

Apache Cookbook

Over 80 recipes to simplify machine learning model implementations with Spark

About This Book

- Solve the day-to-day problems of data science with Spark
- This unique cookbook consists of exciting and intuitive numerical recipes
- Optimize your work by acquiring, cleaning, analyzing, predicting, and visualizing your data

Who This Book Is For

This book is for Scala developers with a fairly good exposure to and understanding of machine learning techniques, but lack practical implementations with Spark. A solid knowledge of machine learning algorithms is assumed, as well as hands-on experience of implementing ML algorithms with Scala. However, you do not need to be acquainted with the Spark ML libraries and ecosystem.

What You Will Learn

- Get to know how Scala and Spark go hand-in-hand for developers when developing ML systems with Spark
- Build a recommendation engine that scales with Spark
- Find out how to build unsupervised clustering systems to classify data in Spark
- Build machine learning systems with the Decision Tree and Ensemble models in Spark
- Deal with the curse of high-dimensionality in big data using Spark
- Implement Text analytics for Search Engines in Spark
- Streaming Machine Learning System implementation using Spark

In Detail

Machine learning aims to extract knowledge from data, relying on fundamental concepts in computer science, statistics, probability, and optimization. Learning about algorithms enables a wide range of applications, from everyday tasks such as product recommendations and spam

filtering to bleeding edge applications such as self-driving cars and personalized medicine. You will gain hands-on experience of applying these principles using Apache Spark, a cluster computing system well suited for large-scale machine learning tasks. This book begins with a quick overview of setting up the necessary IDEs to facilitate the execution of code examples that will be covered. It also highlights some key issues developers face while thinking about Scala for machine learning and during the switch over to Spark. We progress by uncovering the various Spark APIs and the implementation of ML algorithms with developing classification systems, recommendation engines, clustering and learning systems. Towards the final chapters, we'll focus on building high-end applications and explain various unsupervised methodologies and challenges to tackle when implementing with big data ML systems.

Get a hands-on introduction to the Chef, the configuration management tool for solving operations issues in enterprises large and small. Ideal for developers and sysadmins new to configuration management, this guide shows you to automate the packaging and delivery of applications in your infrastructure. You'll be able to build (or rebuild) your infrastructure's application stack in minutes or hours, rather than days or weeks. After teaching you how to write Ruby-based Chef code, this book walks you through different Chef tools and configuration management concepts in each chapter, using detailed examples throughout. All you need to get started is

command-line experience and familiarity with basic system administration. Configure your Chef development environment and start writing recipes Create Chef cookbooks with recipes for each part of your infrastructure Use Test Kitchen to manage sandbox testing environments Manage single nodes with Chef client, and multiple nodes with Chef Server Use data bags for storing shared global data between nodes Simulate production Chef Server environments with Chef Zero Classify different types of services in your infrastructure with roles Model life stages of your application, including development, testing, staging, and production

This book is for intermediate Solr Developers who are willing to learn and implement Pro-level practices, techniques, and solutions. This edition will specifically appeal to developers who wish to quickly get to grips with the changes and new features of Apache Solr 5. Over 50 hands-on recipes to efficiently administer, maintain, and use your Apache Kafka installationAbout This Book- Quickly configure and manage your Kafka cluster- Learn how to use the Apache Kafka cluster and connect it with tools for big data processing- A practical guide to monitor your Apache Kafka installationWho This Book Is ForIf you are a programmer or big data engineer using or planning to use Apache Kafka, then this book is for you. This book has several recipes which will teach you how to effectively use Apache Kafka. You need to have some basic knowledge of Java. If you don't know big data tools, this would be your stepping stone for learning how to consume the data in these kind of

systems. What You Will Learn- Learn how to configure Kafka brokers for better efficiency- Explore how to configure producers and consumers for optimal performance- Set up tools for maintaining and operating Apache Kafka- Create producers and consumers for Apache Kafka in Java- Understand how Apache Kafka can be used by several third party system for big data processing, such as Apache Storm, Apache Spark, Hadoop, and more- Monitor Apache Kafka using tools like graphite and GangliaIn DetailThis book will give you details about how to manage and administer your Apache Kafka Cluster. We will cover topics like how to configure your broker, producer, and consumer for maximum efficiency for your situation. Also, you will learn how to maintain and administer your cluster for fault tolerance. We will also explore tools provided with Apache Kafka to do regular maintenance operations. We shall also look at how to easily integrate Apache Kafka with big data tools like Hadoop, Apache Spark, Apache Storm, and Elasticsearch. Style and approachEasy-to-follow, step-by-step recipes explaining from start to finish how to accomplish real-world tasks.

By introducing in-memory persistent storage, Apache Spark eliminates the need to store intermediate data in filesystems, thereby increasing processing speed by up to 100 times. This book will focus on how to analyze large and complex sets of data. Starting with installing and configuring Apache Spark with various cluster managers, you will cover setting up development environments. You will then cover various recipes to perform interactive queries using Spark SQL and real-

time streaming with various sources such as Twitter Stream and Apache Kafka. You will then focus on machine learning, including supervised learning, unsupervised learning, and recommendation engine algorithms. After mastering graph processing using GraphX, you will cover various recipes for cluster optimization and troubleshooting.

Rails Cookbook is packed with the solutions you need to be a proficient developer with Rails, the leading framework for building the new generation of Web 2.0 applications. Recipes range from the basics, like installing Rails and setting up your development environment, to the latest techniques, such as developing RESTful web services. With applications that are code light, feature-full and built to scale quickly, Rails has revolutionized web development. The Rails Cookbook addresses scores of real-world challenges; each one includes a tested solution, plus a discussion of how and why it works, so that you can adapt the techniques to similar situations. Topics include: Modeling data with the ActiveRecord library Setting up views with ActionView and RHTML templates Building your application's logic into ActionController Testing and debugging your Rails application Building responsive web applications using JavaScript and Ajax Ensuring that your application is security and performs well Deploying your application with Mongrel and Apache Using Capistrano to automate deployment Using the many Rails plugins Working with graphics Whether you're new to Rails or an experienced developer, you'll discover ways to test, debug and secure your

applications, incorporate Ajax, use caching to improve performance, and put your application into production. Want to get ahead of the Web 2.0 curve? This valuable cookbook will save you hundreds of hours when developing applications with Rails.

Enterprise developers face several challenges when it comes to building serverless applications, such as integrating applications and building container images from source. With more than 60 practical recipes, this cookbook helps you solve these issues with Knative—the first serverless platform natively designed for Kubernetes. Each recipe contains detailed examples and exercises, along with a discussion of how and why it works. If you have a good understanding of serverless computing and Kubernetes core resources such as deployment, services, routes, and replicas, the recipes in this cookbook show you how to apply Knative in real enterprise application development. Authors Kamesh Sampath and Burr Sutter include chapters on autoscaling, build and eventing, observability, Knative on OpenShift, and more. With this cookbook, you'll learn how to:

- Efficiently build, deploy, and manage modern serverless workloads
- Apply Knative in real enterprise scenarios, including advanced eventing
- Monitor your Knative serverless applications effectively
- Integrate Knative with CI/CD principles, such as using pipelines for faster, more successful production deployments
- Deploy a rich ecosystem of enterprise integration patterns and connectors in Apache Camel K as Kubernetes and Knative components

Integrating data from multiple sources is essential in the

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

Salt has become one of the major players in automation and configuration management solutions. This book starts with the basics of the tool, the procedures to get up and running with Salt and then moves on to configuring very simple but important details to receive optimal performance from the tool. It also walks you through Salt configurations for different infrastructure components

and the details of the Salt modules for each of the components. The book also provides some common problem scenarios and how to troubleshoot them. With detailed configuration, their explanation and command line outputs of the module execution, Salt Cookbook will help you to get up and running with Salt for all your infrastructural needs.

There's plenty of documentation on installing and configuring the Apache web server, but where do you find help for the day-to-day stuff, like adding common modules or fine-tuning your activity logging? That's easy. The new edition of the Apache Cookbook offers you updated solutions to the problems you're likely to encounter with the new versions of Apache. Written by members of the Apache Software Foundation, and thoroughly revised for Apache versions 2.0 and 2.2, recipes in this book range from simple tasks, such installing the server on Red Hat Linux or Windows, to more complex tasks, such as setting up name-based virtual hosts or securing and managing your proxy server. Altogether, you get more than 200 timesaving recipes for solving a crisis or other deadline conundrums, with topics including: Security Aliases, Redirecting, and Rewriting CGI Scripts, the suexec Wrapper, and other dynamic content techniques Error Handling SSL Performance This book tackles everything from beginner problems to those faced by experienced users. For every

problem addressed in the book, you will find a worked-out solution that includes short, focused pieces of code you can use immediately. You also get explanations of how and why the code works, so you can adapt the problem-solving techniques to similar situations. Instead of poking around mailing lists, online documentation, and other sources, rely on the Apache Cookbook for quick solutions when you need them. Then you can spend your time and energy where it matters most.

This collection of tips, tools, and scripts provides clear, concise, hands-on solutions that can be applied to the challenges facing anyone running a network of Linux servers from small networks to large data centers.

This book is part of Packt's Cookbook series; each chapter looks at a different aspect of working with Apache Solr. The recipes deal with common problems of working with Solr by using easy-to-understand, real-life examples. The book is not in any way a complete Apache Solr reference and you should see it as a helping hand when things get rough on your journey with Apache Solr. Developers who are working with Apache Solr and would like to know how to combat common problems will find this book of great use. Knowledge of Apache Lucene would be a bonus but is not required.

If you are a Java developer or a manager who has experience with Apache Maven and want to extend

your knowledge, then this is the ideal book for you. Apache Maven Cookbook is for those who want to learn how Apache Maven can be used for build automation. It is also meant for those familiar with Apache Maven, but want to understand the finer nuances of Maven and solve specific problems. This book is for system engineers and administrators who have a fundamental understanding of information management systems and infrastructure. It helps if you've already played around with Chef; however, this book covers all the important topics you will need to know. If you don't want to dig through a whole book before you can get started, this book is for you, as it features a set of independent recipes you can try out immediately. Provides experienced mobile developers with details about how each Cordova API works and how to use the APIs in their own applications. Original. Over 150 recipes to design and optimize large scale Apache Cassandra deployments. Over insightful 90 recipes to get lightning-fast analytics with Apache Spark About This Book Use Apache Spark for data processing with these hands-on recipes Implement end-to-end, large-scale data analysis better than ever before Work with powerful libraries such as MLlib, SciPy, NumPy, and Pandas to gain insights from your data Who This Book Is For This book is for novice and intermediate level data science professionals and data analysts who want to solve data science problems with a distributed computing framework. Basic experience with data science implementation tasks is

expected. Data science professionals looking to skill up and gain an edge in the field will find this book helpful. What You Will Learn Explore the topics of data mining, text mining, Natural Language Processing, information retrieval, and machine learning. Solve real-world analytical problems with large data sets. Address data science challenges with analytical tools on a distributed system like Spark (apt for iterative algorithms), which offers in-memory processing and more flexibility for data analysis at scale. Get hands-on experience with algorithms like Classification, regression, and recommendation on real datasets using Spark MLLib package. Learn about numerical and scientific computing using NumPy and SciPy on Spark. Use Predictive Model Markup Language (PMML) in Spark for statistical data mining models. In Detail Spark has emerged as the most promising big data analytics engine for data science professionals. The true power and value of Apache Spark lies in its ability to execute data science tasks with speed and accuracy. Spark's selling point is that it combines ETL, batch analytics, real-time stream analysis, machine learning, graph processing, and visualizations. It lets you tackle the complexities that come with raw unstructured data sets with ease. This guide will get you comfortable and confident performing data science tasks with Spark. You will learn about implementations including distributed deep learning, numerical computing, and scalable machine learning. You will be shown effective solutions to problematic concepts in data science using Spark's data science libraries such as MLLib, Pandas, NumPy, SciPy, and more. These simple and efficient recipes will show you how to implement algorithms and optimize your work. Style and approach This book contains a comprehensive range of recipes designed to help you learn the fundamentals and tackle the difficulties of data science. This book outlines practical steps to produce powerful insights into Big Data

through a recipe-based approach.

This book is for anyone who builds and administers servers, especially in a web operations context. It requires some experience of Linux systems administration, including familiarity with the command line, file system, and text editing. No programming experience is required.

This is a collection of problems, solutions, and practical examples for webmasters, web administrators, programmers, and anyone who works with Apache.

Authoritative Answers to All Your Apache Questions--Now Updated to Cover Apache 2.0 Linux Apache Web Server Administration is the most complete, most advanced guide to the Apache Web server you'll find anywhere. Written by a leading Apache expert--and now updated to cover Apache 2.0--this book teaches you, step-by-step, all the standard and advanced techniques you need to know to administer Apache on a Linux box. Hundreds of clear, consistent examples illustrate these techniques in detail--so you stay on track and accomplish all your goals. Coverage includes: * Compiling Apache from source code * Creating and hosting virtual web sites * Using Server-Side Includes to create Web pages with dynamic content * Using Apache directives to configure your site * Extending Apache using add-on modules * Using the Common Gateway Interface for web programming * Enhancing the performance of CGI programs with FastCGI and mod_perl * Installing Apache support for PHP * Extending Apache to run Java servlets or Java Server Pages * Attaching Apache to a database server * Using URL rewriting for increased request-handling flexibility * Implementing user authentication * Adding Secure Sockets Layer for enhanced system security * Customizing Apache's log formats

The Craig Hunt Linux Library The Craig Hunt Linux Library provides in-depth, advanced coverage of the key topics for Linux administrators. Topics include Samba,

System Administration, DNS Server Administration, Network Servers, Security, and Sendmail. Each book in the series is either written by or meticulously reviewed by Craig Hunt to ensure the highest quality and most complete coverage for networking professionals working specifically in Linux environments.

Apache Cookbook"O'Reilly Media, Inc."

Simplify real-time data processing by leveraging the power of Apache Kafka 1.0 Key Features Use Kafka 1.0 features such as Confluent platforms and Kafka streams to build efficient streaming data applications to handle and process your data Integrate Kafka with other Big Data tools such as Apache Hadoop, Apache Spark, and more Hands-on recipes to help you design, operate, maintain, and secure your Apache Kafka cluster with ease Book Description Apache Kafka provides a unified, high-throughput, low-latency platform to handle real-time data feeds. This book will show you how to use Kafka efficiently, and contains practical solutions to the common problems that developers and administrators usually face while working with it. This practical guide contains easy-to-follow recipes to help you set up, configure, and use Apache Kafka in the best possible manner. You will use Apache Kafka Consumers and Producers to build effective real-time streaming applications. The book covers the recently released Kafka version 1.0, the Confluent Platform and Kafka Streams. The programming aspect covered in the book will teach you how to perform important tasks such as message validation, enrichment and composition. Recipes focusing on optimizing the performance of your Kafka cluster, and integrate Kafka with a variety of third-party tools such as Apache Hadoop, Apache Spark, and Elasticsearch will help ease your day to day collaboration with Kafka greatly. Finally, we cover tasks related to monitoring and securing your Apache Kafka cluster using tools such as Ganglia and

Graphite. If you're looking to become the go-to person in your organization when it comes to working with Apache Kafka, this book is the only resource you need to have. What you will learn -Install and configure Apache Kafka 1.0 to get optimal performance -Create and configure Kafka Producers and Consumers -Operate your Kafka clusters efficiently by implementing the mirroring technique -Work with the new Confluent platform and Kafka streams, and achieve high availability with Kafka -Monitor Kafka using tools such as Graphite and Ganglia -Integrate Kafka with third-party tools such as Elasticsearch, Logstash, Apache Hadoop, Apache Spark, and more Who this book is for This book is for developers and Kafka administrators who are looking for quick, practical solutions to problems encountered while operating, managing or monitoring Apache Kafka. If you are a developer, some knowledge of Scala or Java will help, while for administrators, some working knowledge of Kafka will be useful.

A solution-based guide to put your deep learning models into production with the power of Apache Spark Key Features Discover practical recipes for distributed deep learning with Apache Spark Learn to use libraries such as Keras and TensorFlow Solve problems in order to train your deep learning models on Apache Spark Book Description With deep learning gaining rapid mainstream adoption in modern-day industries, organizations are looking for ways to unite popular big data tools with highly efficient deep learning libraries. As a result, this will help deep learning models train with higher efficiency and speed. With the help of the Apache Spark Deep Learning Cookbook, you'll work through specific recipes to generate outcomes for deep learning algorithms, without getting bogged down in theory. From setting up Apache Spark for deep learning to implementing types of neural net, this book tackles both common and not so

common problems to perform deep learning on a distributed environment. In addition to this, you'll get access to deep learning code within Spark that can be reused to answer similar problems or tweaked to answer slightly different problems. You will also learn how to stream and cluster your data with Spark. Once you have got to grips with the basics, you'll explore how to implement and deploy deep learning models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in Spark, using popular libraries such as TensorFlow and Keras. By the end of the book, you'll have the expertise to train and deploy efficient deep learning models on Apache Spark. What you will learn

- Set up a fully functional Spark environment
- Understand practical machine learning and deep learning concepts
- Apply built-in machine learning libraries within Spark
- Explore libraries that are compatible with TensorFlow and Keras
- Explore NLP models such as Word2vec and TF-IDF on Spark
- Organize dataframes for deep learning evaluation
- Apply testing and training modeling to ensure accuracy
- Access readily available code that may be reusable

Who this book is for If you're looking for a practical and highly useful resource for implementing efficiently distributed deep learning models with Apache Spark, then the Apache Spark Deep Learning Cookbook is for you. Knowledge of the core machine learning concepts and a basic understanding of the Apache Spark framework is required to get the best out of this book. Additionally, some programming knowledge in Python is a plus.

Easy, hands-on recipes to help you understand Hive and its integration with frameworks that are used widely in today's big data world About This Book Grasp a complete reference of different Hive topics. Get to know the latest recipes in development in Hive including CRUD operations Understand Hive internals and integration of Hive with different

frameworks used in today's world. Who This Book Is For The book is intended for those who want to start in Hive or who have basic understanding of Hive framework. Prior knowledge of basic SQL command is also required What You Will Learn Learn different features and offering on the latest Hive Understand the working and structure of the Hive internals Get an insight on the latest development in Hive framework Grasp the concepts of Hive Data Model Master the key concepts like Partition, Buckets and Statistics Know how to integrate Hive with other frameworks such as Spark, Accumulo, etc In Detail Hive was developed by Facebook and later open sourced in Apache community. Hive provides SQL like interface to run queries on Big Data frameworks. Hive provides SQL like syntax also called as HiveQL that includes all SQL capabilities like analytical functions which are the need of the hour in today's Big Data world. This book provides you easy installation steps with different types of metastores supported by Hive. This book has simple and easy to learn recipes for configuring Hive clients and services. You would also learn different Hive optimizations including Partitions and Bucketing. The book also covers the source code explanation of latest Hive version. Hive Query Language is being used by other frameworks including spark. Towards the end you will cover integration of Hive with these frameworks. Style and approach Starting with the basics and covering the core concepts with the practical usage, this book is a complete guide to learn and explore Hive offerings. Describes the history of the Web server platform and covers downloading and compiling, configuring and running the program on UNIX, writing specialized modules, and establishing security routines. Provides information and examples on using CSS to format Web pages, covering such topics as Web typography, links, navigation, page layouts, and Web site design.

Read Online Apache Cookbook

There's plenty of documentation on installing and configuring the Apache web server, but where do you find help for the day-to-day stuff, like adding common modules or fine-tuning your activity logging? That's easy. The new edition of the Apache Cookbook offers you updated solutions to the problems you're likely to encounter with the new versions of Apache.

Apache is far and away the most widely used web server platform in the world. Both free and rock-solid, it runs more than half of the world's web sites, ranging from huge e-commerce operations to corporate intranets and smaller hobby sites, and it continues to maintain its popularity, drawing new users all the time. If you work with Apache on a regular basis, you have plenty of documentation on installing and configuring your server, but where do you go for help with the day-to-day stuff, like adding common modules or fine-tuning your activity logging? The Apache Cookbook is a collection of problems, solutions, and practical examples for webmasters, web administrators, programmers, and everyone else who works with Apache. For every problem addressed in the book, there's a worked-out solution or "recipe"--short, focused pieces of code that you can use immediately. But this book offers more than cut-and-paste code. You also get explanations of how and why the code works, so you can adapt the problem-solving techniques to similar situations. The recipes in the Apache Cookbook range from simple tasks, such as installing the server on Red Hat Linux or Windows, to more complex tasks, such as setting up name-based virtual hosts or securing and managing your proxy server. The two hundred plus recipes in the book cover additional topics such as: Security Aliases, Redirecting, and Rewriting CGI Scripts, the suexec Wrapper, and other dynamic content techniques Error Handling SSL Performance The impressive collection of useful code in this book is a

Read Online Apache Cookbook

guaranteed timesaver for all Apache users, from novices to advanced practitioners. Instead of poking around mailing lists, online documentation, and other sources, you can rely on the Apache Cookbook for quick solutions to common problems, and then you can spend your time and energy where it matters most.

Combine the power of Apache Spark and Python to build effective big data applications

Key Features

- Perform effective data processing, machine learning, and analytics using PySpark
- Overcome challenges in developing and deploying Spark solutions using Python
- Explore recipes for efficiently combining Python and Apache Spark to process data

Book Description

Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. The PySpark Cookbook presents effective and time-saving recipes for leveraging the power of Python and putting it to use in the Spark ecosystem. You'll start by learning the Apache Spark architecture and how to set up a Python environment for Spark. You'll then get familiar with the modules available in PySpark and start using them effortlessly. In addition to this, you'll discover how to abstract data with RDDs and DataFrames, and understand the streaming capabilities of PySpark. You'll then move on to using ML and MLlib in order to solve any problems related to the machine learning capabilities of PySpark and use GraphFrames to solve graph-processing problems. Finally, you will explore how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will be able to use the Python API for Apache Spark to solve any problems associated with building data-intensive applications. What you will learn

- Configure a local instance of PySpark in a virtual environment
- Install and configure Jupyter in local and multi-node environments
- Create DataFrames from JSON and a dictionary using pyspark.sql
- Explore

regression and clustering models available in the ML module
Use DataFrames to transform data used for modeling
Connect to PubNub and perform aggregations on streams
Who this book is for The PySpark Cookbook is for you if you are a Python developer looking for hands-on recipes for using the Apache Spark 2.x ecosystem in the best possible way. A thorough understanding of Python (and some familiarity with Spark) will help you get the best out of the book.

Apache web server has good documentation for installing and configuring, but where do you go for help with the day-to-day stuff, like adding common modules or fine-tuning your activity logging? The third edition of this book provides more than 200 updated recipes for the latest versions of Apache, with practical examples for web administrators, programmers, and anyone else who works with the web server. For every problem addressed in the book, you'll find short, focused pieces of code that you can use immediately. You also get explanations of how and why the code works, so you can adapt the problem-solving techniques to similar situations. Recipes range from simple tasks, such as installation, to complex tasks, such as setting up name-based virtual hosts. This cookbook has solutions for beginners and experienced users alike.

Over 70 recipes to help you use Apache Spark as your single big data computing platform and master its libraries About This Book This book contains recipes on how to use Apache Spark as a unified compute engine Cover how to connect various source systems to Apache Spark Covers various parts of machine learning including supervised/unsupervised learning & recommendation engines Who This Book Is For This book is for data engineers, data scientists, and those who want to implement Spark for real-time data processing. Anyone who is using Spark (or is planning to) will benefit from this book. The book assumes you have a basic knowledge of

Scala as a programming language. What You Will Learn
Install and configure Apache Spark with various cluster managers & on AWS Set up a development environment for Apache Spark including Databricks Cloud notebook Find out how to operate on data in Spark with schemas Get to grips with real-time streaming analytics using Spark Streaming & Structured Streaming Master supervised learning and unsupervised learning using MLlib Build a recommendation engine using MLlib Graph processing using GraphX and GraphFrames libraries Develop a set of common applications or project types, and solutions that solve complex big data problems In Detail While Apache Spark 1.x gained a lot of traction and adoption in the early years, Spark 2.x delivers notable improvements in the areas of API, schema awareness, Performance, Structured Streaming, and simplifying building blocks to build better, faster, smarter, and more accessible big data applications. This book uncovers all these features in the form of structured recipes to analyze and mature large and complex sets of data. Starting with installing and configuring Apache Spark with various cluster managers, you will learn to set up development environments. Further on, you will be introduced to working with RDDs, DataFrames and Datasets to operate on schema aware data, and real-time streaming with various sources such as Twitter Stream and Apache Kafka. You will also work through recipes on machine learning, including supervised learning, unsupervised learning & recommendation engines in Spark. Last but not least, the final few chapters delve deeper into the concepts of graph processing using GraphX, securing your implementations, cluster optimization, and troubleshooting. Style and approach This book is packed with intuitive recipes supported with line-by-line explanations to help you understand Spark 2.x's real-time processing capabilities and deploy scalable big data solutions. This is a

valuable resource for data scientists and those working on large-scale data projects.

Over 60 simple but incredibly effective recipes for taking control of OFBiz.

Over 90 practical, actionable recipes to automate, test, and manage your infrastructure quickly and effectively About This Book Bring down your delivery timeline from days to hours by treating your server configurations and VMs as code, just like you would with software code. Take your existing knowledge and skill set with your existing tools (Puppet, Chef, or Docker) to the next level and solve IT infrastructure challenges. Use practical recipes to use code to provision and deploy servers and applications and have greater control of your infrastructure.

Who This Book Is For This book is for DevOps engineers and developers working in cross-functional teams or operations and would now switch to IAC to manage complex infrastructures. What You Will Learn Provision local and remote development environments with Vagrant Automate production infrastructures with Terraform, Ansible and Cloud-init on AWS, OpenStack, Google Cloud, Digital Ocean, and more Manage and test automated systems using Chef and Puppet Build, ship, and debug optimized Docker containers Explore the best practices to automate and test everything from cloud infrastructures to operating system configuration In Detail Infrastructure as Code (IAC) is a key aspect of the DevOps movement, and this book will show you how to transform the way you work with your infrastructure—by treating it as software. This book is dedicated to helping you discover the essentials of infrastructure automation and its related practices; the over 90 organized practical solutions will demonstrate how to work with some of the very best tools and cloud solutions. You will learn how to deploy repeatable infrastructures and services on AWS, OpenStack, Google Cloud, and Digital Ocean. You

will see both Ansible and Terraform in action, manipulate the best bits from cloud-init to easily bootstrap instances, and simulate consistent environments locally or remotely using Vagrant. You will discover how to automate and test a range of system tasks using Chef or Puppet. You will also build, test, and debug various Docker containers having developers' interests in mind. This book will help you to use the right tools, techniques, and approaches to deliver working solutions for today's modern infrastructure challenges. Style and approach This is a recipe-based book that allows you to venture into some of the most cutting-edge practices and techniques about IAC and solve immediate problems when trying to implement them.

Over 100 practical recipes to make Apache Solr faster, more reliable and return better results.

This book is intended for developers who have some familiarity with Apache Karaf and who want a quick reference for practical, proven tips on how to perform common tasks such as configuring Pax modules deployed in Apache Karaf, Extending HttpService with Apache Karaf. You should have working knowledge of Apache karaf, as the book provides a deeper understanding of the capabilities of Apache Karaf. Java is one of the most widely accepted languages because of its design and programming features. This book offers a range of software development examples in simple and straightforward Java 11 code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently.

Optimized for Kubernetes, Quarkus is designed to help you create Java applications that are cloud first, container native, and serverless capable. With this cookbook, authors Alex Soto Bueno and Jason Porter from Red Hat provide detailed solutions for installing, interacting with, and using Quarkus in the development and production of microservices. The

recipes in this book show midlevel to senior developers familiar with Java enterprise application development how to get started with Quarkus quickly. You'll become familiar with how Quarkus works within the wider Java ecosystem and discover ways to adapt this framework to your particular needs. You'll learn how to: Shorten the development cycle by enabling live reloading in dev mode Connect to and communicate with Kafka Develop with the reactive programming model Easily add fault tolerance to your services Build your application as a Kubernetes-ready container Ease development with OpenAPI and test a native Quarkus application

This book is written in a Cookbook style with short recipes showing developers how to effectively implement EIP without breaking everything in the process. It is concise and to the point, and it helps developers get their data flowing between different components without the need to read through page upon page of theory, while also enabling the reader to learn how to create exciting new projects. Camel Enterprise Integration Cookbook is intended for developers who have some familiarity with Apache Camel and who want a quick lookup reference to practical, proven tips on how to perform common tasks. Every recipe also includes a summary and reference pointers for more details that make it easy for you to get a deeper understanding of the Apache Camel capabilities that you will use day to day.

This book is written in a Cookbook style showing you how to set up and expand your Puppet infrastructure. It progresses through detailed information on the language and features, external tools, reporting, monitoring, and troubleshooting, and concludes with many specific recipes for managing popular applications. The book assumes that the reader already has a working Puppet installation and perhaps has written some basic manifests or adapted some published modules. It also

Read Online Apache Cookbook

requires some experience of Linux systems administration, including familiarity with the command line, file system, and text editing. No programming experience is required.

[Copyright: caec4b56e45e73a7b3d2e04058d9238f](#)