

## Data Warehouse Design Solutions

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.

Foreword by Mark Stephen LaRow, Vice President of Products, MicroStrategy "A unique and authoritative book that blends recent research developments with industry-level practices for researchers, students, and industry practitioners." Il-Yeol Song, Professor, College of Information Science and Technology, Drexel University

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.

Updated new edition of Ralph Kimball's groundbreaking book on dimensional modeling for data warehousing and business intelligence! The first edition of Ralph Kimball's *The Data Warehouse Toolkit* introduced the industry to dimensional modeling, and now his books are considered the most authoritative guides in this space. This new third edition is a complete library of updated dimensional modeling techniques, the most comprehensive collection ever. It covers new and enhanced star schema dimensional modeling patterns, adds two new chapters on ETL techniques, includes new and expanded business matrices for 12 case studies, and more. Authored by Ralph Kimball and Margy Ross, known worldwide as educators, consultants, and influential thought leaders in data warehousing and business intelligence Begins with fundamental design recommendations and progresses through increasingly complex scenarios Presents unique modeling techniques for business applications such 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 and provide fast query response with *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition*.

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

*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 Lawrence Corr 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.

The final edition of the incomparable data warehousing and business intelligence reference, updated and expanded *The Kimball Group Reader, Remastered Collection* is the essential reference for data warehouse and business intelligence design, packed with best practices, design tips, and valuable insight from industry pioneer Ralph Kimball and the Kimball Group. This Remastered Collection represents decades of expert advice and mentoring in data warehousing and business intelligence, and is the final work to be published by the Kimball Group. Organized for quick navigation and easy reference, this book contains nearly 20 years of experience on more than 300 topics, all fully up-to-date and expanded with 65 new articles. The discussion covers the complete data warehouse/business intelligence lifecycle, including project planning, requirements gathering, system architecture, dimensional modeling, ETL, and business intelligence analytics, with each group of articles prefaced by original commentaries explaining their role in the overall Kimball Group methodology. Data warehousing/business intelligence industry's current multi-billion dollar value is due in no small part to the contributions of Ralph Kimball and the Kimball Group. Their publications are the standards on which the industry is built, and nearly all data warehouse hardware and software vendors have adopted their methods in one form or another. This book is a compendium of Kimball Group expertise, and an essential reference for anyone in the field. Learn data warehousing and business intelligence from the field's pioneers Get up to date on best practices and essential design tips Gain valuable knowledge on every stage of the project lifecycle Dig into the Kimball Group methodology with hands-on guidance Ralph Kimball and the Kimball Group have continued to refine their methods and techniques based on thousands of hours of consulting and training. This Remastered Collection of *The Kimball Group Reader* represents their final body of knowledge, and is nothing less than a vital reference for anyone involved in the field.

The first comprehensive handbook on star schema design *The Star Schema Handbook* is a comprehensive guide to dimensional modeling covering both basic and advanced topics. Organized around technical concepts rather than business examples, this is the perfect resource for data warehouse designers or developers. The book is architecture-neutral, providing full coverage of the best practices for the design of star schema and OLAP cubes in any environment. You'll find numerous examples of designs drawn from real-world situations. The book covers schema capabilities, ETL loading, reporting, and more. *The Star Schema Handbook* is a complete reference for dimensional modeling

and star schema design, whether you're new to the field or are seeking advanced information to help you address specific design challenges. The book explores best practices for star schema design and OLAP cubes; it is architecture-neutral, providing guidance on best design practices for any environment. Topics include schema capabilities, ETL loading, reporting, and much more. This comprehensive guide covers both basic and advanced topics, making it the complete reference for all data warehouse designers.

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.

Your all-in-one resource for using Pentaho with MySQL for Business Intelligence and Data Warehousing. Open-source Pentaho provides business intelligence (BI) and data warehousing solutions at a fraction of the cost of proprietary solutions. Now you can take advantage of Pentaho for your business needs with this practical guide written by two major participants in the Pentaho community. The book covers all components of the Pentaho BI Suite. You'll learn to install, use, and maintain Pentaho and find plenty of background discussion that will bring you thoroughly up to speed on BI and Pentaho concepts. Of all available open source BI products, Pentaho offers the most comprehensive toolset and is the fastest growing open source product suite. Explains how to build and load a data warehouse with Pentaho Kettle for data integration/ETL, manually create JFree (Pentaho reporting services) reports using direct SQL queries, and create Mondrian (Pentaho analysis services) cubes and attach them to a JPIVOT cube browser. Review deploying reports, cubes and metadata to the Pentaho platform in order to distribute BI solutions to end-users. Shows how to set up scheduling, subscription and automatic distribution. The companion Web site provides complete source code examples, sample data, and links to related resources.

Open Source Data Warehousing and Business Intelligence is an all-in-one reference for developing open source based data warehousing (DW) and business intelligence (BI) solutions that are business-centric, cross-customer viable, cross-functional, cross-technology based, and enterprise-wide. Considering the entire lifecycle of an open source DW &

The "father of data warehousing" incorporates the latest technologies into his blueprint for integrated decision support systems. Today's corporate IT and data warehouse managers are required to make a small army of technologies work together to ensure fast and accurate information for business managers. Bill Inmon created the Corporate Information Factory to solve the needs of these managers. Since the First Edition, the design of the factory has grown and changed dramatically. This Second Edition, revised and expanded by 40% with five new chapters, incorporates these changes. This step-by-step guide will enable readers to connect their legacy systems with the data warehouse and deal with a host of new and changing technologies, including Web access mechanisms, e-commerce systems, ERP (Enterprise Resource Planning) systems. The book also looks closely at exploration and data mining servers for analyzing customer behavior and departmental data marts for finance, sales, and marketing.

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.

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.

This book constitutes the refereed proceedings of the Second International Conference on Advances in Communication, Network, and Computing, CNC 2011, held in Bangalore, India, in March 2011. The 41 revised full papers, presented together with 50 short papers and 39 poster papers, were carefully reviewed and selected for inclusion in the book. The papers feature current research in the field of Information Technology, Networks, Computational Engineering, Computer and Telecommunication Technology, ranging from theoretical and methodological issues to advanced applications.

"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.

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"

The new edition of the classic bestseller that launched the data warehousing industry covers new approaches and technologies, many of which have been pioneered by Inmon himself In addition to explaining the fundamentals of data warehouse systems, the book covers new topics such as methods for handling unstructured data in a data warehouse and storing data across multiple storage media Discusses the pros and cons of relational versus multidimensional design and how to measure return on investment in planning data warehouse projects Covers advanced topics, including data monitoring and testing Although the book includes an extra 100 pages worth of valuable content, the price has actually been reduced from \$65 to \$55

A cutting-edge response to Ralph Kimball's challenge to the data warehouse community that answers some tough questions about the effectiveness of the relational approach to data warehousing Written by one of the best-known exponents of the Bill Inmon approach to data warehousing Addresses head-on the tough issues raised by Kimball and explains how to choose the best modeling technique for solving common data warehouse design problems Weighs the pros and cons of relational vs. dimensional modeling techniques Focuses on tough modeling problems, including creating and maintaining keys and modeling calendars, hierarchies, transactions, and data quality

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

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.

Learn essential techniques from data warehouse legend Bill Inmon on how to build the reporting environment your business needs now! Answers for many valuable business questions hide in text. How well can your existing reporting environment extract the necessary text from email, spreadsheets, and documents, and put it in a useful format for analytics and reporting? Transforming the traditional data warehouse into an efficient unstructured data warehouse requires additional skills from the analyst, architect, designer, and developer. This book will prepare you to successfully implement an unstructured data warehouse and, through clear explanations, examples, and case studies, you will learn new techniques and tips to successfully obtain and analyze text. Master these ten objectives: • Build an unstructured data warehouse using the 11-step approach • Integrate text and describe it in terms of homogeneity, relevance, medium, volume, and structure • Overcome challenges including blather, the Tower of Babel, and lack

of natural relationships • Avoid the Data Junkyard and combat the “Spider’s Web” • Reuse techniques perfected in the traditional data warehouse and Data Warehouse 2.0, including iterative development • Apply essential techniques for textual Extract, Transform, and Load (ETL) such as phrase recognition, stop word filtering, and synonym replacement • Design the Document Inventory system and link unstructured text to structured data • Leverage indexes for efficient text analysis and taxonomies for useful external categorization • Manage large volumes of data using advanced techniques such as backward pointers • Evaluate technology choices suitable for unstructured data processing, such as data warehouse appliances The following outline briefly describes each chapter’s content: • Chapter 1 defines unstructured data and explains why text is the main focus of this book. The sources for text, including documents, email, and spreadsheets, are described in terms of factors such as homogeneity, relevance, and structure. • Chapter 2 addresses the challenges one faces when managing unstructured data. These challenges include volume, blather, the Tower of Babel, spelling, and lack of natural relationships. Learn how to avoid a data junkyard, which occurs when unstructured data is not properly integrated into the data warehouse. This chapter emphasizes the importance of storing integrated unstructured data in a relational structure. We are cautioned on both the commonality and dangers associated with text based on paper. • Chapter 3 begins with a timeline of applications, highlighting their evolution over the decades. Eventually, powerful yet siloed applications created a “spider’s web” environment. This chapter describes how data warehouses solved many problems, including the creation of corporate data, the ability to get out of the maintenance backlog conundrum, and greater data integrity and data accessibility. There were problems, however, with the data warehouse that were addressed in Data Warehouse 2.0 (DW 2.0), such as the inevitable data lifecycle. This chapter discusses the DW 2.0 architecture, which leads into the role of the unstructured data warehouse. The unstructured data warehouse is defined and benefits are given. There are several features of the conventional data warehouse that can be leveraged for the unstructured data warehouse, including ETL processing, textual integration, and iterative development. • Chapter 4 focuses on the heart of the unstructured data warehouse: Textual Extract, Transform, and Load (ETL). This chapter has separate sections on extracting text, transforming text, and loading text. The chapter emphasizes the issues around source data. There are a wide variety of sources, and each of the sources has its own set of considerations. Extracting pointers are provided, such as reading documents only once and recognizing common and different file types. Transforming text requires addressing many considerations discussed in this chapter, including phrase recognition, stop word filtering, and synonym replacement. Loading text is the final step. There are important points to understand here, too, that are explained in this chapter, such as the importance of the thematic approach and knowing how to handle large volumes of data. Two ETL examples are provided, one on email and one on spreadsheets. • Chapter 5 describes the 11 steps required to develop the unstructured data warehouse. The methodology explained in this chapter is a combination of both traditional system development lifecycle and spiral approaches. • Chapter 6 describes how to inventory documents for maximum analysis value, as well as link the unstructured text to structured data for even greater value. The Document Inventory is discussed, which is similar to a library card catalog used for organizing corporate documents. This chapter explores ways of linking unstructured text to structured data. The emphasis is on taking unstructured data and reducing it into a form of data that is structured. Related concepts to linking, such as probabilistic linkages and dynamic linkages, are discussed. • Chapter 7 goes through each of the different types of indexes necessary to make text analysis efficient. Indexes range from simple indexes, which are fast to create and are good if the analyst really knows what needs to be analyzed before the indexing process begins, to complex combined indexes, which can be made up of any and all of the other kinds of indexes. • Chapter 8 explains taxonomies and how they can be used within the unstructured data warehouse. Both simple and complicated taxonomies are discussed. Techniques to help the reader leverage taxonomies, including using preferred taxonomies, external categorization, and cluster analysis are described. Real world problems are raised, including the possibilities of encountering hierarchies, multiple types, and recursion. The chapter ends with a discussion comparing a taxonomy with a data model. • Chapter 9 explains ways of coping with large amounts of unstructured data. Techniques such as keeping the unstructured data at its source and using backward pointers are discussed. The chapter explains why iterative development is so important. Ways of reducing the amount of data are presented, including screening and removing extraneous data, as well as parallelizing the workload. • Chapter 10 focuses on challenges and some technology choices that are suitable for unstructured data processing. The traditional data warehouse processing technology is reviewed. In addition, the data warehouse appliance is discussed. • Chapters 11, 12, and 13 put all of the previously discussed techniques and approaches in context through three case studies: the Ablatz Medical Group, the Eastern Hills Oil Company, and the Amber Oil Company.

This book helps you to understand Snowflake's unique architecture and ecosystem that places it at the forefront of cloud data warehouses. The recipes present in this book will enable you to develop proficiency in managing data on Snowflake and learn Snowflake's novel features such as data sharing, cloning, and time travel.

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 Master the most agile and resilient design for building analytics applications: the Unified Star Schema (USS) approach. The USS

has many benefits over traditional dimensional modeling. Witness the power of the USS as a single star schema that serves as a foundation for all present and future business requirements of your organization. Data warehouse legend Bill Inmon and business intelligence innovator, Francesco Puppini, explain step-by-step why the Unified Star Schema is the recommended approach for business intelligence designs today, and show through many examples how to build and use this new solution. This book contains two parts. Part I, Architecture, explains the benefits of data marts and data warehouses, covering how organizations progressed to their current state of analytics, and to the challenges that result from current business intelligence architectures. Chapter 1 covers the drivers behind and the characteristics of the data warehouse and data mart. Chapter 2 introduces dimensional modeling concepts, including fact tables, dimensions, star joins, and snowflakes. Chapter 3 recalls the evolution of the data mart. Chapter 4 explains Extract, Transform, and Load (ETL), and the value ETL brings to reporting. Chapter 5 explores the Integrated Data Mart Approach, and Chapter 6 explains how to monitor this environment. Chapter 7 describes the different types of metadata within the data warehouse environment. Chapter 8 progresses through the evolution to our current modern data warehouse environment. Part II, the Unified Star Schema, covers the Unified Star Schema (USS) approach and how it solves the challenges introduced in Part I. There are eight chapters within Part II: · Chapter 9, Introduction to the Unified Star Schema: Learn about its architecture and use cases, as well as how the USS approach differs from the traditional approach. · Chapter 10, Loss of Data: Learn about the loss of data and the USS Bridge. Understand that the USS approach does not create any join, and for this reason, it has no loss of data. · Chapter 11, The Fan Trap: Get introduced to the Oriented Data Model convention, and learn the dangers of a fan trap through an example. Differentiate join and association, and realize that an "in-memory association" is the preferred solution to the fan trap. · Chapter 12, The Chasm Trap: Become familiar with the Cartesian product, and then follow along with an example based on LinkedIn, which illustrates that a chasm trap produces unwanted duplicates. See that the USS Bridge is based on a union, which does not create any duplicates. · Chapter 13, Multi-Fact Queries: Distinguish between multiple facts "with direct connection" versus multiple facts "with no direct connection". Explore how BI tools are capable of building aggregated virtual rows. · Chapter 14, Loops: Learn more about loops and five traditional techniques to solve them. Follow along with an implementation, which will illustrate the solution based on the USS approach. · Chapter 15, Non-Conformed Granularities: Learn about non-conformed granularities, and learn that the Unified Star Schema introduces a solution called "re-normalization". · Chapter 16, Northwind Case Study. Witness how easy it is to detect the pitfalls of Northwind using the ODM convention. Follow along with an implementation of the USS approach on the Northwind database with various BI tools.

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

As it is with building a house, most of the work necessary to build a data warehouse is neither visible nor obvious when looking at the completed product. While it may be easy to plan for a data warehouse that incorporates all the right concepts, taking the steps needed to create a warehouse that is as functional and user-friendly as it is theoreti

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.

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.

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!

PLEASE PROVIDE COURSE INFORMATION PLEASE PROVIDE

*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.

In recent years, the science of managing and analyzing large datasets has emerged as a critical area of research. In the race to answer vital questions and make knowledgeable decisions, impressive amounts of data are now being generated at a rapid pace, increasing the opportunities and challenges associated with the ability to effectively analyze this data. 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

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

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](http://www.qbx-tool.com) which implements the discussed methodologies in a practical way.

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 Geared to IT professionals eager to get into the all-important field of data warehousing, this book explores all topics needed by those who design and implement data warehouses. Readers will learn about planning requirements, architecture, infrastructure, data preparation, information delivery, implementation, and maintenance. They'll also find a wealth of industry examples garnered from the author's 25 years of experience in designing and implementing databases and data warehouse applications for major corporations. Market: IT Professionals, Consultants.

Build a modern data warehouse on Microsoft's Azure Platform that is flexible, adaptable, and fast—fast to snap together, reconfigure, and fast at delivering results to drive good decision making in your business. Gone are the days when data warehousing projects were lumbering dinosaur-style projects that took forever, drained budgets, and produced business

intelligence (BI) just in time to tell you what to do 10 years ago. This book will show you how to assemble a data warehouse solution like a jigsaw puzzle by connecting specific Azure technologies that address your own needs and bring value to your business. You will see how to implement a range of architectural patterns using batches, events, and streams for both data lake technology and SQL databases. You will discover how to manage metadata and automation to accelerate the development of your warehouse while establishing resilience at every level. And you will know how to feed downstream analytic solutions such as Power BI and Azure Analysis Services to empower data-driven decision making that drives your business forward toward a pattern of success. This book teaches you how to employ the Azure platform in a strategy to dramatically improve implementation speed and flexibility of data warehousing systems. You will know how to make correct decisions in design, architecture, and infrastructure such as choosing which type of SQL engine (from at least three options) best meets the needs of your organization. You also will learn about ETL/ELT structure and the vast number of accelerators and patterns that can be used to aid implementation and ensure resilience. Data warehouse developers and architects will find this book a tremendous resource for moving their skills into the future through cloud-based implementations.

**What You Will Learn**

- Choose the appropriate Azure SQL engine for implementing a given data warehouse
- Develop smart, reusable ETL/ELT processes that are resilient and easily maintained
- Automate mundane development tasks through tools such as PowerShell
- Ensure consistency of data by creating and enforcing data contracts
- Explore streaming and event-driven architectures for data ingestion
- Create advanced staging layers using Azure Data Lake Gen 2 to feed your data warehouse

**Who This Book Is For**

Data warehouse or ETL/ELT developers who wish to implement a data warehouse project in the Azure cloud, and developers currently working in on-premise environments who want to move to the cloud, and for developers with Azure experience looking to tighten up their implementation and consolidate their knowledge

[Copyright: c571fb15b8f89f72b79d019fe89fc6c2](#)