

## Apache Kafka 1.0 Cookbook: Over 100 Practical Recipes On Using Distributed Enterprise Messaging To Handle Real Time Data

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 use this 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.

Working with unbounded and fast-moving data streams has historically been difficult. But with Kafka Streams and ksqldb, building stream processing applications is easy and fun. This practical guide shows data engineers how to use these tools to build highly scalable stream processing applications for moving, enriching, and transforming large amounts of data in real time. Mitch Seymour, data services engineer at Mailchimp, explains important stream processing concepts against a backdrop of several interesting business problems. You'll learn the strengths of both Kafka Streams and ksqldb to help you choose the best tool for each unique stream processing project. Non-Java developers will find the ksqldb path to be an especially gentle introduction to stream processing. Learn the basics of Kafka and the pub/sub communication pattern Build stateless and stateful stream processing applications using Kafka Streams and ksqldb Perform advanced stateful operations, including windowed joins and aggregations Understand how stateful processing works under the hood Learn about ksqldb's data integration features, powered by Kafka Connect Work with different types of collections in Kafka and perform push and pull queries Deploy your Kafka Streams and ksqldb applications to production

Apache Kafka 1.0 Cookbook

Oklahoman S.M Barret wrote down and edited Apache Chief Geronimo's story of his life.

Over 100 practical recipes on using distributed enterprise messaging to handle real-time data

Grade: Beyond the Basics

Data Pipelines with Apache Airflow

### Microservices Deployment Cookbook

#### Apache Spark 2.x Machine Learning Cookbook

Learn how to use the Akka framework to build effective applications in Scala About This Book Covers a discussion on Lagom—the newest launched Akka framework that is built to create complex microservices easily The recipe approach of the book allows the reader to know important and independent concepts of Scala and Akka in a seamless manner Provides a comprehensive understanding of the Akka actor model and implementing it to create reactive web applications Who This Book Is For If you are a Scala developer who wants to build scalable and concurrent applications, then this book is for you. Basic knowledge of Akka will help you take advantage of this book. What You Will Learn Control an actor using the ControlAware mailbox Test a fault-tolerant application using the Akka test kit Create a parallel application using futures and agents Package and deploy Akka application inside Docker Deploy remote actors programmatically on different nodes Integrate Streams with Akka actors Install Lagom and create a Lagom project In Detail Akka is an open source toolkit that simplifies the construction of distributed and concurrent applications on the JVM. This book will teach you how to develop reactive applications in Scala using the Akka framework. This book will show you how to build concurrent, scalable, and reactive applications in Akka. You will see how to create high performance applications, extend applications, build microservices with Lagom, and more. We will explore Akka's actor model and show you how to incorporate concurrency into your applications. The book puts a special emphasis on performance improvement and how to make an application available for users. We also make a special mention of message routing and construction. By the end of this book, you will be able to create a high-performing Scala application using the Akka framework. Style and approach This highly practical recipe-based approach will allow you to build scalable, robust, and reactive applications using the Akka framework.

If you're familiar with Gradle's basics elements—possibly through the author's previous O'Reilly book, Building and Testing with Gradle—this more advanced guide provides the recipes, techniques, and syntax to help you master this build automation tool. With clear, concise explanations and lots of ready-to-use code examples, you'll explore four discrete areas of Gradle functionality: file operations, custom Gradle plugins, build lifecycle hooks, and dependency management. Learn how to use Gradle's rich set of APIs and Groovy-based Domain Specific Language to customize build software that actually conforms to your product. By using the techniques in this book, you'll be able to write domain-specific builds that support every other line of code your team creates. Examine Gradle's file API, including copy tasks, pattern matching, content filtering, and the FileCollection interface Understand the process for building and packaging a custom Gradle plug-in Manage build complexity with hook methods and Gradle's rule feature Learn how Gradle handles dependency management natively and through customization Explore Gradle's core plug-ins as well as key examples from the Gradle community

Easy-to-use development frameworks 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 they work. 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 Stutter 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

Build efficient data flow and machine learning programs with this flexible, multi-functional open-source cluster-computing framework Key Features Master the art of real-time big data processing and machine learning Explore a wide range of use-cases to analyze large data Discover ways to optimize your work by using many features of Spark 2.x and Scala Book Description Apache Spark is an in-memory, cluster-based data processing system that provides a wide range of functionalities such as big data processing, analytics, machine learning, and more. With this Learning Path, you can take your knowledge of Apache Spark to the next level by learning how to expand Spark's functionality and building your own data flow and machine learning programs on this platform. You will work with the different modules in Apache Spark, such as interactive querying with Spark SQL, using DataFrames and datasets, implementing streaming analytics with Spark Streaming, and applying machine learning and deep learning techniques on Spark using MLlib and various external tools. By the end of this elaborately designed Learning Path, you will have all the knowledge you need to master Apache Spark, and build your own big data processing and analytics pipeline quickly and without any hassle. This Learning Path includes content from the following Packt products: Mastering Apache Spark 2.x by Romeo Kienzler Scala and Spark for Big Data Analytics by Md. Rezaul Karim, Sriidhar Alla Apache Spark 2.x Machine Learning Cookbook by Siamak Amirghodbi, Meenakshi Rajendran, Broderick Hall, Shuen MeiCookbook What you will learn Get to grips with all the features of Apache Spark 2.x Perform highly optimized real-time big data processing Use ML and DL techniques with Spark MLlib and third-party tools Analyze structured and unstructured data using SparkSQL and GraphX Understand tuning, debugging, and monitoring of big data applications Build scalable and fault-tolerant streaming applications Develop scalable recommendation engines Who this book is for If you are an intermediate-level Spark developer

Looking to master the advanced capabilities and use-cases of Apache Spark 2.x, this Learning Path is ideal for you. Big data professionals who want to learn how to integrate and use the features of Apache Spark and build a strong big data pipeline will also find this Learning Path useful. To grasp the concepts explained in this Learning Path, you must know the fundamentals of Apache Spark and Scala.

Distributed Process Coordination

Over 70 recipes for building AI solutions for smart homes, industrial IoT, and smart cities

Spark Cookbook

Artificial Intelligence for IoT Cookbook

Practical Data Integration for the Web

Customizing Next-Generation Builds

Over 80 recipes to simplify machine learning model implementations with SparkAbout 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 dataWho This Book Is ForThis 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 Sparkin DetailMachine 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 when developing machine learning models 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. 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 3Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystemSharpen your Hadoop skills with real-world case studies and codeBook 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 learnGain an in-depth understanding of distributed computing using Hadoop 3Develop enterprise-grade applications using Apache Spark, Flink, and moreBuild scalable and high-performance Hadoop data pipelines with security, monitoring, and data governanceExplore batch data processing patterns and how to model data in HadoopMaster best practices for enterprises using, or planning to use, Hadoop 3 as a data platformUnderstand security aspects of Hadoop, including authorization and authenticationWho 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.

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 PyTorch. By the end of this 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.

Bridge the gap between basic understanding of Go and use of its advanced features About This Book Discover a number of recipes and approaches to develop modern back-end applications Put to use the best practices to combine the recipes for sophisticated parallel tools This book is based on Go 1.8, which is the latest version Who This Book Is For This book is for web developers, programmers, and enterprise developers. Basic knowledge of the Go language is assumed. Experience with back-end application development is not necessary, but may help understand the motivation behind some of the recipes. What You Will Learn Test your application using advanced testing methodologies Develop an awareness of application structures, interface design, and tooling Create strategies for third-party packages, dependencies, and vendoring Get to know tricks on treating data such as collections Handle errors and cleanly pass them along to calling functions Wrap dependencies in interfaces for ease of portability and testing Explore reactive programming design patterns in Go In Detail Go (a.k.a. Golang) is a statically-typed programming language first developed at Google. It is derived from C with additional features such as garbage collection, type safety, dynamic-typing capabilities, additional built-in types, and a large standard library. This book takes off where basic tutorials on the language leave off. You can immediately put into practice some of the more advanced concepts and libraries offered by the language while avoiding some of the common mistakes for new Go developers. The book covers basic type and error handling. It explores applications that interact with users, such as websites, command-line tools, or via the file system. It demonstrates how to handle advanced topics such as parallelism, distributed systems, and performance tuning. Lastly, it finishes with reactive and serverless programming in Go. Style and approach This guide is a handy reference for developers to quickly look up Go development patterns. It is a companion to other resources and a reference that will be useful long after reading it through the first time. Each recipe includes working, simple, and tested code that can be used as a reference or foundation for your own applications.

A Problem-Solution Approach with PySpark2

Trino: The Definitive Guide

Elasticsearch 8.x Cookbook

Azure Databricks Cookbook

Go Cookbook

Big data processing at scale to unlock unique business insights

"An Airflow bible. Useful for all kinds of users, from novice to expert." - Rambabu Posa, Sai Aashika Consultancy Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. A successful pipeline moves data efficiently, minimizing pauses and blockages between tasks, keeping every process along the way operational. Apache Airflow provides a single customizable environment for building and managing data pipelines, eliminating the need for a hodgepodge collection of tools, snowflake code, and homegrown processes. Using real-world scenarios and examples, Data Pipelines with Apache Airflow teaches you how to simplify and automate data pipelines, reduce operational overhead, and smoothly integrate all the technologies in your stack. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Data pipelines manage the flow of data from initial collection through consolidation, cleaning, analysis, visualization, and more. Apache Airflow provides a single platform you can use to design, implement, monitor, and maintain your pipelines. Its easy-to-use UI, plug-and-play options, and flexible Python scripting make Airflow perfect for any data management task. About the book Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. You'll explore the most common usage patterns, including aggregating multiple data sources, connecting to and from data lakes, and cloud deployment. Part reference and part tutorial, this practical guide covers every aspect of the directed acyclic graphs (DAGs) that power Airflow, and how to customize them for your pipeline's needs. What's inside Build, test, and deploy Airflow pipelines as DAGs Automate monitoring and alerting on datasets using best-of-breed tools Set up Airflow in production environments About the reader For DevOps, data engineers, machine learning engineers, and sysadmins with intermediate Python skills. About the author Bas Harenstak and Julian de Ruiter are data engineers with extensive experience using Airflow to develop pipelines for major companies. Bas is also an Airflow committer. Table of Contents PART 1 - GETTING STARTED 1 Meet Apache Airflow 2 Anatomy of an Airflow DAG 3 Scheduling in Airflow 4 Templating tasks using the Airflow context 5 Defining dependencies between tasks PART 2 - BEYOND THE BASICS 6 Triggering workflows 7 Communicating with external systems 8 Building custom components 9 Testing 10 Running tasks in containers PART 3 - AIRFLOW IN PRACTICE 11 Best practices 12 Operating Airflow in production 13 Securing

Airflow 14 Project: Finding the fastest way to get around NYC PART 4 - IN THE CLOUDS 15 Airflow in the clouds 16 Airflow on AWS 17 Airflow on Azure 18 Airflow in GCP JSON is becoming the backbone for meaningful data interchange over the internet. This format is now supported by an entire ecosystem of standards, tools, and technologies for building truly elegant, useful, and efficient applications. With this hands-on guide, author and architect Tom Marrs shows you how to build enterprise-class applications and services by leveraging JSON tooling and message/document design. JSON at Work provides application architects and developers with guidelines, best practices, and use cases, along with lots of real-world examples and code samples. You'll start with a comprehensive JSON overview, explore the JSON ecosystem, and then dive into JSON's use in the enterprise. Get acquainted with JSON basics and learn how to model JSON data learn how to use JSON with Node.js, Ruby on Rails, and Java Structure JSON documents with JSON Schema to design and test APIs and learn the context of JSON documents with JSON Convert JSON documents to other data formats with JSON Transform tools Compare JSON-based hypertext formats, including HAL and JSON-LD Leverage MongoDB to store and access JSON documents Use Apache Kafka to exchange JSON-based messages between services

Need to move a relational database application to Hadoop? This comprehensive guide introduces you to Apache Hive, Hadoop's data warehouse infrastructure. You'll quickly learn how to use Hive's SQL dialect—HiveQL—to summarize, query, and analyze large datasets stored in Hadoop's distributed filesystem. The example-driven guide shows you how to set up and configure Hive in your environment, provides a detailed overview of Hadoop and MapReduce, and demonstrates how Hive works within the Hadoop ecosystem. You'll also find real-world case studies that describe how companies have used Hive to solve unique problems involving petabytes of data. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes Customize data formats and storage options, from files to external databases Load and extract data from tables—and use queries, grouping, filtering, joining, and other conventional query methods Gain best practices for creating user defined functions (UDFs) Learn Hive patterns you should use and anti-patterns you should avoid Integrate Hive with other data processing programs Use storage handlers for NoSQL databases and other datastores Learn the pros and cons of running Hive on Amazon's Elastic MapReduce

Master over 60 recipes to help you deliver complete, scalable, microservice-based solutions and see the improved business results immediately About This Book Adopt microservices-based architecture and deploy it at scale Build your complete microservice architecture using different recipes for different solutions Identify specific tools for specific scenarios and deliver immediate business results, correlate use cases, and adopt them in your team and organization Who This Book Is For This book is for developers, ops, and DevOps professionals who would like to put microservices to work and improve products, services, and operations. Those looking to build and deploy microservices will find this book useful, as well as managers and people at CXO level looking to adopt microservices in their organization. Prior knowledge of Java is expected. No prior knowledge of microservices is assumed. What You Will Learn Building and testing Spring Boot, Wildfly Swarm, Dropwizard, and SparkJava Containerize your microservices using Docker Deploy microservices using Mesos/Marathon and Kubernetes Implement service discovery and load balancing using Zookeeper, Consul, and Nginx Monitor microservices using Graphite and Grafana Write stream programs with Kafka Streams and Spark Aggregate and manage logs using Kafka Get introduced to DC/OS, Docker Swarm, and YARN In Detail This book will help you understand, deploy, and manage microservices at scale. It is driven by a sample application, helping you gradually build a complete microservice-based ecosystem. Rather than just focusing on writing a microservice, this book addresses various other microservice-related solutions: deployments, clustering, load balancing, logging, streaming, and monitoring. The initial chapters offer insights into how web and enterprise apps can be migrated to scalable microservices. Moving on, you'll see how to Dockerize your application so that it is ready to be shipped and deployed. We will look at how to deploy microservices on Mesos and Marathon and will also deploy microservices on Kubernetes. Next, you will implement service discovery and load balancing for your microservices. We'll also show you how to build asynchronous streaming systems using Kafka Streams and Apache Spark. Finally, we wind up by aggregating your logs in Kafka, creating your own metrics, and monitoring the metrics for the microservice. Style and approach This book follows a recipe-driven approach and shows you how to plug and play with all the various pieces, putting them together to build a complete scalable microservice ecosystem. You do not need to study the chapters in order, as you can directly refer to the content you need for your situation.

Delphi Cookbook

Learning Spark

Apache Spark for Data Science Cookbook

Parqus Cookbook

Apache Kafka 1.0 Cookbook

Apache Spark 2: Data Processing and Real-Time Analytics

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 systems. After mastering graph processing with Spark, you will cover various recipes for cluster optimization and troubleshooting. Data is bigger, arrives faster, and 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

If you are already using Neo4j in your application and want to learn more about our data analysis or database graphs, this is the book for you. This book also caters for your needs if you are looking to migrate your existing application to Neo4j in the future. We assume that you are already familiar with any general purpose programming language and have some familiarity with Neo4j.

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.

Accelerate and scale real-time analytics solutions using the Apache Spark-based analytics service

Clojure Programming Cookbook

Real-Time Data and Stream Processing at Scale

Go Web Development Cookbook

Applying Math with Python

Over 180 recipes to perform fast, scalable, and reliable searches for your enterprise

Summary Kafka Streams in Action teaches you everything you need to know to implement stream processing on data flowing into your Kafka platform, allowing you to focus on getting more from your data without sacrificing time or effort. Foreword by Neha Narkhede, Cocreator of Apache Kafka Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Not all stream-based applications require a dedicated processing cluster. The lightweight Kafka Streams library provides exactly the power and simplicity you need for message handling in microservices and real-time event processing. With the Kafka Streams API, you can filter and transform data streams with just Kafka and your application. About the Book Kafka Streams in Action teaches you to implement stream processing within the Kafka platform. In this easy-to-follow book, you'll explore real-world examples to collect, transform, and aggregate data, work with multiple processors, and handle real-time events. You'll even dive into streaming SQL with KSQL! Practical to the very end, it finishes with testing and operational aspects, such as monitoring and debugging. What's inside Using the KStreams API Filtering, transforming, and splitting data Working with the Processor API Integrating with external systems About the Reader Assumes some experience with distributed systems. No knowledge of Kafka or streaming applications required. About the Author Bill Bejeck is a Kafka Streams contributor and Confluent engineer with over 15 years of software development experience. Table of Contents PART 1 - GETTING STARTED WITH KAFKA STREAMS Welcome to Kafka Streams Kafka quickPART 2 - KAFKA STREAMS DEVELOPMENT Developing Kafka Streams Streams and state the KTable API The Processor APIPART 3 - ADMINISTERING KAFKA STREAMS Monitoring and performance Testing a Kafka Streams applicationPART 4 - ADVANCED CONCEPTS WITH KAFKA STREAMS APPENDICES Appendix A - Additional configuration information Appendix B - Exactly once semantics

Easy-to-use development frameworks 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 they work. 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 Stutter 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

Build efficient data flow and machine learning programs with this flexible, multi-functional open-source cluster-computing framework Key Features Master the art of real-time big data processing and machine learning Explore a wide range of use-cases to analyze large data Discover ways to optimize your work by using many features of Spark 2.x and Scala Book Description Apache Spark is an in-memory, cluster-based data processing system that provides a wide range of functionalities such as big data processing, analytics, machine learning, and more. With this Learning Path, you can take your knowledge of Apache Spark to the next level by learning how to expand Spark's functionality and building your own data flow and machine learning programs on this platform. You will work with the different modules in Apache Spark, such as interactive querying with Spark SQL, using DataFrames and datasets, implementing streaming analytics with Spark Streaming, and applying machine learning and deep learning techniques on Spark using MLlib and various external tools. By the end of this elaborately designed Learning Path, you will have all the knowledge you need to master Apache Spark, and build your own big data processing and analytics pipeline quickly and without any hassle. This Learning Path includes content from the following Packt products: Mastering Apache Spark 2.x by Romeo Kienzler Scala and Spark for Big Data Analytics by Md. Rezaul Karim, Sriidhar Alla Apache Spark 2.x Machine Learning Cookbook by Siamak Amirghodbi, Meenakshi Rajendran, Broderick Hall, Shuen MeiCookbook What you will learn Get to grips with all the features of Apache Spark 2.x Perform highly optimized real-time big data processing Use ML and DL techniques with Spark MLlib and third-party tools Analyze structured and unstructured data using SparkSQL and GraphX Understand tuning, debugging, and monitoring of big data applications Build scalable and fault-tolerant streaming applications Develop scalable recommendation engines Who this book is for If you are an intermediate-level Spark developer

Looking to master the advanced capabilities and use-cases of Apache Spark 2.x, this Learning Path is ideal for you. Big data professionals who want to learn how to integrate and use the features of Apache Spark and build a strong big data pipeline will also find this Learning Path useful. To grasp the concepts explained in this Learning Path, you must know the fundamentals of Apache Spark and Scala.

Distributed Process Coordination

Over 70 recipes for building AI solutions for smart homes, industrial IoT, and smart cities

Spark Cookbook

Artificial Intelligence for IoT Cookbook

Practical Data Integration for the Web

Customizing Next-Generation Builds

Over 80 recipes to simplify machine learning model implementations with SparkAbout 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 dataWho This Book Is ForThis 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 Sparkin DetailMachine 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 when developing machine learning models 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. 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 3Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystemSharpen your Hadoop skills with real-world case studies and codeBook 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 learnGain an in-depth understanding of distributed computing using Hadoop 3Develop enterprise-grade applications using Apache Spark, Flink, and moreBuild scalable and high-performance Hadoop data pipelines with security, monitoring, and data governanceExplore batch data processing patterns and how to model data in HadoopMaster best practices for enterprises using, or planning to use, Hadoop 3 as a data platformUnderstand security aspects of Hadoop, including authorization and authenticationWho 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.

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 PyTorch. By the end of this 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.

Bridge the gap between basic understanding of Go and use of its advanced features About This Book Discover a number of recipes and approaches to develop modern back-end applications Put to use the best practices to combine the recipes for sophisticated parallel tools This book is based on Go 1.8, which is the latest version Who This Book Is For This book is for web developers, programmers, and enterprise developers. Basic knowledge of the Go language is assumed. Experience with back-end application development is not necessary, but may help understand the motivation behind some of the recipes. What You Will Learn Test your application using advanced testing methodologies Develop an awareness of application structures, interface design, and tooling Create strategies for third-party packages, dependencies, and vendoring Get to know tricks on treating data such as collections Handle errors and cleanly pass them along to calling functions Wrap dependencies in interfaces for ease of portability and testing Explore reactive programming design patterns in Go In Detail Go (a.k.a. Golang) is a statically-typed programming language first developed at Google. It is derived from C with additional features such as garbage collection, type safety, dynamic-typing capabilities, additional built-in types, and a large standard library. This book takes off where basic tutorials on the language leave off. You can immediately put into practice some of the more advanced concepts and libraries offered by the language while avoiding some of the common mistakes for new Go developers. The book covers basic type and error handling. It explores applications that interact with users, such as websites, command-line tools, or via the file system. It demonstrates how to handle advanced topics such as parallelism, distributed systems, and performance tuning. Lastly, it finishes with reactive and serverless programming in Go. Style and approach This guide is a handy reference for developers to quickly look up Go development patterns. It is a companion to other resources and a reference that will be useful long after reading it through the first time. Each recipe includes working, simple, and tested code that can be used as a reference or foundation for your own applications.

