution management. Design patterns are reusable code blocks that standardize the approach you use to perform certain types of data integration, logging, and other key data functions. Design patterns solve common problems encountered when developing data integration solutions. Because you do not have to build the code from scratch each time, design patterns improve your efficiency as a Biml developer. In addition to leveraging design patterns in your framework, you will learn how to build a robust metadata store and how to package your framework into Biml bundles for deployment within your enterprise. In the last part of the book, we teach you more advanced Biml features and capabilities, such as SSAS development, T-SQL recipes, documentation autogeneration, and Biml troubleshooting. The Biml Book: Provides practical and applicable examples Teaches you how to use Biml to reduce development time while improving quality Takes you through solutions to common data integration and BI challenges What You'll Learn Master the basics of Business Intelligence Markup Language (Biml) Study patterns for automating SSIS package generation Build a Biml Framework Import and transform database schemas Automate generation of scripts and projects Who This Book Is For BI developers wishing to quickly locate previously tested solutions, Microsoft BI specialists, those seeking more information about solution automation and code generation, and practitioners of Data Integration Lifecycle Management (DILM) in the DevOps enterprise

Agile Data Warehouse Design

In this IBM Redbooks publication we describe and demonstrate dimensional data modeling techniques and technology, specifically focused on business intelligence

and data warehousing. It is to help the reader understand how to design, maintain, and use a dimensional model for data warehousing that can provide the data access and performance required for business intelligence. Business intelligence is comprised of a data warehousing infrastructure, and a query, analysis, and reporting environment. Here we focus on the data warehousing infrastructure. But only a specific element of it, the data model - which we consider the base building block of the data warehouse. Or, more precisely, the topic of data modeling and its impact on the business and business applications. The objective is not to provide a treatise on dimensional modeling techniques, but to focus at a more practical level. There is technical content for designing and maintaining such an environment, but also business content. For example, we use case studies to demonstrate how dimensional modeling can impact the business intelligence requirements for your business initiatives. In addition, we provide a detailed discussion on the query aspects of BI and data modeling. For example, we discuss query optimization and how you can determine performance of the data model prior to implementation. You need a solid base for your data warehousing infrastructure . . . . a solid data model.

Programming Collective Intelligence

With this textbook, Vaisman and Zimányi deliver excellent coverage of data warehousing and business intelligence technologies ranging from the most basic principles to recent findings and applications. To this end, their work is structured into three parts. Part I describes "Fundamental Concepts" including multi-dimensional models; conceptual and logical data warehouse design and MDX and SQL/OLAP. Subsequently, Part II details "Implementation and Deployment," which includes physical data warehouse design; data extraction, transformation, and loading (ETL) and data analytics. Lastly, Part III covers "Advanced Topics" such as spatial data warehouses; trajectory data warehouses; semantic technologies in data warehouses and novel technologies like Map Reduce, column-store databases and in-memory databases. As a key characteristic of the book, most of the topics are presented and illustrated using application tools. Specifically, a case study based on the well-known Northwind database illustrates how the concepts presented in the book can be implemented using Microsoft Analysis Services and Pentaho Business Analytics. All chapters are summarized using review questions and exercises to support comprehensive student learning. Supplemental material to assist instructors using this book as a course text is available at http://cs.ulb.ac.be/DWSDIbook/, including electronic versions of the figures, solutions to all exercises, and a set of slides accompanying each chapter. Overall, students, practitioners and researchers alike will find this book the most comprehensive reference work on data warehouses, with key topics described in a clear and educational style.

The Biml Book

Data warehouses and online analytical processing (OLAP) are emerging key technologies for enterprise decision support systems. They provide sophisticated technologies from data integration, data collection and retrieval, query optimization, and data analysis to advanced user interfaces. New research and technological achievements in the area of data warehousing are implemented in

commercial database management systems, and organizations are developing data warehouse systems into their information system infrastructures. Data Warehouses and OLAP: Concepts, Architectures and Solutions covers a wide range of technical, technological, and research issues. It provides theoretical frameworks, presents challenges and their possible solutions, and examines the latest empirical research findings in the area. It is a resource of possible solutions and technologies that can be applied when designing, implementing, and deploying a data warehouse, and assists in the dissemination of knowledge in this field.

Data Architecture: A Primer for the Data Scientist

The data warehousing bible updated for the new millennium Updated and expanded to reflect the many technological advances occurring since the previous edition, this latest edition of the data warehousing "bible" provides a comprehensive introduction to building data marts, operational data stores, the Corporate Information Factory, exploration warehouses, and Web-enabled warehouses. Written by the father of the data warehouse concept, the book also reviews the unique requirements for supporting e-business and explores various ways in which the traditional data warehouse can be integrated with new technologies to provide enhanced customer service, sales, and support-both online and offline-including near-line data storage techniques.

Managing Data in Motion

PLEASE PROVIDE COURSE INFORMATION PLEASE PROVIDE

Building a Scalable Data Warehouse with Data Vault 2.0

Data Architecture: From Zen to Reality explains the principles underlying data architecture, how data evolves with organizations, and the challenges organizations face in structuring and managing their data. Using a holistic approach to the field of data architecture, the book describes proven methods and technologies to solve the complex issues dealing with data. It covers the various applied areas of data, including data modelling and data model management, data quality, data governance, enterprise information management, database design, data warehousing, and warehouse design. This text is a core resource for anyone customizing or aligning data management systems, taking the Zen-like idea of data architecture to an attainable reality. The book presents fundamental concepts of enterprise architecture with definitions and real-world applications and scenarios. It teaches data managers and planners about the challenges of building a data architecture roadmap, structuring the right team, and building a long term set of solutions. It includes the detail needed to illustrate how the fundamental principles are used in current business practice. The book is divided into five sections, one of which addresses the software-application development process, defining tools, techniques, and methods that ensure repeatable results. Data Architecture is intended for people in business management involved with corporate data issues and information technology decisions, ranging from data architects to IT consultants, IT auditors, and data administrators. It is also an ideal reference tool for those in a higher-level education process involved in data or

information technology management. Presents fundamental concepts of enterprise architecture with definitions and real-world applications and scenarios Teaches data managers and planners about the challenges of building a data architecture roadmap, structuring the right team, and building a long term set of solutions Includes the detail needed to illustrate how the fundamental principles are used in current business practice

Mastering Data Warehouse Design

Want to tap the power behind search rankings, product recommendations, social bookmarking, and online matchmaking? This fascinating book demonstrates how you can build Web 2.0 applications to mine the enormous amount of data created by people on the Internet. With the sophisticated algorithms in this book, you can write smart programs to access interesting datasets from other web sites, collect data from users of your own applications, and analyze and understand the data once you've found it. Programming Collective Intelligence takes you into the world of machine learning and statistics, and explains how to draw conclusions about user experience, marketing, personal tastes, and human behavior in general -- all from information that you and others collect every day. Each algorithm is described clearly and concisely with code that can immediately be used on your web site, blog, Wiki, or specialized application. This book explains: Collaborative filtering techniques that enable online retailers to recommend products or media Methods of clustering to detect groups of similar items in a large dataset Search engine features -- crawlers, indexers, query engines, and the PageRank algorithm Optimization algorithms that search millions of possible solutions to a problem and choose the best one Bayesian filtering, used in spam filters for classifying documents based on word types and other features Using decision trees not only to make predictions, but to model the way decisions are made Predicting numerical values rather than classifications to build price models Support vector machines to match people in online dating sites Non-negative matrix factorization to find the independent features in a dataset Evolving intelligence for problem solving -- how a computer develops its skill by improving its own code the more it plays a game Each chapter includes exercises for extending the algorithms to make them more powerful. Go beyond simple database-backed applications and put the wealth of Internet data to work for you. "Bravo! I cannot think of a better way for a developer to first learn these algorithms and methods, nor can I think of a better way for me (an old AI dog) to reinvigorate my knowledge of the details." -- Dan Russell, Google "Toby's book does a great job of breaking down the complex subject matter of machine-learning algorithms into practical, easy-to-understand examples that can be directly applied to analysis of social interaction across the Web today. If I had this book two years ago, it would have saved precious time going down some fruitless paths." -- Tim Wolters, CTO, Collective Intellect

Data Warehouse Designs

Updated new edition of Ralph Kimball's groundbreaking book ondimensional modeling for data warehousing and businessintelligence! The first edition of Ralph Kimball's The Data WarehouseToolkit introduced the industry to dimensional modeling,and now his books are considered the most authoritative guides inthis space. This new third edition is a complete library of updateddimensional modeling

techniques, the most comprehensive collectionever. It covers new and enhanced star schema dimensional modelingpatterns, adds two new chapters on ETL techniques, includes new andexpanded business matrices for 12 case studies, and more. Authored by Ralph Kimball and Margy Ross, known worldwide aseducators, consultants, and influential thought leaders in datawarehousing and business intelligence Begins with fundamental design recommendations and progressesthrough increasingly complex scenarios Presents unique modeling techniques for business applicationssuch as inventory management, procurement, invoicing, accounting,customer relationship management, big data analytics, and more Draws real-world case studies from a variety of industries,including retail sales, financial services, telecommunications,education, health care, insurance, e-commerce, and more Design dimensional databases that are easy to understand andprovide fast query response with The Data WarehouseToolkit: The Definitive Guide to Dimensional Modeling, 3rdEdition.

ROMANCE ACTION & ADVENTURE MYSTERY & THRILLER BIOGRAPHIES & HISTORY CHILDREN'S YOUNG ADULT FANTASY HISTORICAL FICTION HORROR LITERARY FICTION NON-FICTION SCIENCE FICTION