

Data Warehouse From Architecture

A step-by-step guide to building Web-enabled data warehouses fast, this title helps readers choose the best platforms, technologies, and security techniques. Other topics include CORBA and COM distributed object solutions, data marts, data mining, and OLAP.

Building a Data Warehouse: With Examples in SQL Server describes how to build a data warehouse completely from scratch and shows practical examples on how to do it. Author Vincent Rainardi also describes some practical issues he has experienced that developers are likely to encounter in their first data warehousing project, along with solutions and advice. The relational database management system (RDBMS) used in the examples is SQL Server: the version will not be an issue as long as the user has SQL Server or later. The book is organized as follows. In the beginning of this book (chapters 1 through 6), you learn how to build a data warehouse, for example, defining the architecture, understanding the methodology, gathering the requirements, designing the data models, and creating the databases. Then in chapters 7 through 10, you learn how to populate the data warehouse, for example, extracting from source systems, loading the data stores, maintaining data quality, and utilizing the metadata. After you populate the data warehouse, in chapters 11 through 15, you explore 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. Chapters 16 and 17 wrap up the book. After you have built your data warehouse, before it can be released to production, you need to test it thoroughly. After your application is in production, you need to understand how to administer data warehouse operation and you'll learn a detailed understanding of what it takes to build a data warehouse. The implementation code in SQL Server to build the data warehouse. Dimensional modeling, data extraction methods, data warehouse loading, populating dimension and fact tables, data quality, data warehouse architecture, and database design. Practical data warehousing applications such as business intelligence reports, analytics applications, and customer relationship management. Who this book is for: There are three audiences for the book. The first, the people who implement the data warehouse. This could be considered a field guide for them. The second is database users/admins who want to get a good understanding of what it would take to build a data warehouse. Finally, the third audience is managers who must make decisions about aspects of the data warehousing task before them and use the book to learn about these issues.

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 picture. 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 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 in Big Data

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 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 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.

Building the Data Warehouse

The Data Warehouse Toolkit

The Modern Data Warehouse in Azure

Modeling the Agile Data Warehouse with Data Vault

Ecology of Data Warehouse Second Edition

Data Warehousing in Action

Over time the architecture of data warehouse has evolved towards an architecture known as Data Warehouse (DW) 2.0. In DW 2.0 there have been several advances including the inclusion of unstructured data into the data warehouse, the need for a formal and enterprise wide inclusion of corporate metadata. This course includes an overview to DW 2.0 including: An introduction to DW 2.0. We explore the traditional definition of a data warehouse as subject oriented, integrated, non-volatile, and time variant. We also explore the demands of unstructured data on the data warehouse and what makes the DW 2.0 architecture both unique and powerful. The DW 2.0 Lifecycle. Data can start off as interactive which is very current (up to the second), then integrated (current, hours to five years), and finally archival (older than five years). Archival within DW 2.0. We cover archival, which is when the primary usage of the data is done (that is, probability of access is low) yet the data still needs to be maintained by the organization. Data stored in archive can originate from the big data arena and contain both structured and unstructured data. Metadata is physically and tightly coupled with the data that resides in the archival sector. Data may be periodically retrieved from archival on a project basis for deeper analysis. DW 2.0 Components. We explore each component of the data warehouse architecture including applications, procedures, programs, databases, and transactions. The structures within DW 2.0 are organized by subject area such as Customer and Product. We will also discuss the Operational Data Store (ODS). DW 2.0 Database Design. The DW 2.0 contains different types of data. Therefore, there are different ways to do database design, which are covered within this video segment. We discuss the Interactive Sector, which demands a two to three second response time and 24 x 7 availability. The Integrated Sector of the architecture contains lots of data with this data being used for many different purposes. There is a heavy amount of indexing within the integrated sector. We also explore data mining within the integrated sector. With data mining the requirements are not provided or known, and usually the design resembles a spreadsheet in the form of flat records. We also discuss exploration processing and the role of historical data. DW 2.0 Integrated Design. We cover the integrated sector of DW 2.0 in ...

The first comparative review of the state of the art and best current practice in data warehousing. It covers source and data integration, multidimensional aggregation, query optimisation, update propagation, metadata management, quality assessment, and design optimisation. Also, based on results of the European DWQ project, it offers a conceptual framework by which the architecture and quality of data warehousing efforts can be assessed and improved using enriched metadata management combined with advanced techniques from databases, business modelling, and artificial intelligence. An excellent introduction to the issues of quality and metadata usage for researchers and database professionals in academia and industry. XXXXXXX Neuer Text This book presents the first comparative review of the state-of-the-art and the best current practices of data warehouses. It covers source and data integration, multidimensional aggregation, query optimization, metadata management, quality assessment, and design optimization. A conceptual framework is presented by which the architecture and quality of a data warehouse can be assessed and improved using enriched metadata management combined with advanced techniques from databases, business modeling, and artificial intelligence.

Completely revised, expanded, and updated, this second edition gives extensive new coverage of data integration, management, indexing, cleansing, and transformation. The book covers powerful new multi-dimensional front-ends and conversion tools and gives detailed coverage of lifecycle issues.

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.

Data Warehousing with Service-Oriented Architecture

Handbook of Research on Big Data Storage and Visualization Techniques

Data Lake Architecture

Learn Data Warehousing in 24 Hours

A Primer for the Data Scientist

PLEASE PROVIDE COURSE INFORMATION PLEASE PROVIDE

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.

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 you exist reporting environment create 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 "Spiders' 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 issues. Extracting pointers to source data, identifying 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 things 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 technique 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.

In this book, I introduce a new architectural layer that integrates SOA applications with a near-real-time data warehouse. Through this new "SOA Integration Layer" a new Information Technology (IT) solution is created that realizes even greater benefits to the business as a result of combining these individual information technologies. This new IT solution will connect the SOA integration layer to near-real-time data warehouse and consider the near-real-time data warehouse as a data stage source for other business and system environments. This new IT solution is unique because it focuses on relating SOA application environments to near-real-time data warehouse to support both business operations and business strategic decision making. This study will propose a new prototype model in order to build a SOA integration layer between the user and the data management layer, and to connect this SOA integration layer to the data management layer which includes a near-real-time data warehouse. The expected outcomes from this book will be combining near-real-time data warehousing with SOA architecture to deliver integrated tactical and strategic data.

The Complete Guide to Dimensional Modeling

Corporate Information Factory

Building a Data Warehouse

Data Architecture

Data Warehouse 2.0

Building the Data Lakehouse

The data lakehouse is the next generation of the data warehouse and data lake, designed to meet today's complex and ever-changing analytics, machine learning, and data science requirements. Learn about the features and architecture of the data lakehouse, along with its powerful analytical infrastructure. Appreciate how the universal common connector blends structured, textual, analog, and IoT data. Maintain the lakehouse for future generations through Data Lakehouse Housekeeping and Data Future-proofing. Know how to incorporate the lakehouse into an existing data governance strategy. Incorporate data catalogs, data lineage tools, and open source software into your architecture to ensure your data scientists, analysts, and end users live happily ever after.

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 that is easy to manage and specific to your business. 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 Analytics 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

The first, step-by-step guide to building Web-enabled data warehouses The Web can be an incredibly rich source of customer data, and right now companies across industry sectors are hustling to get up and running with data warehouses capable of capturing the clickstream data from their Web sites. This allows companies to track exactly where a customer is going, or "clicking to," on their site in order to gain meaningful information about that customer's preferences. Following Ralph Kimball's The Data Webhouse Toolkit (0-471-37680-9) where he provides the blueprint, Clickstream Data Warehousing fills developers in on all the technical details that go into building a Web-enabled data warehouse. The authors review all key architectural and design issues that developers need to masterfully build a Webhouse using examples to illustrate key points. Companion Web site features code examples from the book and links to related Web sites.

Agile Data Warehousing for the Enterprise

From Zen to Reality

Data Warehousing Fundamentals

Building the Unstructured Data Warehouse

Data Warehouse Architecture for Army Installations

Building a Data Warehouse

Data Architecture

Data Warehouse 2.0

Building the Data Lakehouse

Data warehousing is one of the hottest topics in the computing industry. Written by Barry Devlin, one of the world's leading experts on data warehousing, this book gives you the insights and experiences gained over 10 years and offers the most comprehensive, practical guide to designing, building, and implementing a successful data warehouse. Included in this vital information is an explanation of the optimal three-tiered architecture for the data warehouse, with a clear division between data and information. Information systems managers will appreciate the full description of the functions needed to implement such an architecture, including reconciling existing, diverse data and deriving consistent, valuable business information.

Data Warehouse From Architecture to Implementation Addison-Wesley Professional Do You Know If Your Data Warehouse Flexible, Scalable, Secure and Will It Stand The Test Of Time And Avoid Being Part Of The Dreaded "Life Cycle"? The Data Vault took the Data Warehouse world by storm when it was released in 2001. Some of the world's largest and most complex data warehouse situations understood the value it gave especially with the capabilities of unlimited scaling, flexibility and security. Here is what industry leaders say about the Data Vault "The Data Vault is the optimal choice for modeling the EDW in the DW 2.0 framework" - Bill Inmon, The Father of Data Warehousing "The Data Vault is foundationally strong and an exceptionally scalable architecture" - Stephen Brobst, CTO, Teradata "The Data Vault should be considered as a potential standard for RDBMS-based analytic data management by organizations looking to achieve a high degree of flexibility, performance and openness" - Doug Lane, Deloitte Analytics Institute "I applaud Dan's contribution to the body of Business Intelligence and Data Warehousing knowledge and recommend this book be read by both data professionals and end users" - Howard Dresner, From the Foreword - Speaker, Author, Leading Research Analyst and Advisor You have in your hands the work, experience and testing of 2 decades of building data warehouses. The Data Vault model and methodology has proven itself in hundreds (perhaps thousands) of solutions in Insurance, Crime-Fighting, Defense, Retail, Finance, Banking, Power, Energy, Education, High-Tech and many more. Learn the techniques and implement them and learn how to build your Data Warehouse faster than you have ever done before while designing it to grow and scale no matter what you throw at it. Ready to "Super Charge Your Data Warehouse"?

Construct and implement a data warehousing plan. In their efforts to collect information that will give them an edge, many companies have amassed vast amounts of data. Often this data becomes unwieldy and difficult to translate into anything useful. Data warehousing, storing all of this data in a system that allows for rapid retrieval of customized information, is the solution. This book, written by a data warehousing authority, is a step-by-step guide to creating and managing a data warehouse from start to finish, reviewing marketing, technology, and design issues.

Big Data, Data Warehouse and Data Vault

Data Warehousing in the Age of Big Data

Data Architecture: A Primer for the Data Scientist

Interactive Data Warehousing

Building a Star Schema

Architecture and Implementation

As huge volumes of data, 24/7 performance requirements, and the need to integrate customer information are outstripping the capacity of many existing data warehouses, the question of how to support this explosive growth of e-business is becoming the hottest topic in data warehousing. Building upon Bill Inmon's proven corporate information factory (CIF) architecture, this team of experts show data warehouse managers and developers how to architect the data warehouse environment to meet the high-performance requirements of data-intensive Web sites as well as the needs of users across the enterprise. Readers will get detailed guidance on how to build the data warehouse or improve existing capabilities in order to: o Capture clickstream data that customers enter at the Webhouse o Interface with HTML, XML, and other Web applications o Integrate clickstream data with legacy database systems o Design the data warehouse for optimum performance and scalability o Store and archive high volumes of data at reasonable costs o Analyze e-business data to spot new marketing and sales opportunities

CUTTING-EDGE CONTENT AND GUIDANCE FROM A DATA WAREHOUSING EXPERT! NOW EXPANDED TO REFLECT FIELD TRENDS Data warehousing has revolutionized the way businesses in a wide variety of industries perform analysis and make strategic decisions. Since the first edition of Data Warehousing Fundamentals, numerous enterprises have implemented data warehouse systems and reaped enormous benefits. Many more are in the process of doing so. Now, this new, revised edition covers the essential fundamentals of data warehousing and intelligence as well as significant new trends in the field. The author provides an enhanced, comprehensive overview of data warehousing together with in-depth explanations of critical issues in planning, design, deployment, and ongoing maintenance. IT professionals eager to get into the field will gain a clear understanding of techniques for data extraction from source systems, data cleansing, data transformations, data warehouse architecture and infrastructure, and the various methods for information delivery. This practical Second Edition highlights the areas of data warehousing and business intelligence where high-impact technological progress has been made. Discussions on developments include data marts, real-time information delivery, data visualization, requirements gathering methods, multi-tier architecture, OLAP applications, Web clickstream analysis, data warehouse appliances, and data mining techniques. The book also contains review questions and exercises for each chapter, appropriate for self-study or classroom work, industry examples of real-world situations, and several appendices with valuable information. Specifically written for professionals responsible for designing, implementing, or maintaining data warehousing systems, Data Warehousing Fundamentals presents agile, thorough, and systematic development principles for the IT professional and anyone working or researching in information management.

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.

U.S. Army installations are enterprises performing tasks such as management of funds, budgeting, estimating, managing facilities, maintaining facilities, providing training, complying with environmental and safety laws and regulations. Installations use many database management and operational systems to conduct these tasks. A data warehouse is a single store of information to answer complex queries from management using cross-functional data to perform advanced data analysis methods and to compare with historical data. In the data warehousing approach, the cleansed and transformed data from several operational systems is stored in a single integrated repository of information. This approach provides easy access to needed data, improves system response time, and enhances data integrity. If designed and developed properly, an Army installation data warehouse has the potential to improve efficiencies and produce a positive return-on-investment.

Building a Data Warehouse for Decision Support

Build Information System Pyramid

Building with Speed and Agility on Microsoft's Cloud Platform

Building a Scalable Data Warehouse with Data Vault 2.0

DATA WAREHOUSING MOBILE CODE D

From Architecture to Implementation

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

Unlike popular belief, Data Warehouse is not a single tool but a collection of software tools. A data warehouse will collect data from diverse sources into a single database. Using Business Intelligence tools, meaningful insights are drawn from this data. The best thing about "Learn Data Warehousing in 1 Day" is that it is small and can be completed in a day. With this e-book, you will be enough knowledge to contribute and participate in a Data warehouse implementation project. The book covers upcoming and promising technologies like Data Lakes, Data Mart, ELT (Extract Load Transform) amongst others. Following are detailed topics included in the book Table Of Content Chapter 1: What Is Data Warehouse? 1. What Is Data Warehouse? 2. Types of Data Warehouse 3. Who needs Data warehouse? 4. Why We Need Data warehouse? 5. Data Warehouse Tools Chapter 2: Data Warehouse Architecture 1. Characteristics of Data warehouse 2. Data Warehouse Architectures 3. Data Warehouse Components 4. Query Tools Chapter 3: ETL Process 1. What is ETL? 2. Why do you need ETL? 3. ETL tools Chapter 4: ETL Vs ELT 1. What is ETL? 2. Difference between ETL vs. ELT Chapter 5: Data Modeling 1. What is Data Modelling? 2. Types of Data Models 3. Characteristics of a physical data model Chapter 6: OLAP 1. What is Online Analytical Processing? 2. Types of OLAP systems 3. Advantages and Disadvantages of OLAP Chapter 7: Multidimensional Olap (MOLAP) 1. What is MOLAP? 2. MOLAP Architecture 3. MOLAP Tools Chapter 8: OLAP Vs OLTP 1. What is the meaning of OLAP? 2. What is the meaning of OLTP? 3. Difference between OLTP and OLAP Chapter 9: Dimensional Modelling 1. What is Dimensional Model? 2. Elements of Dimensional Data Model 3. Attributes 4. Difference between Dimension table vs. Fact table 5. Steps of Dimensional Modelling 6. Rules for Dimensional Modelling Chapter 10: Star and Snowflake Schema 1. What is Multidimensional schemas? 2. What is a Star Schema? 3. What is a Snowflake Schema? 4. Difference between Star Schema and Snowflake Chapter 11: Data Mart 1. What is Data Mart? 2. Type of Data Mart 3. Steps in Implementing a Datamart Chapter 12: Data Mart Vs Data Warehouse 1. What is Data Warehouse? 2. What is Data Mart? 3. Differences between a Data Warehouse and a Data Mart Chapter 13: Data Lake 1. What is Data Lake? 2. Data Lake Architecture 3. Key Data Lake Concepts 4. Maturity stages of Data Lake Chapter 14: Data Lake Vs Data Warehouse 1. What is Data Warehouse? 2. What is Data Lake? 3. Key Difference between the Data Lake and Data Warehouse Chapter 15: What Is Business Intelligence? 1. What is Business Intelligence 2. Why is BI important? 3. How Business Intelligence systems are implemented? 4. Four types of BI users Chapter 16: Data Mining 1. What is Data Mining? 2. Types of Data 3. Data Mining Process 4. Modelling 5. Data Mining Techniques Chapter 17: Data Warehousing Vs Data Mining 1. What is Data Mining? 2. What Is Data Warehouse? 3. Difference between Data mining and Data Warehousing?

Develop modern solutions with Snowflake's unique architecture and integration capabilities: process bulk and real-time data into a data lake; and leverage time travel, cloning, and data-sharing features to optimize data operations Key FeaturesBuild and scale modern data solutions using the all-in-one Snowflake platformPerform advanced cloud analytics for implementing big data and data science solutionsMake quicker and better-informed business decisions by uncovering key insights from your dataBook Description Snowflake is a unique cloud-based data warehousing platform built from scratch to perform data management on the cloud. This book introduces you to Snowflake's unique architecture, which places it at the forefront of cloud data warehouses. You'll explore the compute model available with Snowflake, and find out how Snowflake allows extensive scaling through the virtual warehouses. You will then learn how to configure a virtual warehouse for optimizing cost and performance. Moving on, you'll get to grips with the data ecosystem and discover how Snowflake integrates with other technologies for staging and loading data. As you progress through the chapters, you will leverage Snowflake's capabilities to process a series of SQL statements using tasks to build data pipelines and find out how you can create modern data solutions and pipelines designed to provide high performance and scalability. You will also get to grips with creating role hierarchies, adding custom roles, and setting default roles for users before covering advanced topics such as data sharing, cloning, and performance optimization. By the end of this Snowflake book, you will be well-versed in Snowflake's architecture for building modern analytical solutions and understand best practices for solving commonly faced problems using practical recipes. What you will learnGet to grips with data warehousing techniques aligned with Snowflake's cloud architectureBroaden your skills as a data warehouse designer to cover the Snowflake ecosystemTransfer skills from on-premise data warehousing to the Snowflake cloud analytics platformOptimize performance and costs associated with a Snowflake solutionStage data on object stores and load it into SnowflakeSecure data and share it efficiently for accessManage transactions and extend Snowflake using stored proceduresExtend cloud data applications using Spark ConnectorWho this book is for This book is for data warehouse developers, data analysts, database administrators, and anyone involved in designing, implementing, and optimizing a Snowflake data warehouse. Knowledge of data warehousing and database and cloud concepts will be useful. Basic familiarity with Snowflake is beneficial, but not necessary. Organizations invest incredible amounts of time and money obtaining and then storing big data in data stores called data lakes. But how many of these organizations can actually get the data back out in a useable form? Very few can turn the data lake into an information gold mine. Most wind up with garbage dumps. Data Lake Architecture will explain how to build a useful data lake, where data scientists and data analysts can solve business challenges and identify new business opportunities. Learn how to structure data lakes as well as analog, application, and test-based data ponds to provide maximum business value. Understand the role of the row data pond and when to use an archival data pond. Leverage the four key ingredients for data lake success: metadata, integration mapping, context, and metaprocess. Bill Inmon opened our eyes to the architecture and benefits of a data warehouse, and now he takes us to the next level of data lake architecture.

Object-oriented Data Warehouse Design

Techniques for building modern cloud data warehousing solutions

Snowflake Cookbook

Data Warehousing

Super Charge Your Data Warehouse

A Comprehensive Guide for IT Professionals

The digital age has presented an exponential growth in the amount of data available to individuals looking to draw conclusions based on given or collected information across industries. Challenges associated with the analysis, security, sharing, storage, and visualization of large and complex data sets continue to plague data scientists and analysts alike as traditional data processing applications struggle to adequately manage big data. The Handbook of Research on Big Data Storage and Visualization Techniques is a critical scholarly resource that explores big data analytics and technologies and their role in developing a broad understanding of issues pertaining to the use of big data in multidisciplinary fields. Featuring coverage on a broad range of topics, such as architecture patterns, programing systems, and computational energy, this publication is geared towards professionals, researchers, and students seeking current research and application topics on the subject.

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

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 warehousesystems, the book covers new topics such as methods for handlingunstructured data in a data warehouse and storing data acrossmultiple storage media Discusses the pros and cons of relational versusmultidimensional design and how to measure return on investment

inplanning data warehouse projects Covers advanced topics, including data monitoring andtesting Although the book includes an extra 100 pages worth of valuablecontent, the price has actually been reduced from \$65 to \$55

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.

Data Warehousing for E-Business

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

The Data Warehouse Lifecycle Toolkit

Invaluable Data Modeling Rules to Implement Your Data Vault

Data Warehouse

Clickstream Data Warehousing

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 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 appendices 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 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 requirements

Over the past 5 years, the concept of big data has matured, data science has grown exponentially, and data architecture has become a standard part of organizational decision-making. Throughout all this change, the basic principles that shape the architecture of data have remained the same. There remains a need for people to take a look at the "bigger picture" and to understand where their data fit into the grand scheme of things. Data Architecture: A Primer explains this larger picture into which big data fits, giving data scientists the necessary context for how pieces of the puzzle should fit together. New case studies include expanded coverage of textual management and analytics New chapters on visualization and big data Discussion of new visualizations of the end-state architecture

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 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, 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 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

Market_Desc: · IT, Database, and Data Warehouse Managers and Developers Special Features: · Building the Data Warehouse has sold nearly 40,000 copies in its first 3 editions. Inmon is widely recognized as the Father of the Data Warehouse and remains one of the two leading authorities in the industry he helped to invent. The new edition covers new approaches and technologies, many of which have been pioneered by Inmon himself. Price of this new edition book provides a high-level, conceptual overview of the major components of data warehouse systems, as well as the core approaches used to design and build data warehouses. Topics covered in this book are methods for handling unstructured data in a data warehouse, storing data across multiple storage media, the pros and cons of relational vs. multidimensional design, data monitoring and testing.

Architecture, Analysis, and Design

Designing the Data Lake and Avoiding the Garbage Dump

Fundamentals of Data Warehouses

A Guide for Solution Architects and Project Leaders

BUILDING THE DATA WAREHOUSE (4th Ed.)

With Examples in SQL Server

This dissertation, "Data Warehousing Mobile Code Design" by [REDACTED], Chun-lung, Cheung, was obtained from The University of Hong Kong (Pokfulam, Hong Kong) and is being sold pursuant to Creative Commons: Attribution 3.0 Hong Kong License. The content of this dissertation has not been altered in any way. We have altered the formatting in order to facilitate the ease of printing and reading of the dissertation. All rights not granted by the above license are retained by the author. Abstract: Abstract of thesis entitled Data Warehousing Mobile Code Design submitted by Cheung Chun Lung for the degree of Master of Philosophy at the University of Hong Kong in June 2001 Data warehousing is the process of extracting, integrating, transforming and loading data from source systems into a database called data warehouse. The data warehouse is then used by analytical systems. The advantages of data warehousing are 1) simplification of data collection and preparation, 2) more optimized database design, and 3) avoidance of resource contention. The architectural components of data warehousing are 1) source systems, 2) data warehouse, and 3) analytical systems. The interfacing subsystems are the data integration and query interfaces. The first interface provides means to transform and integrate data from different source systems into the data warehouse. The second interface provides access to warehouse data from analytical applications, and typically provides summarization of data. A new architectural framework, the Java Extensible Warehousing Interface (JEWI), is proposed in this thesis. This framework is based on the Extensible Warehousing Interface approach, which integrates the mobile code paradigm into the data warehousing environment, giving extra features to enhance the existing data warehousing architecture. This approach provides a clean and extensible interface for new features to be "plugged into" the system. This framework plays a role in simplifying the extension of data warehousing functions at runtime. This in turn simplifies the development and maintenance of transformation and summarization algorithms in the data warehousing architecture. Moreover, the JEWI framework was implemented as a set of Java libraries. The libraries were used to build a prototype system to demonstrate our ideas and to evaluate the framework. DOI: 10.5353/th_b2987299 Subjects: Data warehousing

Data Modeling for Agile Data Warehouse using Data Vault Modeling Approach. Includes Enterprise Data Warehouse Architecture. This is a complete guide to the data vault data modeling approach. The book also includes business and program considerations for the agile data warehousing and business intelligence program. There are over 200 diagrams and figures concerning modeling, core business concepts, architecture, business alignment, semantics, and modeling comparisons with 3NF and Dimensional modeling.

This is an introductory guide to the techniques of Data warehousing and business intelligence. Centered on modeling, this devotional book explores the fundamentals of Data warehouse architectures. Using the anatomy analogy and snowflake topology, Taiwei is able to clearly explain multi-layered data warehouse architecture modeling, star/snowflake schema, dynamic ETL, cube design, and recommended approaches in the data warehouse ecology. It is

suitable for database engineers and developers, college students as well as IT managers and professional data architects.