

Pentaho Data Integration Cookbook Second Edition Pulvirenti Adrian Sergio

Market_Desc: Primary MarketEngineering (BE/BTech)/ME/MTech students who are interested to develop conceptual level subject knowledge with examples of industrial strength applications.Secondary MarketMCA/MBA/Business users/business analysts
Special Features: · Foreword by Prof R Natarajan, Former Chairman, AICTE, Former Director, IIT Madras.· Excellent authorship.· Single source of introductory knowledge on business intelligence (BI).· Provides a good start for first-time learners typically from the engineering and management discipline.· Covers the complete life cycle of BI/Analytics Application development project.· Helps develop deeper understanding of the subject with an enterprise context, and discusses its application in businesses.· Explains concepts with the help of illustrations, application to real-life scenarios and provides opportunities to test understanding.· States the pre-requisites for each chapter and different reference sources available.· In addition the book also has the following pedagogical features:· Industrial application case studies.· Crossword puzzles/do it yourself exercises/assignments to help with self-assessment. The solutions to these have also been provided. · Glossary of terms.· References/web links/bibliography - generally at the end of every concept.CD Companion:To ensure that concepts can be practiced for deeper understanding at low cost, the book is accompanied with a CD containing:· Step-by-step Hands-On manual on:ü An open source tool, Pentaho Data Integrator (PDI) to explain the process of extraction of data from multiple varied sources.ü MS Excel to explain the concept of analysis.ü MS Access to generate reports on the analyzed data.· An integrated project that encompasses the complete life cycle of a BI project. About The Book: The book promises to be a single source of introductory knowledge on business intelligence which can be taught in one semester. It will provide a good start for first time learners typically from the engineering and management discipline. Business Intelligence subject cannot be studied in isolation. The book provides a holistic coverage beginning with an enterprise context, developing deeper understanding through the use of tools, touching a few domains where BI is embraced and discussing the problems that BI can help solve. It covers the complete life cycle of BI/Analytics project: Covering operational/transactional data sources, data transformation, data mart/warehouse design-build, analytical reporting, and dashboards. To ensure that concepts can be practiced for deeper understanding at low cost, the book is accompanied with step-by-step hands-on manual in the CD.

Find the right big data solution for your business or organization Big data management is one of the major challenges facing business, industry, and not-for-profit organizations. Data sets such as customer transactions for a mega-retailer, weather patterns monitored by meteorologists, or social network activity can quickly outpace the capacity of traditional data management tools. If you need to develop or manage big data solutions, you'll appreciate how these four experts define, explain, and guide you through this new and often confusing concept. You'll learn what it is, why it matters, and how to choose and implement solutions that work. Effectively managing big data is an issue of growing importance to businesses, not-for-profit organizations, government, and IT professionals Authors are experts in information management, big data, and a variety of solutions Explains big data in detail

and discusses how to select and implement a solution, security concerns to consider, data storage and presentation issues, analytics, and much more. Provides essential information in a no-nonsense, easy-to-understand style that is empowering. Big Data For Dummies cuts through the confusion and helps you take charge of big data solutions for your organization.

Get productive quickly with Pentaho Data Integration. Key Features: Take away the pain of starting with a complex and powerful system. Simplify your data transformation and integration work. Explore, transform, and validate your data with Pentaho Data Integration. Book Description: Pentaho Data Integration (PDI) is an intuitive and graphical environment packed with drag and drop design and powerful Extract-Transform-Load (ETL) capabilities. Given its power and flexibility, initial attempts to use the Pentaho Data Integration tool can be difficult or confusing. This book is the ideal solution. This book reduces your learning curve with PDI. It provides the guidance needed to make you productive, covering the main features of Pentaho Data Integration. It demonstrates the interactive features of the graphical designer, and takes you through the main ETL capabilities that the tool offers. By the end of the book, you will be able to use PDI for extracting, transforming, and loading the types of data you encounter on a daily basis. What you will learn: Design, preview and run transformations in Spoon. Run transformations using the Pan utility. Understand how to obtain data from different types of files. Connect to a database and explore it using the database explorer. Understand how to transform data in a variety of ways. Understand how to insert data into database tables. Design and run jobs for sequencing tasks and sending emails. Combine the execution of jobs and transformations. Who this book is for: This book is for software developers, business intelligence analysts, and others involved or interested in developing ETL solutions, or more generally, doing any kind of data manipulation.

Are you looking for resources to get started with migrating SQL Server Databases to PostgreSQL? This is the book for you. This concise eBook is intended to be a ready reference and a checklist for migrating your databases, applications and services from Microsoft SQL Server to PostgreSQL and discusses all the important information you need to be aware of before beginning a migration project. This eBook is divided into four chapters: Chapter 1: Getting Started 1. Introduction to Open Source Software, its advantages and disadvantages 2. Moving from Licensed software to Open Source software 3. Basics and the history of Microsoft SQL Server and PostgreSQL 4. Choosing the right PostgreSQL Edition and Version for your needs 5. Production, Development & Testing Workloads 6. On-Premise vs Cloud comparison Chapter 2: Platform Comparison 1. In-depth comparison of SQL Server and PostgreSQL 2. Concurrency Control 3. Database Structure comparison 4. Feature comparison 5. Data Type Mapping 6. Built-In Functions and Operators 7. SQL Syntax differences 8. Areas where PostgreSQL has an edge 9. Areas where SQL Server has an edge 10. GUI Client Tools Chapter 3: The Actual Migration 1. Schema migration 2. Data migration 3. Code migration 4. ETL/ELT Tools 5. Reporting Platforms 6. Analytics Platforms 7. Scheduling Options 8. Application/Services Migration Chapter 4: Maintenance and Monitoring 1. Maintenance and Monitoring 2. High Availability, Load Balancing, and Replication 3. Environment Stabilization via Parallel Deployment

Filled with practical, step-by-step instructions and clear explanations for the most important and useful tasks. A practical guide with

easy-to-follow recipes helping developers to quickly and effectively collect data from disparate sources such as databases, files, and applications, and turn the data into a unified format that is accessible and relevant to end users. Any IT professional working on PDI and is a valid support for either learning how to use the command line tools efficiently or for going deeper on some aspects of the command line tools to help you work better.

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.

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

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.

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.

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.

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

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.

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.

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.

"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"--

"Pentaho Data Integration (a.k.a. Kettle) is a full-featured open source ETL (Extract, Transform, and Load) solution. Although PDI is a feature-rich tool, effectively capturing, manipulating, cleansing, transferring, and loading data can get complicated. This book is full of practical examples that will help you to take advantage of Pentaho Data Integration's graphical, drag-and-drop design environment. You will quickly get started with Pentaho Data Integration by following the step-by-step guidance in this book. The useful tips in this book will encourage you to exploit powerful features of Pentaho Data Integration and perform ETL operations with ease."--Resource description p.

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.

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.

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.

Discussing new and existing features, SQL Server designer and administrator Michael Coles takes you on an expert guided tour of Transact-SQL functionality in SQL Server 2008 in his book, Pro T-SQL 2008 Programmer's Guide. Fully functioning examples and downloadable source code bring Coles' technically accurate and engaging treatment of Transact-SQL into your own hands. Step-by-step explanations ensure clarity, and an advocacy of best-practices will steer you down the road to success. Pro T-SQL 2008 Programmer's Guide is every developer's key to making full use of SQL Server 2008's powerful, built-in Transact-SQL language. Transact-SQL is the language developers and DBAs use to interact with SQL Server. It's used for everything from querying data, to writing stored procedures, to managing the database. New features in SQL Server 2008 include a spatial data type, SQLCLR integration, the MERGE statement, a dramatically improved and market-leading XML feature set, and support for encryption—all of which are covered in this book

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.

This second edition of the bestselling Learning XML provides web developers with a concise but grounded understanding of XML (the Extensible Markup Language) and its potential-- not just a whirlwind tour of XML. The author explains the important and relevant XML technologies and their capabilities clearly and succinctly with plenty of real-life projects and useful examples. He outlines the elements of markup--demystifying concepts such as attributes, entities, and namespaces--and provides enough depth and examples to get started. Learning XML is a reliable source for anyone who needs to know XML, but doesn't want to waste time wading through hundreds of web sites or 800 pages of bloated text. For writers producing XML documents, this book clarifies files and the process of creating them with the appropriate structure and format. Designers will learn what parts of XML are most helpful to their team and will get started on creating Document Type Definitions. For programmers, the book makes syntax and structures clear. Learning XML also discusses the stylesheets needed for viewing documents in the next generation of browsers, databases, and other devices. Learning XML illustrates the core XML concepts and language syntax, in addition to important related tools such as the CSS and XSL styling languages and the XLink and XPointer specifications for creating rich link structures. It includes information about three schema languages for validation: W3C Schema, Schematron, and RELAX-NG, which are gaining widespread support from people who need to validate documents but aren't satisfied with DTDs. Also new in this edition is a chapter on XSL-FO, a powerful formatting language for XML. If you need to wade through the acronym soup of XML and start to really use this powerful tool, Learning XML, will give you the roadmap you need.

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.

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, 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.

The book is a practical guide, full of step-by-step examples that are easy to follow and implement. This book is for developers, system administrators, and business intelligence professionals looking to learn how to get more out of their data through Pentaho. In order to best engage with the examples, some knowledge of Java will be required.

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.

Pentaho Data Integration Cookbook Second Edition Packt Publishing Ltd

SALIENT FEATURES OF BOOK • Easy to understand language with simple real life examples. • Primarily focused on Ecommerce and Retail industry. • Stepwise explanation of very basic to the complex of the statistical analysis. • All examples are solved using R and Excel or both. • Step by Step description of Pentaho BI Implementation - ETL, Report, OLAP & Dashboard. • Comprehensive coverage of topics including Probability Theories, Operation Analytics, Digital Marketing, BI Implementation, Web Analytics, Forecasting, Customer Analytics & Optimization. Application focus on current Digital Analytics Practices.

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.

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.

"PostgreSQL" leads users through the internals of an open-source database. Throughout the book are explanations of data structures and algorithms, each backed by a concrete example from the actual source code. Each section contains information about performance implications, debugging techniques, and pointers to more information (on the Web and in book form). Over 70 practical recipes to gain operational data intelligence with Splunk Enterprise About This Book This is the most up-to-date book on Splunk 6.3 and teaches you how to tackle real-world operational intelligence scenarios efficiently Get business insights using machine data using this easy-to-follow guide Search, monitor, and analyze your operational data skillfully using this recipe-based, practical guide Who This Book Is For This book is intended for users of all levels who are looking to leverage the Splunk Enterprise platform as a valuable operational intelligence tool. The recipes provided in this book will appeal to individuals from all facets of business, IT, security, product, marketing, and many more! Also, existing users of Splunk who want to upgrade and get up and running with Splunk 6.3 will find this book invaluable. What You Will Learn Use Splunk to gather, analyze, and report on data Create dashboards and visualizations that make data meaningful Build an operational intelligence application with extensive features and functionality Enrich operational data with lookups and workflows Model and accelerate data and perform pivot-based reporting Build real-time, scripted, and other intelligence-driven alerts Summarize data for longer term trending, reporting, and analysis Integrate advanced JavaScript charts and leverage Splunk's API In Detail Splunk makes it easy for you to take control of your data, and with Splunk Operational Cookbook, you can be confident that you are taking advantage of the Big Data revolution and driving your business with the cutting edge of operational intelligence and business analytics. With more than 70 recipes that demonstrate all of Splunk's features, not only will you find quick solutions to common problems, but you'll also learn a wide range of strategies and uncover new ideas that will make you rethink what operational intelligence means to you and your organization. You'll discover recipes on data processing, searching and reporting, dashboards, and visualizations to make data shareable, communicable, and most importantly meaningful. You'll also find step-by-step demonstrations that walk you through building an operational intelligence application containing vital features essential to understanding data and to help you successfully integrate a data-driven way of thinking in your organization. Throughout the book, you'll dive deeper into Splunk, explore data models and pivots to extend your intelligence capabilities, and perform advanced searching to explore your data in even more sophisticated ways. Splunk is changing the business landscape, so make sure you're taking advantage of it. Style and approach Splunk is an excellent platform that allows you to make sense of machine data with ease. The adoption of Splunk has been huge and everyone who has gone beyond installing Splunk wants to know how to make most of it. This book will not only teach you how to use Splunk in real-world scenarios to get business insights, but will also get existing Splunk users up to date with the latest Splunk 6.3 release.

Learn how to transition from Excel-based business intelligence (BI) analysis to enterprise stacks of open-source BI tools. Select and implement the best free and freemium open-source BI tools for your company's needs and design, implement, and integrate BI automation across the full stack using agile methodologies. Business Intelligence Tools for Small Companies provides hands-on demonstrations of open-source tools suitable for the BI requirements of small businesses. The authors draw on their deep experience as BI consultants, developers, and administrators to guide you through the extract-transform-load/data warehousing (ETL/DWH) sequence of extracting data from an enterprise resource planning (ERP) database freely available on the Internet, transforming the data, manipulating them, and loading them into a relational database. The authors demonstrate how to extract, report, and dashboard key performance indicators (KPIs) in a visually appealing format from the relational database management system (RDBMS). They model the selection and implementation of free and freemium tools such as Pentaho Data Integrator and Talend for ELT, Oracle XE and MySQL/MariaDB for RDBMS, and QlikSense, Power BI, and MicroStrategy Desktop for reporting. This richly illustrated guide models the deployment of a small company BI stack on an inexpensive cloud platform such as AWS.

What You'll Learn You will learn how to manage, integrate, and automate the processes of BI by selecting and implementing tools to: Implement and manage the business intelligence/data warehousing (BI/DWH) infrastructure Extract data from any enterprise resource planning (ERP) tool Process and integrate BI data using open-source extract-transform-load (ETL) tools Query, report, and analyze BI data using open-source visualization and dashboard tools Use a MOLAP tool to define next year's budget, integrating real data with target scenarios Deploy BI solutions and big data experiments inexpensively on cloud platforms Who This Book Is For Engineers, DBAs, analysts, consultants, and managers at small companies with limited resources but whose BI requirements have outgrown the limitations of Excel spreadsheets; personnel in mid-sized companies with established BI systems who are exploring technological updates and more cost-efficient solutions

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.

A step-by-step tutorial on implementing Liferay- based portals to learn performance best practices. The book is good for Liferay portal developers and architects who want to learn performance best practices for implementing Liferay- based solutions. It is assumed that you have a working knowledge of the Liferay portal.

This book is for developers who want to learn how to get the most out of Solr in their applications, whether you are new to the field, have used Solr but don't know everything, or simply want a good reference. It would be helpful to have some familiarity with basic programming concepts, but no prior experience is required.

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.

Managing Data in Motion describes techniques that have been developed for significantly reducing the complexity of managing system interfaces and enabling scalable architectures. Author April Reeve brings over two decades of experience to present a vendor-neutral approach to moving data between computing environments and systems. Readers will learn the techniques,

technologies, and best practices for managing the passage of data between computer systems and integrating disparate data together in an enterprise environment. The average enterprise's computing environment is comprised of hundreds to thousands computer systems that have been built, purchased, and acquired over time. The data from these various systems needs to be integrated for reporting and analysis, shared for business transaction processing, and converted from one format to another when old systems are replaced and new systems are acquired. The management of the "data in motion" in organizations is rapidly becoming one of the biggest concerns for business and IT management. Data warehousing and conversion, real-time data integration, and cloud and "big data" applications are just a few of the challenges facing organizations and businesses today. Managing Data in Motion tackles these and other topics in a style easily understood by business and IT managers as well as programmers and architects. Presents a vendor-neutral overview of the different technologies and techniques for moving data between computer systems including the emerging solutions for unstructured as well as structured data types Explains, in non-technical terms, the architecture and components required to perform data integration Describes how to reduce the complexity of managing system interfaces and enable a scalable data architecture that can handle the dimensions of "Big Data"

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

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.

[Copyright: 6a7220d215c9f58ab014cf6470cc7471](https://www.pentaho.com/copyright)