# Getting Started With Talend Open Studio For Data Integration Bowen Jonathan

This book is for developers who want an alternative way to store and process data within their applications. No previous graph database experience is required; however, some basic database knowledge will help you understand the concepts more easily. A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key Features Get to grips with the newly introduced features and capabilities of Hadoop 3 Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystem Sharpen your Hadoop skills with real-world case studies and code Book Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data processing platform, with a focus on improved scalability and increased efficiency. With this guide, you'll understand advanced concepts of the Hadoop ecosystem tool. You'll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You'll be able to address common challenges like using Kafka efficiently, designing low latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you'll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you'll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you'll be equipped to tackle a range of real-world problems in data pipelines. What you will learn Gain an in-depth understanding of distributed computing using Hadoop 3 Develop enterprise-grade applications using Apache Spark, Flink, and more Build scalable and high-performance Hadoop data pipelines with security, monitoring, and data governance Explore batch data processing patterns and how to model data in Hadoop Master best practices for enterprises using, or planning to use, Hadoop 3 as a data platform Understand security aspects of Hadoop, including authorization and authentication Who this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

Getting Started with Talend Open Studio for Data IntegrationPackt Publishing Ltd

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 Started1.Introduction to Open Source Software, its advantages and disadvantages2.Moving from Licensed software to Open Source software3.Basics and the history of Microsoft SQL Server and PostgreSQL4.Choosing the right PostgreSQL Edition and Version for your needs5.Production, Development & Testing Workloads6.On-Premise vs Cloud comparisonChapter 2: Platform Comparison1.In-depth comparison of SQL Server and PostgreSQL2.Concurrency Control3.Database Structure comparison4.Feature comparison5.Data Type Mapping6.Built-In Functions and Operators7.SQL Syntax differences8.Areas where PostgreSQL has an edge9.Areas where SQL Server has an edge10.GUI Client ToolsChapter 3: The Actual Migration1.Schema migration2.Data migration3.Code migration4.ETL/ELT Tools5.Reporting Platforms6.Analytics Platforms7.Scheduling Options8.Application/Services MigrationChapter 4: Maintenance and Monitoring1.Maintenance and Monitoring2.High Availability, Load Balancing, and Replication3.Environment Stabilization via Parallel Deployment

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

Learn how to write, tune, and port SQL queries and other statements for a Big Data environment, using Impala—the massively parallel processing SQL query engine for Apache Hadoop. The best practices in this practical guide help you design database schemas that not only interoperate with other Hadoop components, and are convenient for administers to manage and monitor, but also accommodate future expansion in data size and evolution of software capabilities. Written by John Russell, documentation lead for the Cloudera Impala project, this book gets you working with the most recent Impala releases quickly. Ideal for database developers and business analysts, the latest revision covers analytics functions, complex types, incremental statistics, subqueries, and submission to the Apache incubator. Getting Started with Impala includes advice from Cloudera's development team, as well as insights from its consulting engagements with customers. Learn how Impala integrates with a wide range of Hadoop components Attain high performance and scalability for huge data sets on production clusters Explore common developer tasks, such as porting code to Impala and optimizing performance Use tutorials for working with billion-row tables, date- and time-based values, and other techniques Learn how to transition from rigid schemas to a flexible model that evolves as needs change Take a deep dive into joins and the roles of statistics

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you.

With this textbook, Vaisman and Zimányi deliver excellent coverage of data warehousing and business intelligence technologies

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

NoSQL Starter is a great resource for someone starting with NoSQL and an indispensable guide for technology decision makers. It is assumed that you have a background in RDBMS modeling and SQL and have had exposure to at least one of the programming languages – Java or JavaScript.Friendly, practical tutorial with lots of hints and tips from several experienced Solr users and developers.

The way developers design, build, and run software has changed significantly with the evolution of microservices and containers. These modern architectures use new primitives that require a different set of practices than most developers, tech leads, and architects are accustomed to. With this focused guide, Bilgin Ibryam and Roland Huß from Red Hat provide common reusable elements, patterns, principles, and practices for designing and implementing cloud-native applications on Kubernetes. Each pattern includes a description of the problem and a proposed solution with Kubernetes specifics. Many patterns are also backed by concrete code examples. This book is ideal for developers already familiar with basic Kubernetes concepts who want to learn common cloud native patterns. You'll learn about the following pattern categories: Foundational patterns cover the core principles and practices for building container-based cloud-native applications. Behavioral patterns explore finer-grained concepts for managing various types of container and platform interactions. Structural patterns help you organize containers within a pod, the atom of the Kubernetes platform. Configuration patterns provide insight into how application configurations can be handled in Kubernetes. Advanced patterns covers more advanced topics such as extending the platform with operators.

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.

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"

Explore the modern market of data analytics platforms and the benefits of using Snowflake computing, the data warehouse built for the cloud. With the rise of cloud technologies, organizations prefer to deploy their analytics using cloud providers such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform. Cloud vendors are offering modern data platforms for building cloud analytics solutions to collect data and consolidate into single storage solutions that provide insights for business users. The core of any analytics framework is the data warehouse, and previously customers did not have many choices of platform to use. Snowflake was built specifically for the cloud and it is a true game changer for the analytics market. This book will

help onboard you to Snowflake, present best practices to deploy, and use the Snowflake data warehouse. In addition, it covers modern analytics architecture and use cases. It provides use cases of integration with leading analytics software such as Matillion ETL, Tableau, and Databricks. Finally, it covers migration scenarios for on-premise legacy data warehouses. What You Will Learn Know the key functionalities of Snowflake Set up security and access with cluster Bulk load data into Snowflake using the COPY command Migrate from a legacy data warehouse to Snowflake integrate the Snowflake data platform with modern business intelligence (BI) and data integration tools Who This Book Is For Those working with data warehouse and business intelligence (BI) technologies, and existing and potential Snowflake users

The database field has experienced a rapid and incessant growth since the development of relational databases. The progress in database systems and applications has produced a diverse landscape of specialized technology areas that have often become the exclusive domain of research specialists. Examples include active databases, temporal databases, object-oriented databases, deductive databases, imprecise reasoning and queries, and multimedia information systems. This book provides a systematic introduction to and an in-depth treatment of these advanced database areas. It supplies practitioners and researchers with authoritative coverage of recent technological advances that are shaping the future of commercial database systems and intelligent information systems. Advanced Database Systems was written by a team of six leading specialists who have made significant contributions to the development of the technology areas covered in the book. Benefiting from the authors' long experience teaching graduate and professional courses, this book is designed to provide a gradual introduction to advanced research topics and includes many examples and exercises to support its use for individual study, desk reference, and graduate classroom teaching.

Summary CMIS and Apache Chemistry in Action is a comprehensive guide to the CMIS standard and related ECM concepts, written by the authors of the standard. In it, you'll tackle hands-on examples for building applications on CMIS repositories from both the client and the server sides. You'll learn how to create new content-centric applications that install and run in any CMIS-compliant repository. About The Technology Content Management Interoperability Services (CMIS) is an OASIS standard for accessing content management systems. It specifies a vendor-and language-neutral way to interact with any compliant content repository. Apache Chemistry provides complete reference implementations of the CMIS standard with robust APIs for developers writing tools, applications, and servers. About This Book CMIS and Apache Chemistry in Action is a comprehensive guide to the CMIS standard and related ECM concepts. In it, you'll find clear teaching and instantly useful examples for building content-centric client and server-side applications that run against any CMIS-compliant repository. In fact, using the CMIS Workbench and the InMemory Repository from Apache Chemistry, you'll have running code talking to a real CMIS server by the end of chapter 1. This book requires some familiarity with content management systems and a standard programming language like Java or C#. No exposure to CMIS or Apache Chemistry is assumed. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's Inside The only CMIS book endorsed by OASIS Complete coverage of the CMIS 1.0 and 1.1 specifications Cookbook-style tutorials and real-world examples About the Authors Florian Müller, Jay Brown, and Jeff Potts are among the original authors, contributors, and leaders of Apache Chemistry and the OASIS CMIS specification. They continue to shape CMIS implementations at Alfresco, IBM, and SAP. Table of Contents PART 1 UNDERSTANDING CMIS Introducing CMIS Exploring the CMIS domain model Creating, updating, and deleting objects with CMIS CMIS metadata: types and properties Query PART 2 HANDS-ON CMIS CLIENT DEVELOPMENT Meet your new project: The Blend The Blend: read and query functionality The Blend: create, update, and delete functionality Using other client libraries Building mobile apps with CMIS PART 3 ADVANCED TOPICS CMIS bindings Security and control Performance Building a CMIS server

Exploit the visualization capabilities of Kibana and build powerful interactive dashboards About This Book Introduction to data-driven architecture and the Elastic stack Build effective dashboards for data visualization and explore datasets with Elastic Graph A comprehensive guide to learning scalable data visualization techniques in Kibana Who This Book Is For If you are a developer, data visualization engineer, or data scientist who wants to get the best of data visualization at scale then this book is perfect for you. A basic understanding of Elasticsearch and Logstash is required to make the best use of this book. What You Will Learn How to create visualizations in Kibana Ingest log data, structure an Elasticsearch cluster, and create visualization assets in Kibana Embed Kibana visualization on web pages Scaffold, develop, and deploy new Kibana & Timelion customizations Build a metrics dashboard in Timelion based on time series data Use the Graph plugin visualization feature and leverage a graph query Create, implement, package, and deploy a new custom plugin Use Prelert to solve anomaly detection challenges In Detail Kibana is an open source data visualization platform that allows you to interact with your data through stunning, powerful graphics. Its simple, browser-based interface enables you to quickly create and share dynamic dashboards that display changes to Elasticsearch queries in real time. In this book, you'll learn how to use the Elastic stack on top of a data architecture to visualize data in real time. All data architectures have different requirements and expectations when it comes to visualizing the data, whether it's logging analytics, metrics, business analytics, graph analytics, or scaling them as per your business requirements. This book will help you master Elastic visualization tools and adapt them to the requirements of your project. You will start by learning how to use the basic visualization features of Kibana 5. Then you will be shown how to implement a pure metric analytics architecture and visualize it using Timelion, a very recent and trendy feature of the Elastic stack. You will learn how to correlate data using the brand-new Graph visualization and build relationships between documents. Finally, you will be familiarized with the setup of a Kibana development environment so that you can build a custom Kibana plugin. By the end of this book you will have all the information needed to take your Elastic stack skills to a new level of data visualization. Style and approach This book takes a comprehensive, step-by-step approach to working with the visualization aspects of the Elastic stack. Every concept is presented in a very easy-to-follow manner that shows you both the logic and method of implementation. Real world cases are referenced to highlight how each of the key concepts can be put to practical use.

Work with petabyte-scale datasets while building a collaborative, agile workplace in the process. This practical book is the canonical reference to Google BigQuery, the query engine that lets you conduct interactive analysis of large datasets. BigQuery enables enterprises to efficiently store, query, ingest, and learn from their data in a convenient framework. With this book, you'll examine how to analyze data at scale to derive insights from large datasets efficiently. Valliappa Lakshmanan, tech lead for Google Cloud Platform, and Jordan Tigani, engineering director for the BigQuery team, provide best practices for modern data warehousing within an autoscaled, serverless public cloud. Whether you want to

explore parts of BigQuery you're not familiar with or prefer to focus on specific tasks, this reference is indispensable. A problem-solution-based guide to help you overcome hurdles effectively while working with kernel APIs, filesystems, networks, threads, and process communications Key Features Learn to apply the latest C++ features (from C++11, 14, 17, and 20) to facilitate systems programming Create robust and concurrent systems that make the most of the available hardware resources Delve into C++ inbuilt libraries and frameworks to design robust systems as per your business needs Book Description C++ is the preferred language for system programming due to its efficient low-level computation, data abstraction, and object-oriented features. System programming is about designing and writing computer programs that interact closely with the underlying operating system and allow computer hardware to interface with the programmer and the user. The C++ System Programming Cookbook will serve as a reference for developers who want to have ready-to-use solutions for the essential aspects of system programming using the latest C++ standards wherever possible. This C++ book starts out by giving you an overview of system programming and refreshing your C++ knowledge. Moving ahead, you will learn how to deal with threads and processes, before going on to discover recipes for how to manage memory. The concluding chapters will then help you understand how processes communicate and how to interact with the console (console I/O). Finally, you will learn how to deal with time interfaces, signals, and CPU scheduling. By the end of the book, you will become adept at developing robust systems applications using C++. What you will learn Get up to speed with the fundamentals including makefile, man pages, compilation, and linking and debugging Understand how to deal with time interfaces, signals, and CPU scheduling Develop your knowledge of memory management Use processes and threads for advanced synchronizations (mutexes and condition variables) Understand interprocess communications (IPC): pipes, FIFOs, message queues, shared memory, and TCP and UDP Discover how to interact with the console (console I/O) Who this book is for This book is for C++ developers who want to gain practical knowledge of systems programming. Though no experience of Linux system programming is assumed, intermediate knowledge of C++ is necessary.

Summary Camel in Action, Second Edition is the most complete Camel book on the market. Written by core developers of Camel and the authors of the highly acclaimed first edition, this book distills their experience and practical insights so that you can tackle integration tasks like a pro. Forewords by James Strachan and Dr. Mark Little Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Apache Camel is a Java framework that implements enterprise integration patterns (EIPs) and comes with over 200 adapters to third-party systems. A concise DSL lets you build integration logic into your app with just a few lines of Java or XML. By using Camel, you benefit from the testing and experience of a large and vibrant open source community. About the Book Camel in Action, Second Edition is the definitive guide to the Camel framework. It starts with core concepts like sending, receiving, routing, and transforming data. It then goes in depth on many topics such as how to develop, debug, test, deal with errors, secure, scale, cluster, deploy, and monitor your Camel applications. The book also discusses how to run Camel with microservices, reactive systems, containers, and in the cloud. What's Inside Coverage of all relevant EIPs Camel microservices with Spring Boot Camel on Docker and Kubernetes Error handling, testing, security, clustering, monitoring, and deployment Hundreds of examples in Java and XML About the Reader Readers should be familiar with Java. This book is accessible to beginners and invaluable to experts. About the Author Claus Ibsen is a senior principal engineer working for Red Hat specializing in cloud and integration. He has worked on Apache Camel for the last nine years where he heads the project. Claus lives in Denmark. Jonathan Anstey is an engineering manager at Red Hat and a core Camel contributor. He lives in Newfoundland, Canada. Table of Contents Part 1 - First steps Meeting Camel Routing with Camel Part 2 - Core Camel Transforming data with Camel Using beans with Camel Enterprise integration patterns Using components Part 3 - Developing and testing Microservices Developing Camel projects Testing RESTful web services Part 4 - Going further with Camel Error handling Transactions and idempotency Parallel processing Securing Camel Part 5 - Running and managing Camel Running and deploying Camel Management and monitoring Part 6 - Out in the wild Clustering Microservices with Docker and Kubernetes Camel tooling Bonus online chapters Available at https://www.manning.com/books/camel-in-?action-second-edition and in electronic versions of this book: Reactive Camel Camel and the IoT by Henryk Konsek

If you are looking to build next-generation AI solutions for work or even for your pet projects, you'll find this cookbook useful. With the help of easy-to-follow recipes, this book will take you through the advanced AI and machine learning approaches and algorithms that are required to build smart models for problem-solving.

Meira Amon is the most reserved of the phoenixes, finding computers easier to deal with than people and their messy emotions. But she wants revenge against the Rotting King Pytheios for the murder of her parents just as much as her sisters. Offering herself as a queen for their ally, dragon shifter King Gorgon of the Black Clan, only makes sense. Just one problem... Samael Veles worked his lowborn ass off to become the fiercest warrior of the Black Clan. He has pledged his life to protect his king at all costs. Yet somehow, in the middle of war, he took one look at a woman in a reflection and gave up his heart. When that woman turns out to be the phoenix promised to mate his king and help bring peace to their kind, he has no choice but to accept it. Until her fire unexpectedly destroys the king after they've wed. Now Samael must choose between his loyalty to his clan and protecting the queen who might be his mate from his people's wrath, one of which will unlock a destiny no one could have ever imagined... Each book in the Inferno Rising series is STANDALONE: * The Rogue King * The Blood King * The Warrior King

Clear your doubts about Business Intelligence and start your new journey KEY FEATURES ? Includes successful methods and innovative ideas to achieve success with BI. ? Vendor-neutral, unbiased, and based on experience. ? Highlights practical challenges in BI journeys. ? Covers financial aspects along with technical aspects. ? Showcases multiple BI organization models and the structure of BI teams. DESCRIPTION The book demystifies misconceptions and

misinformation about BI. It provides clarity to almost everything related to BI in a simplified and unbiased way. It covers topics right from the definition of BI, terms used in the BI definition, coinage of BI, details of the different main uses of BI, processes that support the main uses, side benefits, and the level of importance of BI, various types of BI based on various parameters, main phases in the BI journey and the challenges faced in each of the phases in the BI journey. It clarifies myths about self-service BI and real-time BI. The book covers the structure of a typical internal BI team, BI organizational models, and the main roles in BI. It also clarifies the doubts around roles in BI. It explores the different components that add to the cost of BI and explains how to calculate the total cost of the ownership of BI and ROI for BI. It covers several ideas, including unconventional ideas to achieve BI success and also learn about IBI. It explains the different types of BI architectures, commonly used technologies, tools, and concepts in BI and provides clarity about the boundary of BI w.r.t technologies, tools, and concepts. The book helps you lay a very strong foundation and provides the right perspective about BI. It enables you to start or restart your journey with BI. WHAT YOU WILL LEARN ? Builds a strong conceptual foundation in BI. ? Gives the right perspective and clarity on BI uses, challenges, and architectures. ? Enables you to make the right decisions on the BI structure, organization model, and budget. ? Explains which type of BI solution is required for your business. ? Applies successful BI ideas. WHO THIS BOOK IS FOR This book is a must-read for business managers, BI aspirants, CxOs, and all those who want to drive the business value with data-driven insights. TABLE OF CONTENTS 1. What is Business Intelligence? 2. Why do Businesses need BI? 3. Types of Business Intelligence 4. Challenges in Business Intelligence 5. Roles in Business Intelligence 6. Financials of Business Intelligence 7. Ideas for Success with BI 8. Introduction to IBI 9. BI Architectures 10. Demystify Tech, Tools, and Concepts in BI Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

The inspirational bestseller that ignited a movement and asked us to find our WHY Discover the book that is captivating millions on TikTok and that served as the basis for one of the most popular TED Talks of all time—with more than 56 million views and counting. Over a decade ago, Simon Sinek started a movement that inspired millions to demand purpose at work, to ask what was the WHY of their organization. Since then, millions have been touched by the power of his ideas, and these ideas remain as relevant and timely as ever. START WITH WHY asks (and answers) the questions: why are some people and organizations more innovative, more influential, and more profitable than others? Why do some command greater loyalty from customers and employees alike? Even among the successful, why are so few able to repeat their success over and over? People like Martin Luther King Jr., Steve Jobs, and the Wright Brothers had little in common, but they all started with WHY. They realized that people won't truly buy into a product, service, movement, or idea until they understand the WHY behind it. START WITH WHY shows that the leaders who have had the greatest influence in the world all think, act and communicate the same way—and it's the opposite of what everyone else does. Sinek calls this powerful idea The Golden Circle, and it provides a framework upon which organizations can be built, movements can be led, and people can be inspired. And it all starts with WHY.

This practical book covers both strategies and tactics around managing a data governance initiative to help make the most of your data.

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.

The process of talent development (TD) is essential to success in any sport. Drawing on the latest evidence and a considerable experience base, this book dispels myths about talent development and offers practical advice on the TD pathway from pre-school to elite level. Aimed at practitioners and other stakeholders involved in the TD process – including coaches, scientists, administrators, educators, students, parents, policy makers and senior development athletes – this is the only up-to-date practical guide to TD in sport. Written by experts with more than 20 years' experience in TD training, coaching and research, it covers key topics from deliberate practice and fundamental movement skills to designing and managing a TD pathway. It also includes contributions from professionals working in a wide range of sports, providing real-world insights into important topics including: the recruitment process academy and apprenticeship preparation the coach–athlete relationship what to do to stay ahead considerations for parents and coaches. Talent Development: A Practitioner Guide is an indispensable resource for all those interested in talent identification, talent development and coaching practice in elite sport.

Solve business challenges with Microsoft Power BI's advanced visualization and data analysis techniques Key Features Create effective storytelling reports by implementing simple-to-intermediate Power BI features Develop powerful analytical models to extract key insights for changing business needs Build, publish, and share impressive dashboards for your organization Book Description To succeed in today's transforming business world, organizations need business intelligence capabilities to make smarter decisions faster than ever before. This Power BI book is an entry-level guide that will get you up and running with data modeling, visualization, and analytical techniques from scratch. You'll find this book handy if you want to get well-versed with the extensive Power BI ecosystem. You'll start by covering the basics of business intelligence and installing Power BI. You'll then learn the wide range of Power BI features to unlock business insights. As you progress, the book will take you through how to use Power Query to ingest, cleanse, and shape your data, and use Power BI DAX to create simple to complex calculations. You'll also be able to add a variety of interactive visualizations to your reports to bring your data to life. Finally, you'll gain hands-on experience in creating visually stunning reports that speak to business decision makers, and see how you can securely share these reports and collaborate with others. By the end of this book, you'll be ready to create simple, yet effective, BI reports and dashboards using the latest features of Power BI. What you will learn Explore the different features of Power BI to create interactive dashboards Use the Query Editor to import and transform data Perform simple and complex DAX calculations to enhance analysis Discover business insights and tell a story with your data using Power BI Explore data and learn to manage datasets, dataflows, and data gateways Use workspaces to collaborate with others and publish your reports Who this book is for If you're an IT manager, data analyst, or BI user new to using Power BI for solving business intelligence problems, this book is for you. You'll also find this book useful if you want to migrate from other BI tools to create powerful and interactive dashboards. No experience of working with Power BI is expected.

This book constitutes the refereed proceedings of the First International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability, ARTIIS 2021, held in La Libertad, Ecuador, in November 2021. The 53 full papers and 2 short contributions were carefully reviewed and selected from 155 submissions. The volume covers a variety of topics, such as computer systems organization, software engineering, information storage and retrieval, computing methodologies, artificial intelligence, and others. The papers are logically organized in the following thematic blocks: ?Computing Solutions; Data Intelligence; Ethics, Security, and Privacy; Sustainability.

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.

Would you like to use a consistent visual notation for drawing integration solutions? "Look inside the front cover." Do you want to harness the power of asynchronous systems without getting caught in the pitfalls? "See "Thinking Asynchronously" in the Introduction." Do you want to know which style of application integration is best for your purposes? "See Chapter 2, Integration Styles." Do you want to learn techniques for processing messages concurrently? "See Chapter 10, Competing Consumers and Message Dispatcher." Do you want to learn how you can track asynchronous messages as they flow across distributed systems? "See Chapter 11, Message History and Message Store." Do you want to understand how a system designed using integration patterns can be implemented using Java Web services, .NET message queuing, and a TIBCO-based publish-subscribe architecture? "See Chapter 9, Interlude: Composed Messaging." Utilizing years of practical experience, seasoned experts Gregor Hohpe and Bobby Woolf show how asynchronous messaging has proven to be the best strategy for enterprise integration success. However, building and deploying messaging solutions presents a number of problems for developers. " Enterprise Integration Patterns " provides an invaluable catalog of sixty-five patterns, with real-world solutions that demonstrate the formidable of messaging and help you to design effective messaging solutions for your enterprise. The authors also include examples covering a variety of different integration technologies, such as JMS, MSMQ, TIBCO ActiveEnterprise, Microsoft BizTalk, SOAP, and XSL. A case study describing a bond trading system illustrates the patterns in practice, and the book offers a look at emerging standards, as well as insights into what the future of enterprise integration might hold. This book provides a consistent vocabulary and visual notation framework to describe large-scale integration solutions across many technologies. It also explores in detail the advantages and limitations of asynchronous messaging architectures. The authors present practical advice on designing code that connects an application to a messaging system, and provide extensive information to help you determine when to send a message, how to route it to the proper destination, and how to monitor the health of a messaging system. If you want to know how to manage, monitor, and maintain a messaging system once it is in use, get this book. 0321200683B09122003

The Cambridge Advanced Learner's Dictionary gives the vital support which advanced students need, especially with the essential skills: reading, writing, listening and speaking. In the book: * 170,000 words, phrases and examples * New words: so your English stays up-to-date * Colour headwords: so you can find the word you are looking for quickly * Idiom Finder * 200 'Common Learner Error' notes show how to avoid common mistakes * 25,000 collocations show the way words work together * Colour pictures: 16 full page colour pictures On the CD-ROM: * Sound: recordings in British and American English, plus practice tools to help improve pronunciation * UNIQUE! Smart Thesaurus helps you choose the right word * QUICKfind looks up words for you while you are working or reading on screen * UNIQUE! SUPERwrite gives on screen help with grammar, spelling and collocation when you are writing * Hundreds of interactive exercises

This book is not intended to be a deep-dive on security, programming, reporting, configuration, etc. but more of a high-level

overview on salesforce.com, Force.com development, tools, methodology and other resources. We will not cover all aspects of salesforce.com. We wrote this book to (hopefully) become an invaluable resource for anyone new or inexperienced with the Force.com platform. For new administrators and developers coming on to the platform, we see a gaping hole in existing reference books. This book is not intended for end users or advanced developers. This book is for people somewhere in the middle; let's say a business analyst or a new Java or .NET developer. We would like this to be the first book you pick up after your boss walks in and says, "Let's take a look at salesforce.com and see if we can build some applications in the cloud."This book contains the information that we wish we had when we started developing applications on Force.com.

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

The 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.

A practical cookbook on building portals with GateIn including user security, gadgets, and every type of portlet possible.

The real challenge of programming isn't learning a language's syntax—it's learning to creatively solve problems so you can build something great. In this one-of-a-kind text, author V. Anton Spraul breaks down the ways that programmers solve problems and teaches you what other introductory books often ignore: how to Think Like a Programmer. Each chapter tackles a single programming concept, like classes, pointers, and recursion, and open-ended exercises throughout challenge you to apply your knowledge. You'll also learn how to: –Split problems into discrete components to make them easier to solve –Make the most of code reuse with functions, classes, and libraries –Pick the perfect data structure for a particular job –Master more advanced programming tools like recursion and dynamic memory –Organize your thoughts and develop strategies to tackle particular types of problems Although the book's examples are written in C++, the creative problem-solving concepts they illustrate go beyond any particular language; in fact, they often reach outside the realm of computer science. As the most skillful programmers know, writing great code is a creative art—and the first step in creating your masterpiece is learning to Think Like a Programmer.

This book is written in a concise and easytounderstand manner, and acts as a comprehensive guide on data analytics and integration with Talend big data processing jobs.If you are a chief information officer, enterprise architect, data architect, data scientist, software developer, software engineer, or a data analyst who is familiar with data processing projects and who wants to use Talend to get your first big data job executed in a reliable, quick, and graphical way, then Talend for Big Data is perfect for you.

Primarily designed as a reference book, simple and effective exercises based upon genuine real-world tasks enable the developer to reduce the time to deliver the results. Presentation of the activities in a recipe format will enable the readers to grasp even the complex concepts with consummate ease.Talend Open Studio Cookbook is principally aimed at relative beginners and intermediate Talend Developers who have used the product to perform some simple integration tasks, possibly via a training course or beginner's tutorials.

Named by The Washington Post as one of the 11 Leadership Books to Read in 2018 When it comes to recruiting, motivating, and creating great teams, Patty McCord says most companies have it all wrong. McCord helped create the unique and high-performing culture at Netflix, where she was chief talent officer. In her new book, Powerful: Building a Culture of Freedom and Responsibility, she shares what she learned there and elsewhere in Silicon Valley. McCord advocates practicing radical honesty in the workplace, saying good-bye to employees who don't fit the company's emerging needs, and motivating with challenging work, not promises, perks, and bonus plans. McCord argues that the old standbys of corporate HR—annual performance reviews, retention plans, employee empowerment and engagement programs—often end up being a colossal waste of time and resources. Her road-tested advice, offered with humor and irreverence, provides readers a different path for creating a culture of high performance and profitability. Powerful will change how you think about work and the way a business should be run.