

Teradata Jdbc Driver User Guide

Today, organizations face tremendous challenges with data explosion and information governance. InfoSphere™ Optim™ solutions solve the data growth problem at the source by managing the enterprise application data. The Optim Data Growth solutions are consistent, scalable solutions that include comprehensive capabilities for managing enterprise application data across applications, databases, operating systems, and hardware platforms. You can align the management of your enterprise application data with your business objectives to improve application service levels, lower costs, and mitigate risk. In this IBM® Redbooks® publication, we describe the IBM InfoSphere Optim Data Growth solutions and a methodology that provides implementation guidance from requirements analysis through deployment and administration planning. We also discuss various implementation topics including system architecture design, sizing, scalability, security, performance, and automation. This book is intended to provide various systems development professionals, Data Solution Architects, Data Administrators, Modelers, Data Analysts, Data Integrators, or anyone who has to analyze or integrate data structures, a broad understanding about IBM InfoSphere Optim Data Growth solutions. By being used in conjunction with the product manuals and online help, this book provides guidance about implementing an optimal solution for managing your enterprise application data.

"Before making data available in SAP HANA, you must standardize, integrate, and secure it--that's where data provisioning comes in. In this guide, you'll learn about each of your options, from SAP HANA-based tools like SDI and SDQ to SAP Data Services and SAP LT Replication Server. Whether you'll be provisioning data in batches or in real-time, you'll understand when to use each tool, its requirements, and how it works. A detailed case study will show you how to establish a successful data provisioning practice"--

Data management and analytics simplified with Teradata Key Features Take your understanding of Teradata to the next level and build efficient data warehousing applications for your organization Covers recipes on data handling, warehousing, advanced querying and the administrative tasks in Teradata. Contains practical solutions to tackle common (and not-so-common) problems you might encounter in your day to day activities Book Description Teradata is an enterprise software company that develops and sells its eponymous relational database management system (RDBMS), which is considered to be a leading data warehousing solutions and provides data management solutions for analytics. This book will help you get all the practical information you need for the creation and implementation of your data warehousing solution using Teradata. The book begins with recipes on quickly setting up a development environment so you can work with different types of data structuring and manipulation function. You will tackle all problems related to efficient querying, stored procedure searching, and navigation techniques. Additionally, you'll master various administrative tasks such as user and security management, workload management, high availability, performance tuning, and monitoring. This book is designed to take

you through the best practices of performing the real daily tasks of a Teradata DBA, and will help you tackle any problem you might encounter in the process. What you will learn Understand Teradata's competitive advantage over other RDBMSs. Use SQL to process data stored in Teradata tables. Leverage Teradata's available application utilities and parallelism to play with large datasets Apply various performance tuning techniques to optimize the queries. Acquire deeper knowledge and understanding of the Teradata Architecture. Easy steps to load, archive, restore data and implement Teradata protection features Gain confidence in running a wide variety of Data analytics and develop applications for the Teradata environment Who this book is for This book is for Database administrator's and Teradata users who are looking for a practical, one-stop resource to solve all their problems while handling their Teradata solution. If you are looking to learn the basic as well as the advanced tasks involved in Teradata querying or administration, this book will be handy. Some knowledge of relational database concepts will be helpful to get the best out of this book.

Work with petabyte-scale datasets while building a collaborative, agile workplace in the process. This practical book is the canonical reference to Google BigQuery, the query engine that lets you conduct interactive analysis of large datasets. BigQuery enables enterprises to efficiently store, query, ingest, and learn from their data in a convenient framework. With this book, you'll examine how to analyze data at scale to derive insights from large datasets efficiently. Valliappa Lakshmanan, tech lead for Google Cloud Platform, and Jordan Tigani, engineering director for the BigQuery team, provide best practices for modern data warehousing within an autoscaled, serverless public cloud. Whether you want to explore parts of BigQuery you're not familiar with or prefer to focus on specific tasks, this reference is indispensable.

A Guide for Business and Technology Managers

Storage and Analysis at Internet Scale

HDInsight Essentials - Second Edition

Teradata Utilities

A Practical Guide to Apache Kudu, Impala, and Spark

Oracle WebLogic Server 12c Advanced Administration Cookbook

As business cycles speed up, many customers gain significant competitive advantage from quicker and more accurate business decision-making by using real data. For many customers, choosing the path to co-locate their transactional and analytical workloads on System z® better leverages their existing investment in hardware, software, and skills. We created a project to address a number of best practice questions on how to manage these newer, analytical type workloads, especially when co-located with traditional transactional workloads. The goal of this IBM® Redbooks® publication is to provide technical guidance and performance trade-offs associated with resource management and potentially DB2® data-sharing in a variety of mixed transactional / data warehouse System z topologies. The term co-location used here and in the rest of the book is specifically defined as the practice of housing both transactional (OLTP) and data warehouse (analytical) workloads within the same System z

configuration. We also assumed that key portions of the transactional and data warehouse databases would reside on DB2 for z/OS®. The databases may or may not reside in a DB2 data-sharing environment; we discuss those pros and cons in this book. The intended audience includes DB2 data warehouse architects and practitioners who are facing choices in resource management and system topologies in the data warehouse arena. This specifically includes Business Intelligence (BI) administrators, DB2 database administrators (DBAs) and z/OS performance administrators / systems programmers. In addition, decision makers and architects can utilize this book to assist in making platform and database topology decisions. The book is divided into four parts. Part I, "Introducing the co-location project" covers the System z value proposition and why one should consider System z as the central platform for their data warehousing / business analytics needs. Some topics are risk avoidance via data consolidation, continuous availability, simplified disaster recovery, IBM Smart Analytics Optimizer, reduced network bandwidth requirements, and the unique virtualization and resource management capabilities of System z LPAR, z/VM® and WLM. Part I also provides some of the common System z co-location topologies along with an explanation of the general pros and cons of each. This would be useful input for an architect to understand where a customer is today and where they might consider moving to. Part II, "Project environment" covers the environment, products, workloads, workload drivers, and data models implemented for this study. The environment consisted of a logically partitioned z10™ 32way, running z/VM, Linux®, and z/OS operating system instances. On those instances we ran products such as z/OS DB2 V9, IBM Cognos® Business Intelligence Version 8.4 for Linux on System z, InfoSphere™ Warehouse for System z, InfoSphere Change Data Capture, z/OS WebSphere® V7, Tivoli® Omegamon for DB2 Performance expert. Utilizing these products we created transactional (OLTP), data warehouse query, and data warehouse refresh workloads. All the workloads were based on an existing web-based transactional Bookstore workload, that's currently utilized for internal testing within the System p® and z labs. While some IBM Cognos BI and ISWz product usage and experiences information is covered in this book, we do not go into the depth typically found in IBM Redbooks publications, since there's another book focused specifically on that

IBM® InfoSphere® Guardium® provides the simplest, most robust solution for data security and data privacy by assuring the integrity of trusted information in your data center. InfoSphere Guardium helps you reduce support costs by automating the entire compliance auditing process across heterogeneous environments. InfoSphere Guardium offers a flexible and scalable solution to support varying customer architecture requirements. This IBM Redbooks® publication provides a guide for deploying the Guardium solutions. This book also provides a roadmap process for implementing an InfoSphere Guardium solution that is based on years of experience and best practices that were collected from

various Guardium experts. We describe planning, installation, configuration, monitoring, and administrating an InfoSphere Guardium environment. We also describe use cases and how InfoSphere Guardium integrates with other IBM products. The guidance can help you successfully deploy and manage an IBM InfoSphere Guardium system. This book is intended for the system administrators and support staff who are responsible for deploying or supporting an InfoSphere Guardium environment.

Outside of the world of enterprise computing, there is one database that enables a huge range of software and hardware to flex relational database capabilities, without the baggage and cost of traditional database management systems. That database is SQLite—an embeddable database with an amazingly small footprint, yet able to handle databases of enormous size. SQLite comes equipped with an array of powerful features available through a host of programming and development environments. It is supported by languages such as C, Java, Perl, PHP, Python, Ruby, TCL, and more. The Definitive Guide to SQLite, Second Edition is devoted to complete coverage of the latest version of this powerful database. It offers a thorough overview of SQLite’s capabilities and APIs. The book also uses SQLite as the basis for helping newcomers make their first foray into database development. In only a short time you can be writing programs as diverse as a server-side browser plug-in or the next great iPhone or Android application! Learn about SQLite extensions for C, Java, Perl, PHP, Python, Ruby, and Tcl. Get solid coverage of SQLite internals. Explore developing iOS (iPhone) and Android applications with SQLite. SQLite is the solution chosen for thousands of products around the world, from mobile phones and GPS devices to set-top boxes and web browsers. You almost certainly use SQLite every day without even realizing it!

Perform fast interactive analytics against different data sources using the Presto high-performance, distributed SQL query engine. With this practical guide, you’ll learn how to conduct analytics on data where it lives, whether it’s Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Presto. Initially developed by Facebook, open source Presto is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Presto query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Presto’s use cases and learn about tools that will help you connect to Presto and query data Go deeper: Learn Presto’s internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Presto in production: Secure Presto, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Presto
Data Lake for Enterprises
SQL Programming

Data analysis and graphics with R

InfoSphere DataStage for Enterprise XML Data Integration

Big Data Computing

The Definitive Guide to SQLite

This 6-page laminated guides describes database creation and the creation and use of tables and queries. Many new features are covered as well as a complete overview of this popular software application. This guide comes laminated to ensure continuous use.

If you're considering R for statistical computing and data visualization, this book provides a quick and practical guide to just about everything you can do with the open source R language and software environment. You'll learn how to write R functions and use R packages to help you prepare, visualize, and analyze data. Author Joseph Adler illustrates each process with a wealth of examples from medicine, business, and sports. Updated for R 2.14 and 2.15, this second edition includes new and expanded chapters on R performance, the ggplot2 data visualization package, and parallel R computing with Hadoop. Get started quickly with an R tutorial and hundreds of examples Explore R syntax, objects, and other language details Find thousands of user-contributed R packages online, including Bioconductor Learn how to use R to prepare data for analysis Visualize your data with R's graphics, lattice, and ggplot2 packages Use R to calculate statistical tests, fit models, and compute probability distributions Speed up intensive computations by writing parallel R programs for Hadoop Get a complete desktop reference to R WSO2 Made Simple - dive deep into the core concepts of WSO2 to overcome the challenges faced while using the Enterprise Integrator About This Book Design, create, and publish services in the WSO2 technology Integrate the WSO2 Enterprise Integrator with other components and servers Log and test deployed services Who This Book Is For If you are a Java solutions architect or developer and are keen to understand how to build enterprise applications with WSO2, this book is for you. No prior knowledge of WSO2 is expected. What You Will Learn Configure WSO2 Enterprise Integrator server in a production environment Create SOAP Proxies and REST APIs Interact with WSO2 Message Broker Write services using the new language: Ballerina Schedule automatic tasks for the services you create Manage log messages depending on the log level of the system Integrate with social networks such as Twitter, Facebook, Instagram, and Yammer Test SOAP Services using the Tryit feature and SoapUI tool Work with Quality of Services In Detail WSO2 Enterprise Integrator brings together the most powerful servers provided by the WSO2 company for your SOA infrastructure. As an Enterprise Service Bus (ESB), WSO2 Enterprise Integrator provides greater flexibility and agility to meet growing enterprise demands, whereas, as a Data Services Server (DSS), it provides an easy-to-use platform for integrating data stores, creating composite views across different data sources, and hosting data services. Using real-world scenarios, this book helps you build a solid foundation in developing enterprise applications with powerful data integration capabilities using the WSO2 servers. The book gets you started by brushing up your knowledge about SOA architecture and how it can be implemented through WSO2. It will help build your expertise with the core concepts of ESB such as building proxies, sequences, endpoints, and how to work with these

in WSO2. Going further, you will also get well-acquainted with DSS data service concepts such as configuring data services, tasks, events, testing, and much more. The book will also cover API management techniques. Along with ESB and DSS, you will also learn about business process servers, the rules server and other components that together provide the control and robustness your enterprise applications will need. With practical use cases, the book covers typical daily scenarios you will come across while using these servers to give you hands-on experience. Style and approach The book is a complete guide and helps you get the right start—from understanding SOA architectures to getting valuable experience with two important integration servers such as ESB and DSS. It will include some real-world practical scenarios to help you master the best practices followed right across the industry and overcome the challenges you're likely to face on a daily basis.

HDInsight Essentials - Second Edition Packt Publishing Ltd

Next-Generation Big Data

Learning Apache Spark 2

Mastering Oracle GoldenGate

Concepts, Design and Applications

Presto: The Definitive Guide

Trino: The Definitive Guide

Integrating data from multiple sources is essential in the age of big data, but it can be a challenging and time-consuming task. This handy cookbook provides dozens of ready-to-use recipes for using Apache Sqoop, the command-line interface application that optimizes data transfers between relational databases and Hadoop. Sqoop is both powerful and bewildering, but with this cookbook's problem-solution-discussion format, you'll quickly learn how to deploy and then apply Sqoop in your environment. The authors provide MySQL, Oracle, and PostgreSQL database examples on GitHub that you can easily adapt for SQL Server, Netezza, Teradata, or other relational systems. Transfer data from a single database table into your Hadoop ecosystem Keep table data and Hadoop in sync by importing data incrementally Import data from more than one database table Customize transferred data by calling various database functions Export generated, processed, or backed-up data from Hadoop to your database Run Sqoop within Oozie, Hadoop's specialized workflow scheduler Load data into Hadoop's data warehouse (Hive) or database (HBase) Handle installation, connection, and syntax issues common to specific database vendors

If you've been asked to maintain large and complex Hadoop clusters, this book is a must. Demand for operations-specific material has skyrocketed now that Hadoop is becoming the de facto standard for truly large-scale data processing in the data center. Eric Sammer, Principal Solution Architect at Cloudera, shows you the particulars of running Hadoop in production, from planning, installing, and configuring the system to providing ongoing maintenance. Rather than run through all possible scenarios, this pragmatic operations guide calls out what works, as demonstrated in critical deployments. Get a high-level overview of HDFS and MapReduce: why they exist and how they work Plan a Hadoop deployment, from hardware and OS selection to network requirements Learn setup and configuration details with a list of critical properties Manage resources by sharing a cluster across multiple groups Get a runbook of the

most common cluster maintenance tasks Monitor Hadoop clusters—and learn troubleshooting with the help of real-world war stories Use basic tools and techniques to handle backup and catastrophic failure

Switching database vendors is often considered an exhausting challenge for database administrators and developers. Complexity, total cost, and the risk of downtime are often the reasons that restrain IT decision makers from starting the migration project. The primary goal of this book is to show that, with the proper planning and guidance, converting from MySQL to IBM® DB2® is not only feasible but straightforward. If you picked up this book, you are most likely considering converting to DB2 and are probably aware of several of the advantages of to converting to DB2 data server. In this IBM Redbooks® publication, we discuss in detail how you can take advantage of this industry leading database server. This book is an informative guide that describes how to convert the database system from MySQL™ 5.1 to DB2® V9.7 on Linux® and the steps that are involved in enabling the applications to use DB2 instead of MySQL. This guide also presents the best practices in conversion strategy and planning, conversion tools, porting steps, and practical conversion examples. It is intended for technical staff that is involved in a MySQL to DB2 conversion project.

Get ready to unlock the power of your data. With the fourth edition of this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. You'll learn about recent changes to Hadoop, and explore new case studies on Hadoop's role in healthcare systems and genomics data processing. Learn fundamental components such as MapReduce, HDFS, and YARN Explore MapReduce in depth, including steps for developing applications with it Set up and maintain a Hadoop cluster running HDFS and MapReduce on YARN Learn two data formats: Avro for data serialization and Parquet for nested data Use data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer) Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with Hadoop Learn the HBase distributed database and the ZooKeeper distributed configuration service

A Hands-On Tutorial

IBM i2 Integrated Law Enforcement: Technical Architecture and Deployment Guide

Getting Started with Oracle Data Integrator 11g

Theory and Practice

SAS 9. 3 Intelligence Platform

Data Provisioning for SAP HANA

This is a Cookbook with easy-to-follow recipes, containing practical and detailed examples which are all fully backed up with code, illustrations, and tips to dig deep into Backbone.js. This book is great for JavaScript developers who want to learn how to build advanced frontend applications with the Backbone.js framework. This book can be used in educational institutions to teach students how to build frontend applications in an MVC manner. It's assumed that you have some experience in jQuery, and are familiar with HTML.

A practical guide to implementing your enterprise data lake using Lambda Architecture as the base About This Book Build a full-fledged data lake for your organization with popular big data technologies using the Lambda architecture as the base Delve into the big data technologies required to meet modern day business strategies A highly practical guide to implementing enterprise data lakes with lots of examples and real-world use-cases Who This Book Is For Java developers and architects who would like to implement a data lake for their enterprise will find this book useful. If you want to get hands-on experience with the Lambda Architecture and big data technologies by implementing a practical solution using these technologies, this book will also help you. What You Will Learn Build an enterprise-level data lake using the relevant big data technologies Understand the core of the Lambda architecture and how to apply it in an enterprise Learn the technical details around Sqoop and its functionalities Integrate Kafka with Hadoop components to acquire enterprise data Use flume with streaming technologies for stream-based processing Understand stream- based processing with reference to Apache Spark Streaming Incorporate Hadoop components and know the advantages they provide for enterprise data lakes Build fast, streaming, and high-performance applications using Elasticsearch Make your data ingestion process consistent across various data formats with configurability Process your data to derive intelligence using machine learning algorithms In Detail The term "Data Lake" has recently emerged as a prominent term in the big data industry. Data scientists can make use of it in deriving meaningful insights that can be used by businesses to redefine or transform the way they operate. Lambda architecture is also emerging as one of the very eminent patterns in the big data landscape, as it not only helps to derive useful information from historical data but also correlates real-time data to enable business to take critical decisions. This book tries to bring these two important aspects — data lake and lambda architecture—together. This book is divided into three main sections. The first introduces you to the concept of data lakes, the importance of data lakes in enterprises, and getting you up-to-speed with the Lambda architecture. The second section delves into the principal components of building a data lake using the Lambda architecture. It introduces you to popular big data technologies such as Apache Hadoop, Spark, Sqoop, Flume, and Elasticsearch. The third section is a highly practical demonstration of putting it all together, and shows you how an enterprise data lake can be implemented, along with several real-world use-cases. It also shows you how other peripheral components can be added to the lake to make it more efficient. By the end of this book, you will be able to choose the right big data technologies using the lambda architectural patterns to build your enterprise data lake. Style and approach The book takes a pragmatic approach, showing ways to leverage big data technologies and lambda architecture to build an enterprise-level data lake.

With Microsoft SQL Server 2016, a variety of new features and enhancements to the data platform deliver breakthrough performance, advanced security, and richer, integrated reporting and analytics capabilities. In this ebook, we introduce new security features: Always Encrypted, Row-Level Security, and dynamic data masking; discuss enhancements that enable you to better manage performance and storage: TemDB configuration, query store, and Stretch Database; review several

improvements to Reporting Services; and also describe AlwaysOn Availability Groups, tabular enhancements, and R integration.

Get the most out of the rich development capabilities of SQL Server 2016 to build efficient database applications for your organization

About This Book- Utilize the new enhancements in Transact-SQL and security features in SQL Server 2016 to build efficient database applications- Work with temporal tables to get information about data stored in the table at any point in time- A detailed guide to SQL Server 2016, introducing you to multiple new features and enhancements to improve your overall development experience

Who This Book Is For This book is for database developers and solution architects who plan to use the new SQL Server 2016 features for developing efficient database applications. It is also ideal for experienced SQL Server developers who want to switch to SQL Server 2016 for its rich development capabilities. Some understanding of the basic database concepts and Transact-SQL language is assumed.

What You Will Learn- Explore the new development features introduced in SQL Server 2016- Identify opportunities for In-Memory OLTP technology, significantly enhanced in SQL Server 2016- Use columnstore indexes to get significant storage and performance improvements- Extend database design solutions using temporal tables- Exchange JSON data between applications and SQL Server in a more efficient way- Migrate historical data transparently and securely to Microsoft Azure by using Stretch Database- Use the new security features to encrypt or to have more granular control over access to rows in a table- Simplify performance troubleshooting with Query Store- Discover the potential of R's integration with SQL Server

In Detail Microsoft SQL Server 2016 is considered the biggest leap in the data platform history of the Microsoft, in the ongoing era of Big Data and data science. Compared to its predecessors, SQL Server 2016 offers developers a unique opportunity to leverage the advanced features and build applications that are robust, scalable, and easy to administer. This book introduces you to new features of SQL Server 2016 which will open a completely new set of possibilities for you as a developer. It prepares you for the more advanced topics by starting with a quick introduction to SQL Server 2016's new features and a recapitulation of the possibilities you may have already explored with previous versions of SQL Server. The next part introduces you to small delights in the Transact-SQL language and then switches to a completely new technology inside SQL Server - JSON support. We also take a look at the Stretch database, security enhancements, and temporal tables. The last chapters concentrate on implementing advanced topics, including Query Store, columnstore indexes, and In-Memory OLTP. You will finally be introduced to R and how to use the R language with Transact-SQL for data exploration and analysis. By the end of this book, you will have the required information to design efficient, high-performance database applications without any hassle.

Style and approach This book is a detailed guide to mastering the development features offered by SQL Server 2016, with a unique learn-as-you-do approach. All the concepts are explained in a very easy-to-understand manner and are supplemented with examples to ensure that you-the developer-are able to take that next step in building more powerful, robust applications for your organization with ease.

Mission-Critical Applications, Deeper Insights, Hyperscale Cloud Security Administration Guide

Deployment Guide for InfoSphere Guardium

A Guide for Developers and Administrators

Implementing an InfoSphere Optim Data Growth Solution

Use Tableau Prep to clean, combine, and transform your data for analysis

Explains the security model for the SAS Intelligence Platform and provides instructions for performing security-related administrative tasks. The emphasis is on suite-wide aspects of the security functionality that SAS provides. This title is also available online.

Learn about the fastest-growing open source project in the world, and find out how it revolutionizes big data analytics About This Book Exclusive guide that covers how to get up and running with fast data processing using Apache Spark Explore and exploit various possibilities with Apache Spark using real-world use cases in this book Want to perform efficient data processing at real time? This book will be your one-stop solution. Who This Book Is For This guide appeals to big data engineers, analysts, architects, software engineers, even technical managers who need to perform efficient data processing on Hadoop at real time. Basic familiarity with Java or Scala will be helpful. The assumption is that readers will be from a mixed background, but would be typically people with background in engineering/data science with no prior Spark experience and want to understand how Spark can help them on their analytics journey. What You Will Learn Get an overview of big data analytics and its importance for organizations and data professionals Delve into Spark to see how it is different from existing processing platforms Understand the intricacies of various file formats, and how to process them with Apache Spark. Realize how to deploy Spark with YARN, MESOS or a Stand-alone cluster manager. Learn the concepts of Spark SQL, SchemaRDD, Caching and working with Hive and Parquet file formats Understand the architecture of Spark MLlib while discussing some of the off-the-shelf algorithms that come with Spark. Introduce yourself to the deployment and usage of SparkR. Walk through the importance of Graph computation and the graph processing systems available in the market Check the real world example of Spark by building a recommendation engine with Spark using ALS. Use a Telco data set, to predict customer churn using Random Forests. In Detail Spark juggernaut keeps on rolling and getting more and more momentum each day. Spark provides key capabilities in the form of Spark SQL, Spark Streaming, Spark ML and Graph X all accessible via Java, Scala, Python and R. Deploying the key capabilities is crucial whether it is on a Standalone framework or as a part of existing Hadoop installation and configuring with Yarn and Mesos. The next part of the journey after installation is using key components, APIs, Clustering, machine learning APIs, data pipelines, parallel programming. It is important to understand why each framework component is key, how widely it is being used, its stability and pertinent use cases. Once we understand the individual components, we will take a couple of real life advanced analytics examples such as 'Building a Recommendation system', 'Predicting customer churn' and so on. The

objective of these real life examples is to give the reader confidence of using Spark for real-world problems. Style and approach With the help of practical examples and real-world use cases, this guide will take you from scratch to building efficient data applications using Apache Spark. You will learn all about this excellent data processing engine in a step-by-step manner, taking one aspect of it at a time. This highly practical guide will include how to work with data pipelines, dataframes, clustering, SparkSQL, parallel programming, and such insightful topics with the help of real-world use cases.

Explore common and not-so-common data transformation scenarios and solutions to become well-versed with Tableau Prep and create efficient and powerful data pipelines

Key Features

- Combine, clean, and shape data for analysis using self-service data preparation techniques*
- Become proficient with Tableau Prep for building and managing data flows across your organization*
- Learn how to combine multiple data transformations in order to build a robust dataset*

Book Description Tableau Prep is a tool in the Tableau software suite, created specifically to develop data pipelines. This book will describe, in detail, a variety of scenarios that you can apply in your environment for developing, publishing, and maintaining complex Extract, Transform and Load (ETL) data pipelines. The book starts by showing you how to set up Tableau Prep Builder. You'll learn how to obtain data from various data sources, including files, databases, and Tableau Extracts. Next, the book demonstrates how to perform data cleaning and data aggregation in Tableau Prep Builder. You'll also gain an understanding of Tableau Prep Builder and how you can leverage it to create data pipelines that prepare your data for downstream analytics processes, including reporting and dashboard creation in Tableau. As part of a Tableau Prep flow, you'll also explore how to use R and Python to implement data science components inside a data pipeline. In the final chapter, you'll apply the knowledge you've gained to build two use cases from scratch, including a data flow for a retail store to prepare a robust dataset using multiple disparate sources and a data flow for a call center to perform ad hoc data analysis. By the end of this book, you'll be able to create, run, and publish Tableau Prep flows and implement solutions to common problems in data pipelines. What you will learn

- Perform data cleaning and preparation techniques for advanced data analysis*
- Understand how to combine multiple disparate datasets*
- Prepare data for different Business Intelligence (BI) tools*
- Apply Tableau Prep's calculation language to create powerful calculations*
- Use Tableau Prep for ad hoc data analysis and data science flows*
- Deploy Tableau Prep flows to Tableau Server and Tableau Online*

Who this book is for This book is for business intelligence professionals, data analysts, and Tableau users looking to learn Tableau Prep essentials and create data pipelines or ETL processes using it. Beginner-level knowledge of data management will be beneficial to understand the concepts covered in this Tableau cookbook more effectively.

Cloud Computing: Theory and Practice provides students and IT

professionals with an in-depth analysis of the cloud from the ground up. Beginning with a discussion of parallel computing and architectures and distributed systems, the book turns to contemporary cloud infrastructures, how they are being deployed at leading companies such as Amazon, Google and Apple, and how they can be applied in fields such as healthcare, banking and science. The volume also examines how to successfully deploy a cloud application across the enterprise using virtualization, resource management and the right amount of networking support, including content delivery networks and storage area networks. Developers will find a complete introduction to application development provided on a variety of platforms. Learn about recent trends in cloud computing in critical areas such as: resource management, security, energy consumption, ethics, and complex systems Get a detailed hands-on set of practical recipes that help simplify the deployment of a cloud based system for practical use of computing clouds along with an in-depth discussion of several projects Understand the evolution of cloud computing and why the cloud computing paradigm has a better chance to succeed than previous efforts in large-scale distributed computing
Hadoop For Dummies

R in a Nutshell

Co-locating Transactional and Data Warehouse Workloads on System z Database Systems

Tableau Prep Cookbook

Using real life problems and simple solutions this book will make any issue seem small.

WebLogic Server books can be a bit dry but Dalton keeps the tone light and ensures no matter how complex the problem you always feel like you have someone right there with you helping you along. This book is ideal for those who know the basics of WebLogic but want to dive deeper and get to grips with more advanced topics. So if you are a datacenter operator, system administrator or even a Java developer this book could be exactly what you are looking for to take you one step further with Oracle WebLogic Serv.

Analytics is increasingly an integral part of day-to-day operations at today's leading businesses, and transformation is also occurring through huge growth in mobile and digital channels. Enterprise organizations are attempting to leverage analytics in new ways and transition existing analytics capabilities to respond with more flexibility while making the most efficient use of highly valuable data science skills. The recent growth and adoption of Apache Spark as an analytics framework and platform is very timely and helps meet these challenging demands. The Apache Spark environment on IBM z/OS® and Linux on IBM z Systems™ platforms allows this analytics framework to run on the same enterprise platform as the originating sources of data and transactions that feed it. If most of the data that will be used for Apache Spark analytics, or the most sensitive or quickly changing data is originating on z/OS, then an Apache Spark z/OS based environment will be the optimal choice for performance, security, and governance. This IBM® Redpaper™ publication explores the enterprise analytics market, use of Apache Spark on IBM z Systems™ platforms, integration between Apache Spark and other enterprise data sources, and case studies and examples of what can be achieved with Apache Spark in enterprise environments. It is of interest to data

scientists, data engineers, enterprise architects, or anybody looking to better understand how to combine an analytics framework and platform on enterprise systems.

If you want to discover one of the latest tools designed to produce stunning Big Data insights, this book features everything you need to get to grips with your data. Whether you are a data architect, developer, or a business strategist, HDInsight adds value in everything from development, administration, and reporting.

XML is one of the most common standards for the exchange of information. However, organizations find challenges in how to address the complexities of dealing with hierarchical data types, particularly as they scale to gigabytes and beyond. In this IBM® Redbooks® publication, we discuss and describe the new capabilities in IBM InfoSphere® DataStage® 8.5. These capabilities enable developers to more easily manage the design and processing requirements presented by the most challenging XML sources. Developers can use these capabilities to create powerful hierarchical transformations and to parse and compose XML data with high performance and scalability. Spanning both batch and real-time run times, these capabilities can be used to solve a broad range of business requirements. As part of the IBM InfoSphere Information Server 8.5 release, InfoSphere DataStage was enhanced with new hierarchical transformation capabilities called . XML Stage provides native XML schema support and powerful XML transformation functionality. These capabilities are based on a unique state-of-the-art technology that allows you to parse and compose any complex XML structure from and to a relational form, as well as to a separate hierarchical form. This book is targeted at an audience of systems designers and developers who focus on implementing XML integration support in their environments.

Pentaho 5.0 Reporting By Example Beginner's Guide

Apache Spark for the Enterprise: Setting the Business Free

The Ultimate Guide with Exercises, Tips and Tricks to Learn SQL

Apache Sqoop Cookbook

Over 85 recipes to implement efficient data warehousing solutions

MySQL to DB2 Conversion Guide

Master Oracle GoldenGate technology on multiple database platforms using this step-by-step implementation guide. Learn about advanced features to use in building a robust, high-availability replication system. Provided are detailed illustration of Oracle GoldenGate concepts, GoldenGate tools and add-ons, as well as illustrative examples. The book covers Oracle GoldenGate for Oracle database, and also discusses setup and configuration for other common databases such as IBM DB2, SYBASE ASE, MySQL, and Microsoft SQL Server. The technology landscape is fast-changing, and Mastering Oracle GoldenGate stays current by covering the new features included in Oracle GoldenGate 12c. The book covers both classic capture and integrated capture, as well as delivery. Also covered are Oracle GoldenGate security and performance tuning, to keep your system secure and performing at its best. You will learn to monitor your GoldenGate system using tools that come with Oracle GoldenGate management pack, as well as using shell scripts. Troubleshooting is well-illustrated with examples: Covering Oracle GoldenGate technology

across common database brands Discussing high-performing and secure replication environments Speaking to replication in Big Data and cloud computing environments What You Will Learn Implement Oracle GoldenGate for real time replication Secure and tune your replication environment for high performance Administer your Oracle GoldenGate environment Learn troubleshooting approaches with help of examples Make use of GoldenGate Management Pack and its API Feed live data into Big Data and cloud-based systems Who This Book Is For Database professionals who have chosen to ride the Oracle GoldenGate roller coaster for real-time replication solutions. The book is for beginners as well as professionals who are willing to master the leading replication technology in the industry. It is an excellent choice for professionals who are implementing or maintaining Oracle GoldenGate replication environments on any of the major database management system platforms.

IBM® i2® Integrated Law Enforcement is an IBM Smarter Cities® solution that addresses the needs of modern-day law enforcement agencies. It is a solution framework that provides the individual capabilities of the products that comprise the solution and extended capabilities developed through the synergistic integration of those product components. As a framework, IBM i2 Integrated Law Enforcement allows for the continuous expansion of capabilities by putting together building blocks within the system and integrating with new, external systems. In doing so, an organization can respond and adapt to its changing needs. Simply stated, the configuration, integration, and implementation of IBM i2 Integrated Law Enforcement and its components provide the tools for more effective law enforcement. This IBM Redpaper™ publication explains the technology and the architecture on which the solution is built. Most importantly, this paper enables technical teams to install, configure, and deploy an instance of the i2 Integrated Law Enforcement solution using the product i2 Intelligent Law Enforcement V1.0.1. This paper is targeted to solution architects, system and deployment engineers, security specialists, data management experts, system analysts, software developers and test engineers, and system administrators. Readers of this paper will benefit from the IBM Redguide™ publication "Integrated Law Enforcement: A Holistic Approach to Solving Crime", REDP-5116.

This book unravels the mystery of Big Data computing and its power to transform business operations. The approach it uses will be helpful to any professional who must present a case for realizing Big Data computing solutions or to those who could be involved in a Big Data computing project. It provides a framework that enables business and

technical managers to make optimal decisions necessary for the successful migration to Big Data computing environments and applications within their organizations.

Combine high volume data movement, complex transformations and real-time data integration with the robust capabilities of ODI in this practical guide.

Big Data Analytics with Java

SOA and data services with WSO2 Enterprise Integrator

Hadoop: The Definitive Guide

Access XP

Hadoop Operations

Unlocking Hadoop for Your Relational Database

Utilize this practical and easy-to-follow guide to modernize traditional enterprise data warehouse and business intelligence environments with next-generation big data technologies. Next-Generation Big Data takes a holistic approach, covering the most important aspects of modern enterprise big data. The book covers not only the main technology stack but also the next-generation tools and applications used for big data warehousing, data warehouse optimization, real-time and batch data ingestion and processing, real-time data visualization, big data governance, data wrangling, big data cloud deployments, and distributed in-memory big data computing. Finally, the book has an extensive and detailed coverage of big data case studies from Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard. What You'll Learn Install Apache Kudu, Impala, and Spark to modernize enterprise data warehouse and business intelligence environments, complete with real-world, easy-to-follow examples, and practical advice Integrate HBase, Solr, Oracle, SQL Server, MySQL, Flume, Kafka, HDFS, and Amazon S3 with Apache Kudu, Impala, and Spark Use StreamSets, Talend, Pentaho, and CDAP for real-time and batch data ingestion and processing Utilize Trifacta, Alteryx, and Datameer for data wrangling and interactive data processing Turbocharge Spark with Alluxio, a distributed in-memory storage platform Deploy big data in the cloud using Cloudera Director Perform real-time data visualization and time series analysis using Zoomdata, Apache Kudu, Impala, and Spark Understand enterprise big data topics such as big data governance, metadata management, data lineage, impact analysis, and policy enforcement, and how to use Cloudera Navigator to perform

common data governance tasks Implement big data use cases such as big data warehousing, data warehouse optimization, Internet of Things, real-time data ingestion and analytics, complex event processing, and scalable predictive modeling Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard Who This Book Is For BI and big data warehouse professionals interested in gaining practical and real-world insight into next-generation big data processing and analytics using Apache Kudu, Impala, and Spark; and those who want to learn more about other advanced enterprise topics Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications

Helps you find your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving Shows you how to improve the value of your Hadoop cluster, maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster From programmers challenged with building and maintaining affordable, scaleable data systems to administrators who must deal with huge volumes of information effectively and efficiently, this how-to has something to help you with Hadoop.

The second edition of this bestselling title is a perfect blend of theoretical knowledge and practical application. It progresses gradually from basic to advance concepts in database management systems, with numerous solved exercises to make learning easier and interesting. New to this edition are discussions on more commercial database management systems.

Teradata Cookbook

*Data Warehousing, Analytics, and Machine Learning at Scale
A Desktop Quick Reference*

SQL Server 2016 Developer's Guide

Google BigQuery: The Definitive Guide

SQL at Any Scale, on Any Storage, in Any Environment

Summary R in Action, Second Edition presents both the R language and the examples that make it so useful for business developers. Focusing on practical solutions, the book offers a crash course in statistics and covers elegant methods for dealing with messy and incomplete data that are difficult to analyze using traditional methods. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on time series analysis, cluster analysis, and classification methodologies, including decision trees, random forests, and support vector machines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Business pros and researchers thrive on data, and R speaks the language of data analysis. R is a powerful programming language for statistical computing. Unlike general-purpose tools, R provides thousands of modules for solving just about any data-crunching or presentation challenge you're likely to face. R runs on all important platforms and is used by thousands of major corporations and institutions worldwide. About the Book R in Action, Second Edition teaches you how to use the R language by presenting examples relevant to

scientific, technical, and business developers. Focusing on practical solutions, the book offers a crash course in statistics, including elegant methods for dealing with messy and incomplete data. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on forecasting, data mining, and dynamic report writing. What's Inside Complete R language tutorial Using R to manage, analyze, and visualize data Techniques for debugging programs and creating packages OOP in R Over 160 graphs About the Author Dr. Rob Kabacoff is a seasoned researcher and teacher who specializes in data analysis. He also maintains the popular Quick-R website at statmethods.net. Table of Contents PART 1 GETTING STARTED Introduction to R Creating a dataset Getting started with graphs Basic data management Advanced data management PART 2 BASIC METHODS Basic graphs Basic statistics PART 3 INTERMEDIATE METHODS Regression Analysis of variance Power analysis Intermediate graphs Resampling statistics and bootstrapping PART 4 ADVANCED METHODS Generalized linear models Principal components and factor analysis Time series Cluster analysis Classification Advanced methods for missing data PART 5 EXPANDING YOUR SKILLS Advanced graphics with ggplot2 Advanced programming Creating a package Creating dynamic reports Advanced graphics with the lattice package available online only from manning.com/kabacoff2

This book is useful to teradata beginners.

Learn the basics of analytics on big data using Java, machine learning and other big data tools About This Book Acquire real-world set of tools for building enterprise level data science applications Surpasses the barrier of other languages in data science and learn create useful object-oriented codes Extensive use of Java compliant big data tools like apache spark, Hadoop, etc. Who This Book Is For This book is for Java developers who are looking to perform data analysis in production environment. Those who wish to implement data analysis in their Big data applications will find this book helpful. What You Will Learn Start from simple analytic tasks on big data Get into more complex tasks with predictive analytics on big data using machine learning Learn real time analytic tasks Understand the concepts with examples and case studies Prepare and refine data for analysis Create charts in order to understand the data See various real-world datasets In Detail This book covers case studies such as sentiment analysis on a tweet dataset, recommendations on a movielens dataset, customer segmentation on an ecommerce dataset, and graph analysis on actual flights dataset. This book is an end-to-end guide to implement analytics on big data with Java. Java is the de facto language for major big data environments, including Hadoop. This book will teach you how to perform analytics on big data with production-friendly Java. This book basically divided into two sections. The first part is an introduction that will help the readers get acquainted with big data environments, whereas the second part will contain a hardcore discussion on all the concepts in

analytics on big data. It will take you from data analysis and data visualization to the core concepts and advantages of machine learning, real-life usage of regression and classification using Naive Bayes, a deep discussion on the concepts of clustering, and a review of simple neural networks on big data using deepLearning4j or plain Java Spark code. This book is a must-have book for Java developers who want to start learning big data analytics and want to use it in the real world. Style and approach The approach of book is to deliver practical learning modules in manageable content. Each chapter is a self-contained unit of a concept in big data analytics. Book will step by step builds the competency in the area of big data analytics. Examples using real world case studies to give ideas of real applications and how to use the techniques mentioned. The examples and case studies will be shown using both theory and code.

The big tech companies are increasingly relying on the database management systems to store and maintain the massive volume of data generated by our digital lives. The Relational Database Management System (RDBMS) is extensively used by these tech giants to not only store the large volume of data but as an advanced tool to gain insight from massive volume of data generated by our increasingly digital lives. The Structured Query Language (SQL) is the language of choice to define, manipulate, control and query the data within a RDBMS. This book is written to serve as your personal guide so you can efficiently and effectively learn and write SQL statements or queries to retrieve from and update data on relational databases such as MySQL. You will be able to install the free and open MySQL user interface with the instructions provided in this book. This will allow you to get hands-on practice utilizing a variety of exercises included in this book, so you will be able to create not only correct but efficient SQL queries to succeed at work and ace those job interview questions. Some of the highlights of this book are: - Foundational concepts of SQL language as well as 5 fundamental types of SQL queries namely - Learn the thumb rules for building SQL syntax or query - A variety of SQL data types that are a pre-requisite for learning SQL - Overview of a wide range of user interfaces available with MySQL servers - Learn how to create an effective database on the MySQL server - Learn the concept of temporary tables, derived tables and how you can create a new table from an existing one - Learn how to create new user accounts, update the user password as needed, grant and revoke access privileges - Learn CREATE VIEW, MERGE, TEMPTABLE, UNDEFINED, Updatable SQL Views and ALTER VIEW - The properties of SQL transactions as well as various SQL transaction statements with controlling clauses Don't miss the opportunity to quickly learn a programming language like SQL. Don't you think it can be that easy? If you really want to have proof of all this, don't waste any more time! Grab your copy now!

R in Action

Cloud Computing

WSO2 Developer's Guide

Introducing Microsoft SQL Server 2016