Data Virtualization For Business Intelligence Systems Revolutionizing Data Integration For Data Warehouses The Morgan Kaufmann Series On Business Intelligence

Open Source BI solutions have many advantages over traditional proprietary software, from offering lower initial costs to more flexible support and integration options; but, until now, there has been no comprehensive guide to the complete offerings of the OS BI market. Writing for IT managers and business analysts without bias toward any BI suite, industry insider Lyndsay Wise covers the benefits and challenges of all available open source BI systems and tools, enabling readers to identify the solutions and technologies that best meet their business needs. Wise compares and contrasts types of OS BI and proprietary tools on the market, including Pentaho, Jaspersoft, RapidMiner, SpagoBI, BIRT, and many more. Real-world case studies and project templates clarify the steps involved in implementing open source BI, saving new users the time and trouble of developing their own solutions from scratch. For business managers who are hard pressed to indentify the best BI solutions and software for their companies, this book provides a practical guide to evaluating the ROI of open source versus traditional BI deployments. The only book to provide complete coverage of all open source BI systems and tools specifically for business managers, without bias toward any OS BI suite A practical, step-by-step guide to implementing OS BI solutions that maximize ROI Comprehensive coverage of all open source systems and tools, including architectures, data integration, support, optimization, data mining, data warehousing, and interoperability Case studies and project templates enable readers to evaluate the benefits and tradeoffs of all OS BI options without having to spend time developing their own solutions from scratch

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"

Run queries and analysis on big data clusters across relational and non relational databases KEY FEATURES ? Connect to Hadoop, Azure, Spark, Oracle, Teradata, Cassandra, MongoDB, CosmosDB, MySQL, PostgreSQL, MariaDB, and SAP HANA. ? Numerous techniques on how to query data and troubleshoot Polybase for better data analytics. ? Exclusive coverage on Azure Synapse Analytics and building Big Data clusters.

DESCRIPTION This book brings exciting coverage on establishing and managing data virtualization using polybase. This book teaches how to configure polybase on almost all relational and nonrelational databases. You will learn to set up the test environment for any tool or software instantly without hassle. You will practice how to design and build some of the high performing data warehousing solutions and that too in a few minutes of time. You will almost become an expert in connecting to all databases including hadoop, cassandra, MySQL, PostgreSQL, MariaDB and Oracle database. This book also brings exclusive coverage on how to build data clusters on Azure and using Azure Synapse Analytics. By the end of this book, you just don't administer the polybase for managing big data clusters but rather you learn to optimize and boost the performance for enabling data analytics and ease of data accessibility. WHAT YOU WILL LEARN ? Learn to configure Polybase and process Transact SQL queries with ease. ? Create a Docker container with SQL Server 2019 on Windows and Polybase. ? Establish SQL Server instance with any other software or tool using Polybase ? Connect with Cassandra, MongoDB, MySQL, PostgreSQL, MariaDB, and IBM DB2. WHO THIS BOOK IS FOR This book is for database developers and administrators familiar with the SQL language and command prompt. Managers and decision-makers will also find this book useful. No prior knowledge of any other technology or language is required. TABLE OF CONTENTS 1. What is Data Virtualization (Polybase) 2. History of Polybase 3. Polybase current state 4. Differences with other technologies 5. Usage 6. Future 7. SQL Server 8. Hadoop Cloudera and Hortonworks 9. Windows Azure Storage Blob 10. Spark 11. From Azure Synapse Analytics 12. From Big Data Clusters 13. Oracle 14. Teradata 15. Cassandra 16. MongoDB 17. CosmosDB 18. MySQL 19. PostgreSQL 20. MariaDB 21. SAP HANA 22. IBM DB2 23. Excel

The overall objective of this book is to show that data management is an exciting and valuable capability that is worth time and effort. More specifically it aims to achieve the following goals: 1. To give a “gentle” introduction to the field of DM by explaining and illustrating its core concepts, based on a mix of theory, practical frameworks such as TOGAF, ArchiMate, and DMBOK, as well as results from real-world assignments. 2. To offer guidance on how to build an effective DM capability in an organization.This is illustrated by various use cases, linked to the previously mentioned theoretical exploration as well as the stories of practitioners in the field. The primary target groups are: busy professionals who “are actively involved with managing data”. The book is also aimed at (Bachelor’s/ Master’s) students with an interest in data management. The book is industry-agnostic and should be applicable in different industries such as government, finance, telecommunications etc. Typical roles for which this book is intended: data governance office/ council, data owners, data stewards, people involved with data governance (data governance board), enterprise architects, data architects, process managers, business analysts and IT analysts. The book is divided into three main parts: theory, practice, and closing remarks. Furthermore, the chapters are as short and to the point as possible and also make a clear distinction between the main text and the examples. If the reader is already familiar with the topic of a chapter, he/she can easily skip it and move on to the next.

InfoWorld is targeted to Senior IT professionals. Content is segmented into Channels and Topic Centers. InfoWorld also celebrates people, companies, and projects.

The book ‘Data Intensive Computing Applications for Big Data’ discusses the technical concepts of big data, data intensive computing through machine learning, soft computing and parallel computing paradigms. It brings together researchers to report their latest results or progress in the development of the above mentioned areas. Since there are few books on this specific subject, the editors aim to provide a common platform for researchers working in this area to exhibit their novel findings. The book is intended as a reference work for advanced undergraduates and graduate students, as well as multidisciplinary, interdisciplinary and transdisciplinary research workers and scientists on the subjects of big data and cloud/parallel

and distributed computing, and explains didactically many of the core concepts of these approaches for practical applications. It is organized into 24 chapters providing a comprehensive overview of big data analysis using parallel computing and addresses the complete data science workflow in the cloud, as well as dealing with privacy issues and the challenges faced in a data-intensive cloud computing environment. The book explores both fundamental and high-level concepts, and will serve as a manual for those in the industry, while also helping beginners to understand the basic and advanced aspects of big data and cloud computing.

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 & BI implementation, its comprehensive coverage spans from basic concepts all the way through to customization. Highlighting the key differences between open source and vendor DW and BI technologies, the book identifies end-to-end solutions that are scalable, high performance, and stable. It illustrates the practical aspects of implementing and using open source DW and BI technologies to supply you with valuable on-the-project experience that can help you improve implementation and productivity. Emphasizing analysis, design, and programming, the text explains best-fit solutions as well as how to maximize ROI. Coverage includes data warehouse design, real-time processing, data integration, presentation services, and real-time reporting. With a focus on real-world applications, the author devotes an entire section to powerful implementation best practices that can help you build customer confidence while saving valuable time, effort, and resources.

Windows Azure is the Microsoft's cloud platform: a growing collection of complimentary services-compute, storage, data, networking, and app-that help you move faster, do more, and save money. But that's just scratching the surface. Azure is the only major cloud platform ranked by Gartner as an industry leader for both infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS). This powerful combination of managed and unmanaged services lets you build, deploy, and manage applications any way you like for unmatched productivity. Some cloud providers make you choose between your datacenter and the public cloud. Not Azure. Its enterprise-proven hybrid cloud solutions give you the best of both worlds, expanding your IT options without added complexity. With Azure, data storage, backup, and recovery become more efficient and economical. It's also easier to build applications that span both on-premises and the cloud. You'll share the same enterprise-tested platform that powers Skype, Office 365, Bing, and Xbox. Azure offers a 99.95% availability SLA, 24x7 tech support, and round-the-clock service health monitoring. That's why more than 57% of Fortune 500 companies rely on Azure today. Azure can quickly scale up or down to match demand, so you only pay for what you use. Per-minute billing and a commitment to match competitor prices for popular infrastructure services like compute, storage and bandwidth means you're always getting unbeatable price for performance. Azure runs on a growing global network of Microsoft-managed datacenters across 13 regions, giving you a wide range of options for running applications and ensuring your customers always get great performance. Azure is the first multinational cloud provider in mainland China and is continuing to expand to new regions around the globe. With Windows Azure is possible: Provision Windows and Linux Virtual Machines and applications in minutes. Use the same virtual machines and management tools in Azure that you use on-premises. Build and deploy a wide variety of modern applications for Android, iOS, and Windows that take full advantage of the cloud-including web, mobile, media and line-of-business solutions. Automatically scale up and down to meet any need. Azure provides managed SQL and NoSQL data services and built-in support for gaining insights from your data. Leverage the full power of SQL Server in the cloud and use HDInsight to build

Hadoop clusters to analyze data. Manage user accounts, synchronize with existing on-premises directories, and get single sign on across Azure, Office 365 and hundreds of popular software-as-a-service applications including Salesforce, DocuSign, Google Apps, Box, Dropbox, and more.

There has never been a Master Data Management Guide like this. It contains 24 answers, much more than you can imagine; comprehensive answers and extensive details and references, with insights that have never before been offered in print. Get the information you need--fast! This all-embracing guide offers a thorough view of key knowledge and detailed insight. This Guide introduces what you want to know about Master Data Management. A quick look inside of some of the subjects covered: Teradata - History, Master data management Issues, Data virtualization - Topics, Data steward, Business intelligence Amount and quality of available data, Identity resolution, Microsoft SQL Server - SQL Server 2008 R2, Master data - Master Data Defined, Identity resolution - Master data management, Entity-attribute-value model - History of EAV database systems, Fact (data warehouse) - Hybrid design, Single customer view, Data profiling, MIKE2.0 Methodology, Business intelligence Business intelligence and data warehousing, Enterprise software, Master data management Solutions, Customer data integration - History of customer data integration, Software AG - History, Data governance Data governance conferences, Information technology management IT managers, and much more...

Use this guide to one of SQL Server 2019's latest and most impactful features—Big Data Clusters—that combines large volumes of non-relational data for analysis along with data stored relationally inside a SQL Server database. Big Data Clusters is a feature set covering data virtualization, distributed computing, and relational databases and provides a complete AI platform across the entire cluster environment. This book shows you how to deploy, manage, and use Big Data Clusters. For example, you will learn how to combine data stored on the HDFS file system together with data stored inside the SQL Server instances that make up the Big Data Cluster. Filled with clear examples and use cases, SQL Server Big Data Clusters Revealed provides everything necessary to get started working with SQL Server 2019 Big Data Clusters. You will learn about the architectural foundations that are made up from Kubernetes, Spark, HDFS, and SQL Server on Linux. You then are shown how to configure and deploy Big Data Clusters in on-premises environments or in the cloud. Next, you are taught about querying. You will learn to write queries in Transact-SQL—taking advantage of skills you have honed for years—and with those queries you will be able to examine and analyze data from a wide variety of sources such as Apache Spark. Through the theoretical foundation provided in this book and easy-to-follow example scripts and notebooks, you will be ready to use and unveil the full potential of SQL Server 2019: combining different types of data spread across widely disparate sources into a single view that is useful for business intelligence

and machine learning analysis. What You Will Learn Install, manage, and troubleshoot Big Data Clusters in cloud or on-premise environments Analyze large volumes of data directly from SQL Server and/or Apache Spark Manage data stored in HDFS from SQL Server as if it were relational data Implement advanced analytics solutions through machine learning and AI Expose different data sources as a single logical source using data virtualization Who This Book Is For For data engineers, data scientists, data architects, and database administrators who want to employ data virtualization and big data analytics in their environment

Use this guide to one of SQL Server 2019's most impactful features—Big Data Clusters. You will learn about data virtualization and data lakes for this complete artificial intelligence (AI) and machine learning (ML) platform within the SQL Server database engine. You will know how to use Big Data Clusters to combine large volumes of streaming data for analysis along with data stored in a traditional database. For example, you can stream large volumes of data from Apache Spark in real time while executing Transact-SQL queries to bring in relevant additional data from your corporate, SQL Server database. Filled with clear examples and use cases, this book provides everything necessary to get started working with Big Data Clusters in SQL Server 2019. You will learn about the architectural foundations that are made up from Kubernetes, Spark, HDFS, and SQL Server on Linux. You then are shown how to configure and deploy Big Data Clusters in on-premises environments or in the cloud. Next, you are taught about querying. You will learn to write queries in Transact-SQL—taking advantage of skills you have honed for years—and with those queries you will be able to examine and analyze data from a wide variety of sources such as Apache Spark. Through the theoretical foundation provided in this book and easy-to-follow example scripts and notebooks, you will be ready to use and unveil the full potential of SQL Server 2019: combining different types of data spread across widely disparate sources into a single view that is useful for business intelligence and machine learning analysis. What You Will Learn Install, manage, and troubleshoot Big Data Clusters in cloud or on-premise environments Analyze large volumes of data directly from SQL Server and/or Apache Spark Manage data stored in HDFS from SQL Server as if it were relational data Implement advanced analytics solutions through machine learning and AI Expose different data sources as a single logical source using data virtualization Who This Book Is For Data engineers, data scientists, data architects, and database administrators who want to employ data virtualization and big data analytics in their environments

Information Management: Gaining a Competitive Advantage with Data is about making smart decisions to make the most of company information. Expert author William McKnight develops the value proposition for information in the enterprise and succinctly outlines the numerous forms of data storage. Information Management will enlighten you, challenge your preconceived notions, and help

activate information in the enterprise. Get the big picture on managing data so that your team can make smart decisions by understanding how everything from workload allocation to data stores fits together. The practical, hands-on guidance in this book includes: Part 1: The importance of information management and analytics to business, and how data warehouses are used Part 2: The technologies and data that advance an organization, and extend data warehouses and related functionality Part 3: Big Data and NoSQL, and how technologies like Hadoop enable management of new forms of data Part 4: Pulls it all together, while addressing topics of agile development, modern business intelligence, and organizational change management Read the book cover-to-cover, or keep it within reach for a quick and useful resource. Either way, this book will enable you to master all of the possibilities for data or the broadest view across the enterprise. Balances business and technology, with non-product-specific technical detail Shows how to leverage data to deliver ROI for a business Engaging and approachable, with practical advice on the pros and cons of each domain, so that you learn how information fits together into a complete architecture Provides a path for the data warehouse professional into the new normal of heterogeneity, including NoSQL solutions

Artificial intelligence (AI) describes machines/computers that mimic cognitive functions that humans associate with other human minds, such as learning and problem solving. As businesses have evolved to include more automation of processes, it has become more vital to understand AI and its various applications. Additionally, it is important for workers in the marketing industry to understand how to coincide with and utilize these techniques to enhance and make their work more efficient. The Handbook of Research on Applied AI for International Business and Marketing Applications is a critical scholarly publication that provides comprehensive research on artificial intelligence applications within the context of international business. Highlighting a wide range of topics such as diversification, risk management, and artificial intelligence, this book is ideal for marketers, business professionals, academicians, practitioners, researchers, and students.

Harness the power of PolyBase data virtualization software to make data from a variety of sources easily accessible through SQL queries while using the T-SQL skills you already know and have mastered. PolyBase Revealed shows you how to use the PolyBase feature of SQL Server 2019 to integrate SQL Server with Azure Blob Storage, Apache Hadoop, other SQL Server instances, Oracle, Cosmos DB, Apache Spark, and more. You will learn how PolyBase can help you reduce storage and other costs by avoiding the need for ETL processes that duplicate data in order to make it accessible from one source. PolyBase makes SQL Server into that one source, and T-SQL is your golden ticket. The book also covers PolyBase scale-out clusters, allowing you to distribute PolyBase queries among several SQL Server instances, thus improving performance. With great flexibility comes great complexity, and this book shows you where to look when

queries fail, complete with coverage of internals, troubleshooting techniques, and where to find more information on obscure cross-platform errors. Data virtualization is a key target for Microsoft with SQL Server 2019. This book will help you keep your skills current, remain relevant, and build new business and career opportunities around Microsoft's product direction. What You Will Learn Install and configure PolyBase as a stand-alone service, or unlock its capabilities with a scale-out cluster Understand how PolyBase interacts with outside data sources while presenting their data as regular SQL Server tables Write queries combining data from SQL Server, Apache Hadoop, Oracle, Cosmos DB, Apache Spark, and more Troubleshoot PolyBase queries using SQL Server Dynamic Management Views Tune PolyBase queries using statistics and execution plans Solve common business problems, including "cold storage" of infrequently accessed data and simplifying ETL jobs Who This Book Is For SQL Server developers working in multi-platform environments who want one easy way of communicating with, and collecting data from, all of these sources

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

This book constitutes the thoroughly refereed conference proceedings of the BIRTE workshops listed below, which were held in in conjunction with VLDB, the International Conference on Very Large Data Bases: 9th International Workshop on Business Intelligence for the Real-Time Enterprise, BIRTE 2015, held in Kohala Coast, Hawaii, in August 2015, 10th International Workshop on Enabling Real-Time Business Intelligence, BIRTE 2016, held in New Delhi, India, in September 2016, 11th International Workshop on Real-Time Business

Intelligence and Analytics, BIRTE 2017, held in Munich, Germany, in August 2017. The BIRTE workshop series provides a forum for the discussion and advancement of the science and engineering enabling real-time business intelligence and the novel applications that build on these foundational techniques. The book includes five selected papers from BIRTE 2015; five selected papers from BIRTE 2016; and three selected papers from BIRTE 2017.

Big data is certainly one of the biggest buzz phrases in IT today. Combined with virtualization and cloud computing, big data is a technological capability that will force data centers to significantly transform and evolve within the next five years. Similar to virtualization, big data infrastructure is unique and can create an architectural upheaval in the way systems, storage, and software infrastructure are connected and managed. Unlike previous business analytics solutions, the real-time capability of new big data solutions can provide mission critical business intelligence that can change the shape and speed of enterprise decision making forever. Hence, the way in which IT infrastructure is connected and distributed warrants a fresh and critical analysis.

Following the footsteps of the first edition, the second edition of Business Intelligence is a full overview of what comprises business intelligence. It is intended to provide an introduction to the concepts to uncomplicate the learning process when implementing a business intelligence program. Over a relatively long lifetime (7 years), the current edition of book has received numerous accolades from across the industry for its straightforward introduction to both business and technical aspects of business intelligence. As an author, David Loshin has a distinct ability to translate challenging topics into a framework that is easily digestible by managers, business analysts, and technologists alike. In addition, his material has developed a following (such as the recent Master Data Management book) among practitioners and key figures in the industry (both analysts and vendors) and that magnifies our ability to convey the value of this book. Guides managers through developing, administering, or simply understanding business intelligence technology Keeps pace with the changes in best practices, tools, methods and processes used to transform an organization's data into actionable knowledge Contains a handy, quick-reference to technologies and terminology.

Big Data Imperatives, focuses on resolving the key questions on everyone’s mind: Which data matters? Do you have enough data volume to justify the usage? How you want to process this amount of data? How long do you really need to keep it active for your analysis, marketing, and BI applications? Big data is emerging from the realm of one-off projects to mainstream business adoption; however, the real value of big data is not in the overwhelming size of it, but more in its effective use. This book addresses the following big data characteristics: Very large, distributed aggregations of loosely structured data – often incomplete and inaccessible Petabytes/Exabytes of data Millions/billions of people providing/contributing to the context behind the data Flat schema's with few complex interrelationships Involves time-stamped events Made up of incomplete data Includes connections between data elements that must be probabilistically inferred Big Data Imperatives explains 'what big data can do'. It can batch process millions and billions of records both unstructured and structured much faster and cheaper. Big data analytics provide a platform to merge all analysis which enables data analysis to be more accurate, well-rounded, reliable and focused on a

specific business capability. Big Data Imperatives describes the complementary nature of traditional data warehouses and big-data analytics platforms and how they feed each other. This book aims to bring the big data and analytics realms together with a greater focus on architectures that leverage the scale and power of big data and the ability to integrate and apply analytics principles to data which earlier was not accessible. This book can also be used as a handbook for practitioners; helping them on methodology,technical architecture, analytics techniques and best practices. At the same time, this book intends to hold the interest of those new to big data and analytics by giving them a deep insight into the realm of big data.

Information modelling and knowledge bases have become ever more essential in recent years because of the need to handle and process the vast amounts of data which now form part of everyday life. The machine to machine communication of the Internet of Things (IoT), in particular, can generate unexpectedly large amounts of raw data. This book presents the proceedings of the 27th International Conference on Information Modelling and Knowledge Bases (EJC2017), held in Krabi, Thailand, in June 2017. The EJC conferences originally began in 1982 as a co-operative initiative between Japan and Finland, but have since become a world-wide research forum bringing together researchers and practitioners in information modelling and knowledge bases for the exchange of scientific results and achievements. Of the 42 papers submitted, 29 were selected for publication here, and these cover a wide range of information-modelling topics, including the theory of concepts, semantic computing, data mining, context-based information retrieval, ontological technology, image databases, temporal and spatial databases, document data management, software engineering, cross-cultural computing, environmental analysis, social networks, and WWW information. The book will be of interest to all those whose work involves dealing with large amounts of data.

With the benefit of advanced analytics such as online analytical processing (OLAP), data mining, and text analytics, the IBM® InfoSphere® Warehouse Enterprise Edition brings sophisticated business intelligence (BI) to warehouse users. InfoSphere Warehouse allows you to run extreme concurrent query volumes that can help answer questions for all types of business users, while consistently meeting service level requirements. Combined with a virtualization platform and a solid BI solution, such as IBM Cognos®, you can deliver BI cloud services with improved flexibility and speed to your clients, thereby presenting a new avenue for which your services can be offered. This IBM Redbooks® publication discusses the deployment of a BI cloud solution. It includes details such as understanding the architecture of a cloud, planning implementation, integrating various software components, and understanding the preferred practices of running a cloud deployment. Essentially, this book can be used as a guide by anyone who is interested in deploying a virtualized environment for a BI cloud solution.

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. Big Data: Concepts, Methodologies, Tools, and Applications is a multi-

volume compendium of research-based perspectives and solutions within the realm of large-scale and complex data sets. Taking a multidisciplinary approach, this publication presents exhaustive coverage of crucial topics in the field of big data including diverse applications, storage solutions, analysis techniques, and methods for searching and transferring large data sets, in addition to security issues. Emphasizing essential research in the field of data science, this publication is an ideal reference source for data analysts, IT professionals, researchers, and academics.

In light of the rising cost of healthcare and the overall challenges associated with delivering quality care to patients across regions, scientists and pharmacists are exploring new initiatives in drug discovery and design. One such initiative is the adoption of information technology and software applications to improve healthcare and pharmaceutical processes. Software Innovations in Clinical Drug Development and Safety is a comprehensive resource analyzing the integration of software engineering for the purpose of drug discovery, clinical trials, genomics, and drug safety testing. Taking a multi-faceted approach to the application of computational methods to pharmaceutical science, this publication is ideal for healthcare professionals, pharmacists, computer scientists, researchers, and students seeking the latest information on the architecture and design of software in clinical settings, the impact of clinical technologies on business models, and the safety and privacy of patients and patient data. This timely resource features a well-rounded discussion on topics pertaining to the integration of computational methods in pharmaceutical science and practice including, the impact of software integration on business models, patient safety concerns, software architecture and design, and data security.

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

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 &

Get the end-to-end instruction you need to design, develop, and deploy more effective data integration, reporting, and analysis solutions using SQL Server 2008—whether you're new to business intelligence (BI) programming or a seasoned pro. With real-world examples and insights from an expert team, you'll master the concepts, tools, and techniques for building solutions that deliver intelligence—and business value—exactly where users want it. Discover how to: Manage the development life cycle and build a BI team Dig into SQL Server Analysis Services, Integration Services, and Reporting Services Navigate the Business Intelligence Development Studio (BIDS) Write queries that rank, sort, and drill down on sales data Develop extract, transform, and load (ETL) solutions Add a source code control system Help secure packages for deployment via encryption and credentials Use MDX and DMX Query Designers to build reports based on OLAP cubes and data mining models Create and implement custom objects using .NET code View reports in Microsoft Office Excel and Office SharePoint Serverook

Get a head-start on learning one of SQL Server 2019's latest and most impactful features-Big Data Clusters-that combines large volumes of non-relational data for analysis along with data stored relationally inside a SQL Server database. This book provides a first look at Big Data Clusters based upon SQL Server 2019 Release Candidate 1. Start now and get a jump on your competition in learning this important new feature. Big Data Clusters is a feature set covering data virtualization, distributed computing, and relational databases and provides a complete AI platform across the entire cluster environment. This book shows you how to deploy, manage, and use Big Data Clusters. For example, you will learn how to combine data stored on the HDFS file system together with data stored inside the SQL Server instances that make up the Big Data Cluster. Filled with clear examples and use cases, this book provides everything necessary to get started working with Big Data Clusters in SQL Server 2019 using Release Candidate 1. You will learn about the architectural foundations that are made up from Kubernetes, Spark, HDFS, and SQL Server on Linux. You then are shown how to configure and deploy Big Data Clusters in on-premises environments or in the cloud. Next, you are taught about querying. You will learn to write queries in Transact-SQL-taking advantage of skills you have honed for years-and with those queries you will be able to examine and analyze data from a wide variety of sources such as Apache Spark. Through the theoretical foundation provided in this book and easy-to-follow example scripts and notebooks, you will be ready to use and unveil the full potential of SQL Server 2019: combining different types of data spread across widely disparate sources into a single view that is useful for business intelligence and machine learning analysis. What You Will Learn Install, manage, and troubleshoot Big Data Clusters in cloud or on-premise environments Analyze large volumes of data directly from SQL Server and/or

Apache Spark Manage data stored in HDFS from SQL Server as if it were relational data Implement advanced analytics solutions through machine learning and AI Expose different data sources as a single logical source using data virtualization Who This Book Is For For data engineers, data scientists, data architects, and database administrators who want to employ data virtualization and big data analytics in their environment.

This book contains practical steps business users can take to implement data management in a number of ways, including data governance, data architecture, master data management, business intelligence, and others. It defines data strategy, and covers chapters that illustrate how to align a data strategy with the business strategy, a discussion on valuing data as an asset, the evolution of data management, and who should oversee a data strategy. This provides the user with a good understanding of what a data strategy is and its limits. Critical to a data strategy is the incorporation of one or more data management domains. Chapters on key data management domains—data governance, data architecture, master data management and analytics, offer the user a practical approach to data management execution within a data strategy. The intent is to enable the user to identify how execution on one or more data management domains can help solve business issues. This book is intended for business users who work with data, who need to manage one or more aspects of the organization’s data, and who want to foster an integrated approach for how enterprise data is managed. This book is also an excellent reference for students studying computer science and business management or simply for someone who has been tasked with starting or improving existing data management.

This book presents a detailed review of high-performance computing infrastructures for next-generation big data and fast data analytics. Features: includes case studies and learning activities throughout the book and self-study exercises in every chapter; presents detailed case studies on social media analytics for intelligent businesses and on big data analytics (BDA) in the healthcare sector; describes the network infrastructure requirements for effective transfer of big data, and the storage infrastructure requirements of applications which generate big data; examines real-time analytics solutions; introduces in-database processing and in-memory analytics techniques for data mining; discusses the use of mainframes for handling real-time big data and the latest types of data management systems for BDA; provides information on the use of cluster, grid and cloud computing systems for BDA; reviews the peer-to-peer techniques and tools and the common information visualization techniques, used in BDA.

There has never been a Data Integration Guide like this. Data Integration 25 Success Secrets is not about the ins and outs of Data Integration. Instead, it answers the top 25 questions that we are asked and those we come across in our forums, consultancy and education programs. It tells you exactly how to deal with those questions, with tips that have never before been offered in print. Get

the information you need--fast! This comprehensive guide offers a thorough view of key knowledge and detailed insight. This Guide introduces everything you want to know to be successful with Data Integration. A quick look inside of the subjects covered: Global Business Intelligence Can Help Businesses and Companies, The New SQL Server 2005 DTS: Introducing the Server Integration Services as the New Transformation, What are some of the specific components of PaaS?, Integrating SOA Into The Mainframe, Uses of Business Intelligence Tools, The Advantages of Attending Enterprise Architecture Conferences, Data Modeling Techniques, What are Business Process Management Tools, Benefits of MDM, The MCITP Developer: Microsoft SQL Server 2005 Exam 70-441, Data Warehouse Metadata and Its Benefits, The MDM strategy comprises different elements., Jumper 2.0 i, Need a Tool? Try Business Intelligence Software, Updates on Data Virtualization, Terminology, Platform as a Service (PaaS), MDM Component Layer Model, Special Features, MDM Myths, IT Metadata Management Products: History and Uses, Leading to Business Intelligence Best Practices, Rights Metadata, The Importance of Creating Metadata Registries, There are six distinct levels of MDM maturity., and much more...

Business Intelligence is an essential tool used by enterprises for strategic, tactical and operational decision making. Business Intelligence most often needs to correlate data from disparate data sources to derive insights. Unifying data from disparate data sources and providing a unifying view of data is generally known as data integration. Traditionally enterprises employed ETL and data warehouses for data integration. However in last few years a technology known as "Data Virtualization" has found some acceptance as an alternative data integration solution. "Data Virtualization" is a federated database termed as composite database by McLeod/Heimbigner's in 1985. Till few years back Data Virtualization weren't considered as an alternative for ETL but was rather thought of as a technology for niche integration challenges. In this paper we hypothesize that for many BI applications "data virtualization" is a better cost effective data integration strategy. We analyze the system architecture of "Data warehouse" and "Data Virtualization" solutions. We further employ System Dynamics Model to compare few key metrics like "Time to Market" and "Cost of "Data warehouse" and "Data Virtualization" solutions. We also look at the impact of "Enterprise Data Standardization" on data integration.

Data Virtualization for Business Intelligence SystemsRevolutionizing Data Integration for Data WarehousesElsevier

Annotation In this book, Rick van der Lans explains how data virtualization servers work, what techniques to use to optimize access to various data sources and how these products can be applied in different projects.

Integrating Hadoop leverages the discipline of data integration and applies it to the Hadoop open-source software framework for storing data on clusters of commodity hardware. It is packed with the need-to-know for managers, architects, designers, and developers responsible for populating Hadoop in the enterprise, allowing you to

harness big data and do it in such a way that the solution: · Complies with (and even extends) enterprise standards · Integrates seamlessly with the existing information infrastructure · Fills a critical role within enterprise architecture. Integrating Hadoop covers the gamut of the setup, architecture and possibilities for Hadoop in the organization, including: · Supporting an enterprise information strategy · Organizing for a successful Hadoop rollout · Loading and extracting of data in Hadoop · Managing Hadoop data once it's in the cluster · Utilizing Spark, streaming data, and master data in Hadoop processes - examples are provided to reinforce concepts.

This book constitutes the refereed conference proceedings of the 14th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2015, held in Delft, The Netherlands, in October 2015. The 40 revised full papers presented together with 1 keynote panel were carefully reviewed and selected from 65 submissions. They are organized in the following topical sections: adoption; big and open data; e-business, e-services,, and e-society; and witness workshop.

This collection of different views on how digitalization is influencing various industrial sectors addresses essential topics like big data and analytics, fintech and insuretech, cloud and mobility technologies, disruption and entrepreneurship. The technological advances of the 21st century have been massively impacted by the digital upheaval: there is no future without digitalization. The sale of products and services has left the classical point of sale and now takes place on a variety of channels. Whether in the automotive industry, travel and traffic, in cities, or the financial industry – newly designed ecosystems are being created everywhere; data is being generated and analyzed in real time; and companies are competing for mobile access channels to customers in order to gain knowledge about their individual contexts and preferences. In turn, customers can now publicly share their opinions, experiences and knowledge as User Generated Content, allowing them to impact the market and empowering them to build or destroy trust.

Achieve best-in-class metrics and get more from your data with JMP JMP Connections is the small- and medium-sized business owner's guide to exceeding customer expectations by getting more out of your data using JMP. Uniquely bifunctional, this book is divided into two parts: the first half of the book shows you what JMP can do for you. You'll discover how to wring every last drop of insight out of your data, and let JMP parse reams of raw numbers into actionable insight that leads to better strategic decisions. You'll also discover why it works so well; clear explanations break down the Connectivity platform and metrics in business terms to demystify data analysis and JMP while giving you a macro view of the benefits that come from optimal implementation. The second half of the book is for your technical team, demonstrating how to implement specific solutions relating to data set development and data virtualization. In the end, your organization reduces Full Time Equivalents while increasing productivity and competitiveness. JMP is a powerful tool for business, but many organizations aren't even scratching the surface of what their data can do for them. This book provides the information and technical guidance your business needs to achieve more. Learn what a JMP Connectivity Platform can do for your business Understand Metrics-on-Demand, Real-Time Metrics, and their implementation Delve into technical implementation with information on configuration and management, version control, data visualization, and more Make better business decisions by getting more and better information from your

data Business leadership relies on good information to make good business decisions—but what if you could increase the quality of the information you receive, while getting more of what you want to know and less of what you don't need to know? How would that affect strategy, operations, customer experience, and other critical areas? JMP can help with that, and JMP Connections provides real, actionable guidance on getting more out of JMP.

Websites are a central part of today’s business world; however, with the vast amount of information that constantly changes and the frequency of required updates, this can come at a high cost to modern businesses. Web Data Mining and the Development of Knowledge-Based Decision Support Systems is a key reference source on decision support systems in view of end user accessibility and identifies methods for extraction and analysis of useful information from web documents. Featuring extensive coverage across a range of relevant perspectives and topics, such as semantic web, machine learning, and expert systems, this book is ideally designed for web developers, internet users, online application developers, researchers, and faculty.

Business intelligence (BI) software allows you to view different components of a business using a single visual platform, which makes comprehending mountains of data easier. BI is everywhere. Applications that include reports, analytics, statistics, and historical and predictive modeling are all examples of business intelligence. Currently, we are in the second generation of business intelligence software—called BI 2.0—which is focused on writing business intelligence software that is predictive, adaptive, simple, and interactive. As computers and software have evolved, more data can be presented to end users with increasingly visually rich techniques. Rich Internet application (RIA) technologies such as Microsoft Silverlight can be used to transform traditional user-interfaces filled with boring data into fully interactive analytical applications that quickly deliver insight from large data sets. Furthermore, RIAs now include 3D spatial-design capabilities that move beyond a simple list or grid and allow for interesting layouts of aggregated data. BI 2.0 implemented via an RIA technology can truly bring out the power of business intelligence and deliver it to an average user on the Web. Silverlight 4 Business Intelligence Software provides developers, designers, and architects with a solid foundation in business intelligence design and architecture concepts for Microsoft Silverlight. This book covers key business intelligence design concepts and how they can be applied without an existing BI infrastructure. Author Bart Czernicki provides you with examples of how to build small BI applications that are interactive, highly visual, statistical, predictive—and most importantly—intuitive to the end-user. Business intelligence isn’t just for the executive branch of a Fortune 500 company—it is for the masses. Let Silverlight 4 Business Intelligence Software show you how to unlock the rich intelligence you already have.

Clear your doubts about Business Intelligence and start your new journey KEY FEATURES ? Includes successful methods and innovative ideas to achieve success with BI. ? Vendor-neutral, unbiased, and based on experience. ? Highlights practical challenges in BI journeys. ? Covers financial aspects along with technical aspects. ? Showcases multiple BI organization models and the structure of BI teams. DESCRIPTION The book demystifies misconceptions and misinformation about BI. It provides clarity to almost everything related to BI in a simplified and unbiased way. It covers topics right from the definition of BI, terms used in the BI definition, coinage of

BI, details of the different main uses of BI, processes that support the main uses, side benefits, and the level of importance of BI, various types of BI based on various parameters, main phases in the BI journey and the challenges faced in each of the phases in the BI journey. It clarifies myths about self-service BI and real-time BI. The book covers the structure of a typical internal BI team, BI organizational models, and the main roles in BI. It also clarifies the doubts around roles in BI. It explores the different components that add to the cost of BI and explains how to calculate the total cost of the ownership of BI and ROI for BI. It covers several ideas, including unconventional ideas to achieve BI success and also learn about IBI. It explains the different types of BI architectures, commonly used technologies, tools, and concepts in BI and provides clarity about the boundary of BI w.r.t technologies, tools, and concepts. The book helps you lay a very strong foundation and provides the right perspective about BI. It enables you to start or restart your journey with BI. WHAT YOU WILL LEARN ? Builds a strong conceptual foundation in BI. ? Gives the right perspective and clarity on BI uses, challenges, and architectures. ? Enables you to make the right decisions on the BI structure, organization model, and budget. ? Explains which type of BI solution is required for your business. ? Applies successful BI ideas. WHO THIS BOOK IS FOR This book is a must-read for business managers, BI aspirants, CxOs, and all those who want to drive the business value with data-driven insights. TABLE OF CONTENTS 1. What is Business Intelligence? 2. Why do Businesses need BI? 3. Types of Business Intelligence 4. Challenges in Business Intelligence 5. Roles in Business Intelligence 6. Financials of Business Intelligence 7. Ideas for Success with BI 8. Introduction to IBI 9. BI Architectures 10. Demystify Tech, Tools, and Concepts in BI