Pentaho Reporting 35 For Java Developers Net City

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.

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.

Oracle 10g has become the most complex database ever created and Oracle tuning has become increasingly complex. This book provides a complete step-by-step approach for holistic Oracle tuning and it is the accumulated knowledge from tuning thousands of Oracle databases. Incorporating the principles of artificial intelligence, Oracle10g has developed a sophisticated mechanism for capturing and tracking database performance over time periods. This new complexity has introduced dozens of new v$ and DBA views, plus dozens of Automatic Workload Repository (AWR) tables. The AWR and its interaction with the Automatic Database Diagnostic Monitor (ADDM) is a revolution in database tuning. By understanding the internal workings of the AWR tables, the senior DBA can develop time-series tuning models to predict upcoming outages and dynamically change the instance to accommodate the impending resource changes. This is not a book for beginners. Targeted at the senior Oracle DBA, this book dives deep into the internals of the v$ views, the AWR table structures and the new DBA history views. Packed with ready-to-run scripts, you can quickly monitor and identify the most challenging performance issues.

This book focuses on teaching by example. Every chapter provides an overview, and then dives right into hands-on examples so you can see and play with the solution in your own environment. This book is for Java developers who don't need any prior experience with Liferay portal. Although Liferay portal makes heavy use of open source frameworks, no prior experience of using these is assumed.

Learning Pentaho Data Integration 8 CE

Practical Java Machine Learning

DATA WAREHOUSING

Hybrid Computational Intelligence

Pentaho Solutions

Undocumented Secrets of MATLAB-Java Programming

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 relationaldatabases 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-Tranform-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, your 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. Build machine learning (ML) solutions for Java development. This book shows you that when designing ML apps, data is the key driver and must be considered throughout all phases of the project life cycle. Practical Java Machine Learning helps you understand the importance of data and how to organize it for use within your ML project. You will be introduced to tools which can help you identify and manage your data including JSON, visualization, NoSQL databases, and cloud platforms including Google Cloud Platform and Amazon Web Services. Practical Java Machine Learning includes multiple projects, with particular focus on the Android mobile platform and features such as sensors, camera, and connectivity, each of which produce data that can power unique machine learning solutions. You will learn to build a variety of applications that demonstrate the capabilities of the Google Cloud Platform machine learning API, including data visualization for Java; document classification using the Weka ML environment; audio file classification for Android using ML with spectrogram voice data; and machine learning using device sensor data. After reading this book, you will come away with case study examples and projects that you can take away as templates for re-use and exploration for your own machine learning programming projects with Java. What You Will LearnIdentify, organize, and architect the data required for ML projects Deploy ML solutions in conjunction with cloud providers such as Google and Amazon Determine which algorithm is the most appropriate for a specific ML problem Implement Java ML solutions on Android mobile devices Create Java ML solutions to work with sensor data Build Java streaming based solutionsWho This Book Is For Experienced Java developers who have not implemented machine learning techniques before.

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.

There's a great deal of wisdom in a crowd, but how do you listen to a thousand people talking at once? Identifying the wants, needs, and knowledge of internet users can be like listening to a mob. In the Web 2.0 era, leveraging the collective power of user contributions, interactions, and feedback is the key to market dominance. A new category of powerful programming techniques lets you discover the patterns, inter-relationships, and individual profiles-the collective intelligence--locked in the data people leave behind as they surf websites, post blogs, and interact with other users. Collective Intelligence in Action is a hands-on guidebook for implementing collective intelligence concepts using Java. It is the first Java-based book to emphasize the underlying algorithms and technical implementation of vital data gathering and mining techniques like analyzing trends, discovering relationships, and making predictions. It provides a pragmatic approach to personalization by combining content-based analysis with collaborative approaches. This book is for Java developers implementing Collective Intelligence in real, high-use applications. Following a running example in which you harvest and use information from blogs, you learn to develop software that you can embed in your own applications. The code examples are immediately reusable and give the Java developer a working collective intelligence toolkit. Along the way, you work with, a number of APIs and open-source toolkits including text analysis and search using Lucene, web-crawling using Nutch, and applying machine learning algorithms using WEKA and the Java Data Mining (JDM) standard. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

Osworkflow

Mondrian in Action

Java/J2EE Job Interview Companion

Socket. IO Real-Time Web Application Development

The Definitive Reference

Data Mining

Your all-in-one resource for using Pentaho with MySQL forBusiness Intelligence and Data Warehousing Open-source Pentaho provides business intelligence (BI) and datawarehousing solutions at a fraction of

the cost of proprietarysolutions. Now you can take advantage of Pentaho for your businessneeds with this practical guide written by two major participantsin the Pentaho community. The book covers all components of the Pentaho BI Suite. You'lllearn to install, use, and maintain Pentaho-and find plenty ofbackground discussion that will bring you thoroughly up to speed onBI and Pentaho concepts. Of all available open source BI products, Pentaho offers themost comprehensive toolset and is the fastest growing open sourceproduct suite Explains how to build and load a data warehouse with PentahoKettle for data integration/ETL, manually create JFree (pentahoreporting services) reports using direct SQL queries, and createMondrian (Pentaho analysis services) cubes and attach them to aJPivot cube browser Review deploying reports, cubes and metadata to the Pentahoplatform in order to distribute BI solutions to end-users Shows how to set up scheduling, subscription and automaticdistribution The companion Web site provides complete source code examples,sample data, and links to related resources.

This book is for moderate to advanced PostgreSQL database professionals who wish to extend PostgreSQL, utilizing the most updated features of PostgreSQL 9.4. For a better understanding of this book, familiarity with writing SQL, a basic idea of query tuning, and some coding experience in your preferred language is expected.

The recent pursuits emerging in the realm of big data processing, interpretation, collection and organization have emerged in numerous sectors including business, industry and government organizations. Data sets such as customer transactions for a mega-retailer, weather monitoring, intelligence gathering, quickly outpace the capacities of traditional techniques and tools of data analysis. The 3V (volume, variability and velocity) challenges led to the emergence of new techniques and tools in data visualization, acquisition, and serialization. Soft Computing being regarded as a plethora of technologies of fuzzy sets (or Granular Computing), neurocomputing and evolutionary optimization brings forward a number of unique features that might be instrumental to the development of concepts and algorithms to deal with big data. This carefully edited volume provides the reader with an updated, in-depth material on the emerging principles, conceptual underpinnings, algorithms and practice of Computational Intelligence in the realization of concepts and implementation of big data architectures, analysis, and interpretation as well as data analytics. The book is aimed at a broad audience of researchers and practitioners including those active in various disciplines in which big data, their analysis and optimization are of genuine relevance. One focal point is the systematic exposure of the concepts, design methodology, and detailed algorithms. In general, the volume adheres to the top-down strategy starting with the concepts and motivation and then

proceeding with the detailed design that materializes in specific algorithms and representative applications. The material is self-contained and provides the reader with all necessary prerequisites and augments some parts with a step-by-step explanation of more advanced concepts supported by a significant amount of illustrative numeric material and some application scenarios to motivate the reader and make some abstract concepts more tangible.

Over 70 recipes to solve ETL problems using Pentaho Kettle.

A Guidebook for Successful Development and Design

Beginning Apache Pig

Pentaho Reporting 3.5 for Java Developers

Principles and Paradigms

Information Granularity, Big Data, and Computational Intelligence

Python Tutorial

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

Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, offers a thorough grounding in machine learning concepts as well as practical advice on applying machine learning tools and techniques in real-world data mining situations. This highly anticipated third edition of the most acclaimed work on data mining and machine learning will teach you everything you need to know about preparing inputs, interpreting outputs, evaluating results, and the algorithmic methods at the heart of successful data mining. Thorough updates reflect the technical changes and modernizations that have taken place in the field since the last edition, including new material on Data Transformations, Ensemble Learning, Massive Data Sets, Multi-instance Learning, plus a new version of the popular Weka machine learning software developed by the authors. Witten, Frank, and Hall include both tried-and-true techniques of today as well as methods at the leading edge of contemporary research. The book is targeted at information systems practitioners, programmers, consultants, developers, information technology managers, specification writers, data analysts, data modelers, database R&D professionals, data warehouse engineers, data mining professionals. The book will also be useful for professors and students of upper-level undergraduate and graduate-level data mining and machine learning courses who want to incorporate data mining as part of their data management knowledge base and expertise. Provides a thorough grounding

in machine learning concepts as well as practical advice on applying the tools and techniques to your data mining projects Offers concrete tips and techniques for performance improvement that work by transforming the input or output in machine learning methods Includes downloadable Weka software toolkit, a collection of machine learning algorithms for data mining tasks—in an updated, interactive interface. Algorithms in toolkit cover: data pre-processing, classification, regression, clustering, association rules, visualization

This book highlights state-of-the-art research on big data and the Internet of Things (IoT), along with related areas to ensure efficient and Internet-compatible IoT systems. It not only discusses big data security and privacy challenges, but also energy-efficient approaches to improving virtual machine placement in cloud computing environments. Big data and the Internet of Things (IoT) are ultimately two sides of the same coin, yet extracting, analyzing and managing IoT data poses a serious challenge. Accordingly, proper analytics infrastructures/platforms should be used to analyze IoT data. Information technology (IT) allows people to upload, retrieve, store and collect information, which ultimately forms big data. The use of big data analytics has grown tremendously in just the past few years. At the same time, the IoT has entered the public consciousness, sparking people's imaginations as to what a fully connected world can offer. Further, the book discusses the analysis of real-time big data to derive actionable intelligence in enterprise applications in several domains, such as in industry and agriculture. It explores possible automated solutions in daily life, including structures for smart cities and automated home systems based on IoT technology, as well as health care systems that manage large amounts of data (big data) to improve clinical decisions. The book addresses the security and privacy of the IoT and big data technologies, while also revealing the impact of IoT technologies on several scenarios in smart cities design. Intended as a comprehensive introduction, it offers in-depth analysis and provides scientists, engineers and professionals the latest techniques, frameworks and strategies used in IoT and big data technologies.

Hybrid Computational Intelligence: Challenges and Utilities is a comprehensive resource that begins with the basics and

main components of computational intelligence. It brings together many different aspects of the current research on HCI technologies, such as neural networks, support vector machines, fuzzy logic and evolutionary computation, while also covering a wide range of applications and implementation issues, from pattern recognition and system modeling, to intelligent control problems and biomedical applications. The book also explores the most widely used applications of hybrid computation as well as the history of their development. Each individual methodology provides hybrid systems with complementary reasoning and searching methods which allow the use of domain knowledge and empirical data to solve complex problems.

Pentaho 8 Reporting for Java Developers

Projects with Google Cloud Platform and Amazon Web Services

PostgreSQL Server Programming - Second Edition

Pentaho Data Integration Beginner's Guide

Data Lake for Enterprises

Data Mining: Practical Machine Learning Tools and Techniques

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms from the Python Web site, https: //www.python.org/, and may be freely distributed. The same site also contains distributions of and pointers to many free third party Python modules, programs and tools, and additional documentation. The Python interpreter is easily extended with new functions and data types implemented in C or C++ (or other languages callable from C). Python is also suitable as an extension language for customizable applications. This tutorial introduces the reader informally to the basic concepts and features of the python language and system. It helps to have a Python interpreter handy for hands-on experience, but all examples are self contained, so the tutorial can be read off-line as well. For a description of standard objects and modules, see library-index. reference-index gives a more formal definition of the language. To write extensions in C or C++, read extending-index and c-api-index. There are also several books covering Python in depth. This tutorial does not attempt to be comprehensive and cover every single feature, or even every commonly used feature. Instead, it introduces many of Python's most noteworthy features, and will give you a good idea of the language's flavor and style. After reading it, you will be able to read and write Python modules and programs, and you will be ready to learn more about the various Python library modules described in library-index. The Glossary is also worth going through.

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.

HBase is a remarkable tool for indexing mass volumes of data, but getting started with this distributed database and its ecosystem can be daunting. With this hands-on guide, you ' ll learn how to architect, design, and deploy your own HBase applications by examining real-world solutions. Along with HBase principles and cluster deployment guidelines, this book includes in-depth case studies that demonstrate how large companies solved specific use cases with HBase. Authors Jean-Marc Spaggiari and Kevin O ' Dell also provide draft solutions and code examples to help you implement your own versions of those use cases, from master data management (MDM) and document storage to near real-time event processing. You ' ll also learn troubleshooting techniques to help you avoid common deployment mistakes. Learn exactly what HBase does, what its ecosystem includes, and how to set up your environment Explore how real-world HBase instances were deployed and put into production Examine documented use cases for tracking healthcare claims, digital advertising, data management, and product quality Understand how HBase works with tools and techniques such as Spark, Kafka, MapReduce, and the Java API Learn how to identify the causes and understand the consequences of the most common HBase issues

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.

Big Data

MySQL Management and Administration with Navicat

Collective Intelligence in Action

Pentaho Data Integration 4 Cookbook

Practical Machine Learning Tools and Techniques with Java Implementations

Big Data Processing Made Easy

For a variety of reasons, the MATLAB®-Java interface was never fully documented. This is really quite unfortunate: Java is one of the most widely used programming languages, having many times the number of programmers and programming resources as MATLAB. Also unfortunate is the popular claim that while MATLAB is a fine programming platform for prototyping, it is not suitable for real-world, modern-looking applications. Undocumented Secrets of MATLAB®-Java Programming aims to correct this misconception. This book shows how using Java can significantly improve MATLAB program appearance and functionality, and that this can be done easily and even without any prior Java knowledge. Readers are led step-by-step from simple to complex customizations. Code snippets, screenshots, and numerous online references are provided to enable the utilization of this book as both a sequential tutorial and as a random-access reference suited for immediate use. Java-savvy readers will find it easy to tailor code samples for their particular needs; for Java newcomers, an introduction to Java and numerous online references are provided. This book demonstrates how The MATLAB programming environment relies on Java for numerous tasks, including networking, data-processing algorithms and graphical user-interface (GUI) We can use MATLAB for easy access to external Java functionality, either third-party or user-created Using Java, we can extensively customize the MATLAB environment and application GUI, enabling the creation of visually appealing and usable applications

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.

Learn to use Apache Pig to develop lightweight big data applications easily and quickly. This book shows you many optimization techniques and covers every context where Pig is used in big data analytics. Beginning Apache Pig shows you how Pig is easy to learn and requires relatively little time to develop big data applications.The book is divided into four parts: the complete features of Apache Pig; integration with other tools; how to solve complex business problems; and optimization of tools.You'll discover topics such as MapReduce and why it cannot meet every business need; the features of Pig Latin such as data types for each load, store, joins, groups, and ordering; how Pig workflows can be created; submitting Pig jobs using Hue; and working with Oozie. You'll also see how to extend the framework by writing UDFs and custom load, store, and filter functions. Finally you'll cover different optimization techniques such as gathering statistics about a Pig script, joining strategies, parallelism, and the role of data formats in good performance. What You Will Learn• Use all the features of Apache Pig• Integrate Apache Pig with other tools• Extend Apache Pig• Optimize Pig Latin code• Solve different use cases for Pig LatinWho This Book Is ForAll levels of IT professionals: architects, big data enthusiasts, engineers, developers, and big data administrators

Many corporations are finding that the size of their data sets are outgrowing the capability of their systems to store and process them. The data is becoming too big to manage and use with traditional tools. The solution: implementing a big data system. As Big Data Made Easy: A Working Guide to the

Complete Hadoop Toolset shows, Apache Hadoop offers a scalable, fault-tolerant system for storing and processing data in parallel. It has a very rich toolset that allows for storage (Hadoop), configuration (YARN and ZooKeeper), collection (Nutch and Solr), processing (Storm, Pig, and Map Reduce), scheduling (Oozie), moving (Sqoop and Avro), monitoring (Chukwa, Ambari, and Hue), testing (Big Top), and analysis (Hive). The problem is that the Internet offers IT pros wading into big data many versions of the truth and some outright falsehoods born of ignorance. What is needed is a book just like this one: a wide-ranging but easily understood set of instructions to explain where to get Hadoop tools, what they can do, how to install them, how to configure them, how to integrate them, and how to use them successfully. And you need an expert who has worked in this area for a decade—someone just like author and big data expert Mike Frampton. Big Data Made Easy approaches the problem of managing massive data sets from a systems perspective, and it explains the roles for each project (like architect and tester, for example) and shows how the Hadoop toolset can be used at each system stage. It explains, in an easily understood manner and through numerous examples, how to use each tool. The book also explains the sliding scale of tools available depending upon data size and when and how to use them. Big Data Made Easy shows developers and architects, as well as testers and project managers, how to: Store big data Configure big data Process big data Schedule processes Move data among SQL and NoSQL systems Monitor data Perform big data analytics Report on big data processes and projects Test big data systems Big Data Made Easy also explains the best part, which is that this toolset is free. Anyone can download it and—with the help of this book—start to use it within a day. With the skills this book will teach you under your belt, you will add value to your company or client immediately, not to mention your career.

Oracle Tuning

A Guide for Java Developers and Architects to Integrating Open-source Business Process Management Release 3. 6. 6rc1

Concepts, Techniques, Products and Applications

Pentaho Analytics for MongoDB Cookbook

This book offers a thorough grounding in machine learning concepts combined with practical advice on applying machine learning tools and techniques in real-world data mining situations. Clearly written and effectively illustrated, this book is ideal for anyone involved at any level in the work of extracting usable knowledge from large collections of data. Complementing the book's instruction is fully functional machine learning software.

This book covers all aspects of OSWorkflow for Java developers and system architects, from basics of Business Process Management and installing OSWorkflow to developing complex Java applications and integrating this open-source Java workflow engine with the third-party components Drools for business rules, Quartz for task scheduling, and Pentaho for dashboards. Authored by an active developer of the OSWorkflow project, it gives step-by-step instructions, explaining the basics and clarifying and reinforcing principles with real-life examples. OSWorkflow is a pure Java open-source workflow engine for technical users, who can focus on the business logic and rules without Petri Net or finite state machine coding and easily integrate OSWorkflow into applications to create simple or complex workflows as needed. Because OSWorkflow provides a relatively low-level but highly flexible workflow implementation for Java developers, it is not a quick plug-and-play solution for non-technical users.

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

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

Second Edition

Internet of Things and Big Data Analytics Toward Next-Generation Intelligence

Data Provisioning for SAP HANA

Business Intelligence and Data Warehousing with Pentaho and MySQL
Pentaho Data Integration Cookbook
Building Open Source ETL Solutions with Pentaho Data Integration

Practical DWR Web 2.0 Projects addresses the needs of most developers who would rather learn by example and by doing. This book contains several projects that developers can sink their teeth into doing. Written by accomplished Ajax and Java Web developer and author, Frank Zammetti, this book explores DWR and all it offers. It features six full, working applications that use DWR. This book allows you to lean by example, as you tear the applications apart, you see what makes them tick, and even discover how to extend them at your own pace.

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

The Third Edition of this well-received text analyses the fundamental concepts of data warehousing, data marts, and OLAP. The author discusses, in an easy-to-understand language, important topics such as data mining, how to build a data warehouse, and potential applications of data warehousing technology in government. Besides, the text compares and contrasts the currently available software tools used to design and develop data warehouses. While retaining the six existing case studies, it gives four new case studies: HARBOR, A Highly Available Data Warehouse A Typical Business Data Warehouse for a Trading Company Customer Data Warehouse for the World ' s First and Largest Online Bank in the United Kingdom A German Supermarket EDEKA ' s Data Warehouse The book, which is a blend of principles and real-life case studies, is intended as a text for students of B.Tech/M.Tech (Computer Science and Engineering), B.Tech/M.Tech (Information Technology), MBA, M.Sc. (Computer Science), M.Sc. (Information Technology), and MCA. It should also be of considerable utility and worth to software professionals and database practitioners.

This book develops a broad range of knowledge in ERP implementation and usage for textile and apparel vertical. Covered are two major areas in ERP: the basics about ERP and the technology and functioning of it and usage of ERP for textile and apparel vertical specifically. Also addressed are concerns of the industry, mainly on how to select the ERP, what to expect from ERP, and how it will be beneficial to the industry.

Pentaho Kettle Solutions
Big Data Made Easy
A Working Guide to the Complete Hadoop Toolset
Practical Machine Learning Tools and Techniques
Architecting HBase Applications
Challenges and Applications

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.

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.

Big Data: Principles and Paradigms captures the state-of-the-art research on the architectural aspects, technologies, and applications of Big Data. The book identifies potential future directions and technologies that facilitate insight into numerous scientific, business, and consumer applications. To help realize Big Data's full potential, the book addresses numerous challenges, offering the conceptual and technological solutions for tackling them. These challenges include life-cycle data management, large-scale storage, flexible processing infrastructure, data modeling, scalable machine learning, data analysis algorithms, sampling techniques, and privacy and ethical issues. Covers computational platforms supporting Big Data applications Addresses key principles underlying Big Data computing Examines key developments supporting next generation Big Data platforms Explores the challenges in Big Data computing and ways to overcome them Contains expert contributors from both academia and industry

A complete guide to Pentaho Kettle, the Pentaho Data lntegration 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.

Open source business analytics

Real-Time Data and Stream Processing at Scale

Kafka: The Definitive Guide

Pentaho 5.0 Reporting By Example Beginner's Guide

ERP for Textiles and Apparel Industry

Practical DWR 2 Projects

Socket.io Real-time Web Application Development.

400+ Java/J2EE Interview questions with clear and concise answers for: job seekers (junior/senior developers, architects, team/technical leads), promotion seekers, pro-active learners and interviewers. Lulu top 100 best seller. Increase your earning potential by learning, applying and succeeding. Learn the fundamentals relating to Java/J2EE in an easy to understand questions and answers approach. Covers 400+ popular interview Q&A with lots of diagrams, examples, code snippets, cross referencing and comparisons. This is not only an interview guide but also a quick reference guide, a refresher material and a roadmap covering a wide range of Java/J2EE related topics. More Java J2EE interview questions and answers & resume resources at http: //www.lulu.com/java-succes

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

"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"--

Learning Pentaho CTools

Using OpenRefine
An end-to-end guide to exploring, transforming, and integrating your data across multiple sources
Liferay Portal Systems Development