

Pentaho 50 Reporting By Example Beginners Guide

Create reports and solve common report problems with minimal fuss. About This Book Use this unique book to master the basics and advanced features of Pentaho 8 Reporting. A book showing developers and analysts with IT skills how to create and use the best possible reports using the Pentaho platform. Written with a very practical approach: full of tutorials and practical examples (source code included). Who This Book Is For This book is written for two types of professionals and students: Information Technologists with a basic knowledge of Databases and Java Developers with medium seniority. Developers will be interested to discover how to embed reports in a third-party Java application. What You Will Learn The basics of Pentaho Reporting (Designer and SDK) and its initial setup. Develop the most attractive reports on top of a wide range of data sources. Perform detailed customization of layout, parameterization, internationalization, behaviors, and more for your custom reports developed with Pentaho Reporting. Integrate Pentaho reports into third-party Java application with full control over interactions, layout, and behavior in general. Use Pentaho reports in the other components of the Pentaho Suite (BA Platform and PDI). In Detail This hands-on tutorial, filled with exercises and examples, introduces the reader to a variety of concepts within Pentaho Reporting. With screenshots that show you how reports look at design time as well as how they should look when rendered as PDF, Excel, HTML, Text, Rich-Text-File, XML, and CSV, this book also contains complete example source code that you can copy and paste into your environment to get up-and-running quickly. Updated to cover the features of Pentaho 8, this book will teach you everything you need to know to build fast, efficient reports using Pentaho. If your interest lies in the technical details of creating reports and you want to see how to solve common reporting problems with a minimum of fuss, this is the book for you. Style and approach A step-by-step guide covering technical topics relating to environments, best practices, and source code, to enable the reader to assemble the best reports and use them in existing Java applications.

As the importance of the social sector has increased in recent years, utilizing social enterprise aids in the development of knowledge, research, and practices in order to achieve an organizations' goals. Therefore, an understanding of ICT implementation in regards to social enterprise is crucial for effectiveness. Social E-Enterprise: Value Creation through ICT provides research on the understanding of ICT in the social enterprise field as it emerges as a major component of both business model and developed economy. This reference source focuses on the role of information communication technology as it promotes the development of the social sector.

Using Agile methods, you can bring far greater innovation, value, and quality to any data warehousing (DW), business intelligence (BI), or analytics project. However, conventional Agile methods must be carefully adapted to address the unique characteristics of DW/BI projects. In Agile Analytics, Agile pioneer Ken Collier shows how to do just that. Collier introduces platform-agnostic Agile solutions for integrating infrastructures consisting of diverse operational, legacy, and specialty systems that mix commercial and custom code. Using working examples, he shows how to manage analytics development teams with widely diverse skill sets and how to support enormous and fast-growing data volumes. Collier's techniques offer optimal value whether your projects involve "back-end" data management, "front-end" business analysis, or both. Part I focuses on Agile project management techniques and delivery team coordination, introducing core practices that shape the way your Agile DW/BI project community can collaborate toward success Part II presents technical methods for enabling continuous delivery of business value at production-quality levels, including evolving superior designs; test-driven DW development; version control; and project automation Collier brings together proven solutions you can apply right now--whether you're an IT decision-maker, data warehouse professional, database administrator, business intelligence specialist, or database developer. With his help, you can mitigate project risk, improve business alignment, achieve better results--and have fun along the way.

Pentaho Data Integration Cookbook Second Edition is written in a cookbook format, presenting examples in the style of recipes. This allows you to go directly to your topic of interest, or follow topics throughout a chapter to gain a thorough in-depth knowledge. Pentaho Data Integration Cookbook Second Edition is designed for developers who are familiar with the basics of Kettle but who wish to move up to the next level. It is also aimed at advanced users that want to learn how to use the new features of PDI as well as and best practices for working with Kettle.

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

This revised edition covers all aspects of public health informatics and discusses the creation and management of an information technology infrastructure that is essential in linking state and local organizations in their efforts to gather data for the surveillance and prevention. Public health officials will have to understand basic principles of information resource management in order to make the appropriate technology choices that will guide the future of their organizations. Public health continues to be at the forefront of modern medicine, given the importance of implementing a population-based health approach and to addressing chronic health conditions. This book provides informatics principles and examples of practice in a public health context. In doing so, it clarifies the ways in which newer information technologies will improve individual and community health status. This book's primary purpose is to consolidate key information and promote a strategic approach to information systems and development, making it a resource for use by faculty and students of public health, as well as the practicing public health professional. Chapter

highlights include: The Governmental and Legislative Context of Informatics; Assessing the Value of Information Systems; Ethics, Information Technology, and Public Health; and Privacy, Confidentiality, and Security. Review questions are featured at the end of every chapter. Aside from its use for public health professionals, the book will be used by schools of public health, clinical and public health nurses and students, schools of social work, allied health, and environmental sciences.

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

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

Get up and running with the Pentaho Data Integration tool using this hands-on, easy-to-read guide. About This Book Manipulate your data by exploring, transforming, validating, and integrating it using Pentaho Data Integration 8 CE. A comprehensive guide exploring the features of Pentaho Data Integration 8 CE. Connect to any database engine, explore the databases, and perform all kind of operations on relational databases. Who This Book Is For This book is a must-have for software developers, business intelligence analysts, IT students, or anyone involved or interested in developing ETL solutions. If you plan on using Pentaho Data Integration for doing any data manipulation task, this book will help you as well. This book is also a good starting point for data warehouse designers, architects, or anyone who is responsible for data warehouse projects and needs to load data into them. What You Will Learn Explore the features and capabilities of Pentaho Data Integration 8 Community Edition. Install and get started with PDI. Learn the ins and outs of Spoon, the graphical designer tool. Learn to get data from all kind of data sources, such as plain files, Excel spreadsheets, databases, and XML files. Use Pentaho Data Integration to perform CRUD (create, read, update, and delete) operations on relational databases. Populate a data mart with Pentaho Data Integration. Use Pentaho Data Integration to organize files and folders, run daily processes, deal with errors, and more. In Detail Pentaho Data Integration (PDI) is an intuitive and graphical environment packed with drag-and-drop design and powerful Extract-Transform-Load (ETL) capabilities. This book shows and explains the new interactive features of Spoon, the revamped look and feel, and the newest features of the tool including transformations and jobs Executors and the invaluable Metadata Injection capability. We begin with the installation of PDI software and then move on to cover all the key PDI concepts. Each of the chapter introduces new features, enabling you to gradually get practicing with the tool. First, you will learn to do all kind of data manipulation and work with simple plain files. Then, the book teaches you how you can work with relational databases inside PDI. Moreover, you will be given a primer on data warehouse concepts and you will learn how to load data in a data warehouse. During the course of this book, you will be familiarized with its intuitive, graphical and drag-and-drop design environment. By the end of this book, you will learn everything you need to know in order to meet your data manipulation requirements. Besides, you will be given best practices and advises for designing and deploying your projects. Style and approach Step by step guide filled with practical, real world scenarios and examples.

Summary Mondrian in Action teaches business users and developers how to use Mondrian and related tools for strategic business analysis. You'll learn how to design and populate a data warehouse and present the data via a multidimensional model. You'll follow examples showing how to create a Mondrian schema and then expand it to add basic security based on the users' roles. About the Technology Mondrian is an open source, lightning-fast data analysis engine designed to help you explore your business data and perform speed-of-thought analysis. Mondrian can be integrated into a wide

variety of business analysis applications and learning it requires no specialized technical knowledge. About this Book Mondrian in Action teaches you to use Mondrian for strategic business analysis. In it, you'll learn how to organize and present data in a multidimensional manner. You'll follow apt and thoroughly explained examples showing how to create a Mondrian schema and then expand it to add basic security based on users' roles. Developers will discover how to integrate Mondrian using its olap4j Java API and web service calls via XML for Analysis. Written for developers building data analysis solutions. Appropriate for tech-savvy business users and DBAs needing to query and report on data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's Inside Mondrian from the ground up—no experience required A primer on business analytics Using Mondrian with a variety of leading applications Optimizing and restricting business data for fast, secure analysis About the Authors William D. Back is an Enterprise Architect and Director of Pentaho Services. Nicholas Goodman is a Business Intelligence pro who has authored training courses on OLAP and Mondrian. Julian Hyde founded Mondrian and is the project's lead developer. Table of Contents Beyond reporting: business analytics Mondrian: a first look Creating the data mart Multidimensional modeling: making analytics data accessible How schemas grow Securing data Maximizing Mondrian performance Dynamic security Working with Mondrian and Pentaho Developing with Mondrian Advanced analytics

Getting Started With Amazon Redshift is a step-by-step, practical guide to the world of Redshift. Learn to load, manage, and query data on Redshift. This book is for CIOs, enterprise architects, developers, and anyone else who needs to get familiar with RedShift. The CIO will gain an understanding of what their technical staff is working on; the technical implementation personnel will get an in-depth view of the technology, and what it will take to implement their own solutions.

The big data era is upon us: data are being generated, analyzed, and used at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. Since the value of data explodes when it can be linked and fused with other data, addressing the big data integration (BDI) challenge is critical to realizing the promise of big data. BDI differs from traditional data integration along the dimensions of volume, velocity, variety, and veracity. First, not only can data sources contain a huge volume of data, but also the number of data sources is now in the millions. Second, because of the rate at which newly collected data are made available, many of the data sources are very dynamic, and the number of data sources is also rapidly exploding. Third, data sources are extremely heterogeneous in their structure and content, exhibiting considerable variety even for substantially similar entities. Fourth, the data sources are of widely differing qualities, with significant differences in the coverage, accuracy and timeliness of data provided. This book explores the progress that has been made by the data integration community on the topics of schema alignment, record linkage and data fusion in addressing these novel challenges faced by big data integration. Each of these topics is covered in a systematic way: first starting with a quick tour of the topic in the context of traditional data integration, followed by a detailed, example-driven exposition of recent innovative techniques that have been proposed to address the BDI challenges of volume, velocity, variety, and veracity. Finally, it presents merging topics and opportunities that are specific to BDI, identifying promising directions for the data integration community.

Find the right people with the right skills. This book clarifies best practices for creating high-functioning data integration teams, enabling you to understand the skills and requirements, documents, and solutions for planning, designing, and monitoring both one-time migration and daily integration systems. The growth of data is exploding. With multiple sources of information constantly arriving across enterprise systems, combining these systems into a single, cohesive, and documentable unit has become more important than ever. But the approach toward integration is much different than in other software disciplines, requiring the ability to code, collaborate, and disentangle complex business rules into a scalable model. Data migrations and integrations can be complicated. In many cases, project teams save the actual migration for the last weekend of the project, and any issues can lead to missed deadlines or, at worst, corrupted data that needs to be reconciled post-deployment. This book details how to plan strategically to avoid these last-minute risks as well as how to build the right solutions for future integration projects. What You Will Learn Understand the “language” of integrations and how they relate in terms of priority and ownership Create valuable documents that lead your team from discovery to deployment Research the most important integration tools in the market today Monitor your error logs and see how the output increases the cycle of continuous improvement Market across the enterprise to provide valuable integration solutions Who This Book Is For The executive and integration team leaders who are building the corresponding practice. It is also for integration architects, developers, and business analysts who need additional familiarity with ETL tools, integration processes, and associated project deliverables.

The pervasiveness of and universal access to modern Information and Communication Technologies has enabled a popular new paradigm in the dissemination of information, art, and ideas. Now, instead of relying on a finite number of content providers to control the flow of information, users can generate and disseminate their own content for a wider audience. Open Source Technology: Concepts, Methodologies, Tools, and Applications investigates examples and methodologies in user-generated and freely-accessible content available through electronic and online media. With applications in education, government, entertainment, and more, the technologies explored in these volumes will provide a comprehensive reference for web designers, software developers, and practitioners in a wide variety of fields and disciplines.

Acquire finesse with CTools features and build rich and custom analytics solutions using Pentaho About This Book Learn everything you need to know to make the most of CTools Create interactive and remarkable dashboards using the CTools Understand how to use and create data visualizations that can make the difference The author of our book works for Pentaho as a Senior Consultant Acts as a follow-up to Packt's previously published products on Pentaho such as

Pentaho Business Analytics Cookbook, Pentaho Analytics for MongoDB, Pentaho Data Integration Cookbook - Second Edition, and Pentaho Reporting [Video] Our book is based on the latest version of Pentaho, that is, 6.0 Who This Book Is For If you are a CTools developer and would like to expand your knowledge and create attractive dashboards and frameworks, this book is the go-to-guide for you. A basic knowledge of JavaScript and Cascading Style Sheets (CSS) is highly recommended. What You Will Learn Install Community Tools on Pentaho; and understand the necessary concepts and considerations when creating an exciting dashboard design Get data from many different Pentaho datasources and deliver it in different formats (CSV, XLS, XML, or JSON) Use the Community Data Access (CDA) as the data abstraction layer and understand the concepts in the Community Dashboard Framework (CDF) Create a Community Dashboard Editor (CDE) dashboard and make the most of the main components Create and make use of widgets and use duplicate components to have data-driven sections on the dashboard Customize and create interaction between all components, including charts, using the Community Charts Components Create and embed dashboards in a better and new way Create plugins and make use of parameters inside Pentaho without writing code In Detail Pentaho and CTools are two of the fastest and most rapidly growing tools for practical solutions not found in any other tool available on the market. Using Pentaho allows you to build a complete analytics solution, and CTools brings an advanced flexibility to customizing them in a remarkable way. CTools provides its users with the ability to utilize Web technologies and data visualization concepts, and make the most of best practices to create a huge visual impact. The book starts with the basics of the framework and how to get data to your dashboards. We'll take you all the way through to create your custom and advanced dashboards that will create an effective visual impact and provide the best user experience. You will be given deep insights into the lifecycle of dashboards and the working of various components. Further, you will create a custom dashboard using the Community Dashboards Editor and use datasources to load data on the components. You will also create custom content using Query, the Freeform Addins Popup, and text components. Next, you will make use of widgets to create similar sections and duplicate components to reproduce other components on a dashboard. You will then learn to build a plugin without writing Java code, use Sparkl as a CPK plugin manager, and understand the application of deployment and version control to dashboard development. Finally, you will learn tips and tricks that can be very useful while embedding dashboards into other applications. This guide is an invaluable tutorial if you are planning to use custom and advanced dashboards among the solutions that you are building with Pentaho. Style and approach This book is a pragmatic, easy-to-follow guide that provides theoretical concepts, ideas, and tricks to better understand the necessary theoretical concepts. It also provides you with a set of highly intriguing samples of dashboards with customized code within them that can be utilized for future projects.

It will be a step-by-step tutorial that will discuss best practices. The book is structured in such a way that it can be read both from start to end or can be dipped into. If you are a developer who is looking to learn a fast and easy way to learn to develop your business intelligence apps with QlikView, then this book is for you. If you are a power-user in a QlikView environment, then you will find quicker ways of working with QlikView. You should know the basics of business intelligence before you pick up this book. This book covers QlikView Desktop Personal Edition. Deployments to QlikView Server/Publisher are out of scope for this book.

For fifty years, Medicare and Medicaid have stood at the center of a contentious debate surrounding American government, citizenship, and health care entitlement. In *Medicare and Medicaid at 50*, leading scholars in politics, government, economics, health policy, and history offer a comprehensive assessment of the evolution of these programs and their impact on society -- from their origins in the Great Society era to the current battles over the Affordable Care Act ("Obamacare"). These highly accessible essays examine Medicare and Medicaid from their origins as programs for the elderly and poor to their later role as a safety net for the middle class. Along the way, they have served as touchstones for heated debates about economics, social welfare, and the role of government. *Medicare and Medicaid at 50* addresses key questions for understanding the past and future of health policy in America, including:

- What were the origins for these initiatives, and how were they transformed over time?
- What marks have Medicare and Medicaid left on society?
- In what ways have these programs produced innovation, even in eras of retrenchment?
- How did Medicaid, once regarded as a poor person's program, expand its benefits and coverage over the decades to become the platform for the ACA's future expansion?

The volume's contributors go on to examine the powerful role of courts in these transformations, along with the shifting roles of Congress, public opinion, and state governors in the programs' ongoing evolution. From Lyndon Johnson to Barack Obama on the left, and from Ronald Reagan to George W. Bush on the right, American political leaders have tied their political fortunes to the fate of America's entitlement programs; *Medicare and Medicaid at 50* helps explain why, and how those ongoing debates are likely to shape the future of the Affordable Care Act.

This book gives step-by-step instructions on how to do things. The basics are explained first and then examples help to clarify and reinforce the principles. The book is aimed at experienced Java developers and system architects who want to develop complex Java applications using the OSWorkflow workflow engine. OSWorkflow is a flexible low-level workflow implementation for developers and architects; it is not a quick "plug-and-play" solution for non-technical end users.

The LNCS journal *Transactions on Large-Scale Data- and Knowledge-Centered Systems* focuses on data management, knowledge discovery, and knowledge processing, which are core and hot topics in computer science. Since the 1990s, the Internet has become the main driving force behind application development in all domains. An increase in the demand for resource sharing (e.g., computing resources, services, metadata, data sources) across different sites connected through networks has led to an evolution of data- and knowledge management systems from centralized systems to decentralized systems enabling large-scale distributed applications providing high scalability. This, the 48th issue of *Transactions on Large-Scale Data- and Knowledge-Centered Systems*, contains 8 invited papers dedicated to the memory of Prof. Dr. Roland Wagner. The topics covered include distributed database systems, NewSQL, scalable

transaction management, strong consistency, caches, data warehouse, ETL, reinforcement learning, stochastic approximation, multi-agent systems, ontology, model-driven development, organisational modelling, digital government, new institutional economics and data governance.

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.

Over 50 recipes to learn how to use Pentaho Analytics and MongoDB to create powerful analysis and reporting solutions

About This Book Create reports and stunning dashboards with MongoDB data Accelerate data access and maximize productivity with unique features of Pentaho for MongoDB A step-by-step recipe-based guide for making full use of Pentaho suite tools with MongoDB Who This Book Is For This book is intended for data architects and developers with a basic level of knowledge of MongoDB. Familiarity with Pentaho is not expected. What You Will Learn Extract, load, and transform data from MongoDB collections to other datasources Design Pentaho Reports using different types of connections for MongoDB Create a OLAP mondrian schema for MongoDB Explore your MongoDB data using Pentaho Analyzer Utilize the drag and drop web interface to create dashboards Use Kettle Thin JDBC with MongoDB for analysis Integrate advanced dashboards with MondoDB using different types of connections Publish and run a report on Pentaho BI server using a web interface

In Detail MongoDB is an open source, schemaless NoSQL database system. Pentaho as a famous open source Analysis tool provides high performance, high availability, and easy scalability for large sets of data. The variant features in Pentaho for MongoDB are designed to empower organizations to be more agile and scalable and also enables applications to have better flexibility, faster performance, and lower costs. Whether you are brand new to online learning or a seasoned expert, this book will provide you with the skills you need to create turnkey analytic solutions that deliver insight and drive value for your organization. The book will begin by taking you through Pentaho Data Integration and how it works with MongoDB. You will then be taken through the Kettle Thin JDBC Driver for enabling a Java application to interact with a database. This will be followed by exploration of a MongoDB collection using Pentaho Instant view and creating reports with MongoDB as a datasource using Pentaho Report Designer. The book will then teach you how to explore and visualize your data in Pentaho BI Server using Pentaho Analyzer. You will then learn how to create advanced dashboards with your data. The book concludes by highlighting contributions of the Pentaho Community.

Style and approach A comprehensive, recipe-based guide to take complete advantage of the Pentaho Analytics for MongoDB.

Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition, offers a thorough grounding in machine learning concepts, along with practical advice on applying these tools and techniques in real-world data mining situations. This highly anticipated fourth edition of the most acclaimed work on data mining and machine learning teaches readers everything they need to know to get going, from preparing inputs, interpreting outputs, evaluating results, to the algorithmic methods at the heart of successful data mining approaches. Extensive updates reflect the technical changes and modernizations that have taken place in the field since the last edition, including substantial new chapters on probabilistic methods and on deep learning. Accompanying the book is a new version of the popular WEKA machine learning software from the University of Waikato. Authors Witten, Frank, Hall, and Pal include today's techniques coupled with the methods at the leading edge of contemporary research. Please visit the book companion website at <http://www.cs.waikato.ac.nz/ml/weka/book.html> It contains Powerpoint slides for Chapters 1-12. This is a very comprehensive teaching resource, with many PPT slides covering each chapter of the book Online Appendix on the Weka workbench; again a very comprehensive learning aid for the open source software that goes with the book Table of contents, highlighting the many new sections in the 4th edition, along with reviews of the 1st edition, errata, etc. Provides a thorough grounding in machine learning concepts, as well as practical advice on applying the tools and techniques to data mining projects Presents concrete tips and techniques for performance improvement that work by transforming the input or output in machine learning methods Includes a downloadable WEKA software toolkit, a comprehensive collection of machine learning algorithms for data mining tasks-in an easy-to-use interactive interface Includes open-access online courses that introduce practical applications of the material in the book

This book focuses on teaching you by example. The book walks you through every aspect of Pentaho Data Integration, giving systematic instructions in a friendly style, allowing you to learn in front of your computer, playing with the tool. The extensive use of drawings and screenshots make the process of learning Pentaho Data Integration easy. Throughout the book, numerous tips and helpful hints are provided that you will not find anywhere else. This book is a must-have for software developers, database administrators, IT students, and everyone involved or interested in developing ETL solutions, or, more generally, doing any kind of data manipulation. Those who have never used Pentaho Data Integration will benefit most from the book, but those who have, they will also find it useful. This book is also a good starting point for database administrators, data warehouse designers, architects, or anyone who is responsible for data warehouse projects and needs to load data into them.

Create advanced reports, including cross tabs, sub-reports, and charts that connect to practically any data source using open source Pentaho Reporting.

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO

and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries A complete guide to Pentaho Kettle, the Pentaho Data Integration toolset for ETL This practical book is a complete guide to installing, configuring, and managing Pentaho Kettle. If you're a database administrator or developer, you'll first get up to speed on Kettle basics and how to apply Kettle to create ETL solutions—before progressing to specialized concepts such as clustering, extensibility, and data vault models. Learn how to design and build every phase of an ETL solution. Shows developers and database administrators how to use the open-source Pentaho Kettle for enterprise-level ETL processes (Extracting, Transforming, and Loading data) Assumes no prior knowledge of Kettle or ETL, and brings beginners thoroughly up to speed at their own pace Explains how to get Kettle solutions up and running, then follows the 34 ETL subsystems model, as created by the Kimball Group, to explore the entire ETL lifecycle, including all aspects of data warehousing with Kettle Goes beyond routine tasks to explore how to extend Kettle and scale Kettle solutions using a distributed “cloud” Get the most out of Pentaho Kettle and your data warehousing with this detailed guide—from simple single table data migration to complex multisystem clustered data integration tasks.

Readers gain an overview of the core principles of IS and how it is practiced today as the concise, nine-chapter FUNDAMENTALS OF INFORMATION SYSTEMS, 9th edition combines the latest research with the most current coverage available. The book addresses analytics; big data; business intelligence; cloud computing; cybercrime; information system security; e-commerce; enterprise systems; ethical, legal, and social issues of information systems; mobile computing solutions; and systems acquisition and development. Readers learn how information systems can increase profits and reduce costs as they explore the latest information on artificial intelligence, change management, consumerization of information systems, data governance, energy and environmental concerns, global IS challenges, Internet of Everything, Internet censorship and net neutrality, IS careers, and virtual teams. Maximize success as an employee, decision maker, and business leader with this streamlined, contemporary resource. Important Notice: Media content referenced within the product description or the product text may not be available in the ebook version.

Master the tools you thought you knew and discover the features you never knew existed.

Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds.

Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

Over 70 recipes to solve ETL problems using Pentaho Kettle.

All Hell Breaking Loose is an eye-opening examination of climate change from the perspective of the U.S. military. The Pentagon, unsentimental and politically conservative, might not seem likely to be worried about climate change—still linked, for many people, with polar bears and coral reefs. Yet of all the major institutions in American society, none take climate change as seriously as the U.S. military. Both as participants in climate-triggered conflicts abroad, and as first responders to hurricanes and other disasters on American soil, the armed services are already confronting the impacts of global warming. The military now regards climate change as one of the top threats to American national security—and is busy developing strategies to cope with it. Drawing on previously obscure reports and government documents, renowned security expert Michael Klare shows that the U.S. military sees the climate threat as imperiling the country on several fronts at once. Droughts and food shortages are stoking conflicts in ethnically divided nations, with “climate refugees” producing worldwide havoc. Pandemics and other humanitarian disasters will increasingly require extensive military involvement. The melting Arctic is creating new seaways to defend. And rising seas threaten American cities and military bases themselves. While others still debate the causes of global warming, the Pentagon is intensely focused on its effects. Its response makes it clear that where it counts, the immense impact of climate change is not in doubt.

Learn all you need to know about seven key innovations disrupting business analytics today. These innovations—the open source business model, cloud analytics, the Hadoop ecosystem, Spark and in-memory analytics, streaming analytics, Deep Learning, and self-service analytics—are radically changing how businesses use data for competitive advantage. Taken together, they are disrupting the business analytics value chain, creating new opportunities. Enterprises who seize the opportunity will thrive and prosper, while others struggle and decline: disrupt or be disrupted. Disruptive Business Analytics provides strategies to profit from disruption. It shows you how to organize for insight, build and provision an open source stack, how to practice lean data warehousing, and how to assimilate disruptive innovations into an organization. Through a short history of business analytics and a detailed survey of products and services, analytics authority Thomas W. Dinsmore provides a practical explanation of the most compelling innovations available today. What You'll Learn Discover how the open source business model works and how to make it work for you See how cloud computing completely changes the economics of analytics Harness the power of Hadoop and its ecosystem Find out why Apache Spark is everywhere Discover the potential of streaming and real-time analytics Learn what Deep Learning can do and why it matters See how self-service analytics can change the way organizations do business Who This Book Is For Corporate actors at all levels of responsibility for analytics: analysts, CIOs, CTOs, strategic decision makers, managers, systems architects, technical marketers, product developers, IT personnel, and consultants.

The book is styled on a Cookbook, containing recipes - combined with free datasets - which will turn readers into proficient OpenRefine users in the fastest possible way. This book is targeted at anyone who works on or handles a large amount of data. No prior knowledge of OpenRefine is required, as we start from the very beginning and gradually reveal more advanced features. You don't even need your own dataset, as we provide example data to try out the book's recipes.

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

