Learning Apache Kafka Second Edition Garg Nishant

Start from scratch and learn how to administer Apache Kafka effectively for messaging In Detail Kafka is one of those systems that is very simple to describe at a high level but has an incredible depth of technical detail when you dig deeper. Learning Apache Kafka Second Edition provides you with step-by-step, practical examples that help you take advantage of the real power of Kafka and handle hundreds of megabytes of messages per second from multiple clients. This book teaches you everything you need to know, right from setting up Kafka clusters to understanding basic blocks like producer, broker, and consumer blocks. Once you are all set up, you will then explore additional settings and configuration changes to achieve ever more complex goals. You will also learn how Kafka is designed internally and what configurations make it more effective. Finally, you will learn how Kafka works with other tools such as Hadoop, Storm, and so on. What You Will Learn Set up both single- and multi-node Kafka clusters and start sending messages Understand the internals of Kafka's design and learn about message compression and replication in Kafka Explore additional settings and configuration changes to achieve ever more complex goals Write Kafka message producers and custom consumers using the Kafka API Integrate Kafka with Apache Hadoop and Storm Integrate Kafka with other tools for logging, packaging, and so on Administer Kafka effectively and consistently with cluster management tools Downloading the example code for this book. You can download the example code files for all Packt books you have purchased from your account at http://www.PacktPub.com. If you purchased this book elsewhere, you can visit http://www.PacktPub.com/support and register to have the files e-mailed directly to you.
Ready to build cloud native applications? Get a hands-on introduction to daily life as a developer crafting code on OpenShift, the open source container application platform from Red Hat. Creating and packaging your apps for deployment on modern distributed systems can be daunting. Too often, adding infrastructure value can complicate development. With this practical guide, you'll learn how to build, deploy, and manage a multitiered application on OpenShift. Authors Joshua Wood and Brian Tannous, principal developer advocates at Red Hat, demonstrate how OpenShift speeds application development. With the Kubernetes container orchestrator at its core, OpenShift simplifies and automates the way you build, ship, and run code. You'll learn how to use OpenShift and the Quarkus Java framework to develop and deploy apps using proven enterprise technologies and practices that you can apply to code in any language. Learn the development cycles for building and deploying on OpenShift, and the tools that drive them Use OpenShift to build, deploy, and manage the ongoing lifecycle of an n-tier application Create a continuous integration and deployment pipeline to build and deploy application source code on OpenShift Automate scaling decisions with metrics and trigger lifecycle events with webhooks
Design and administer fast, reliable enterprise messaging systems with Apache Kafka About This Book Build efficient real-time streaming applications in Apache Kafka to process data streams of data Master the core Kafka APIs to set up Apache Kafka clusters and start writing message producers and consumers A comprehensive guide to help you get a solid grasp of the Apache Kafka concepts in Apache Kafka with pracitcalpractical examples Who This Book Is For If you want to learn how to use Apache Kafka and the different tools in the Kafka ecosystem in the easiest possible manner, this book is for you. Some programming experience with Java is required to get the most out of this book What You Will Learn Learn the basics of Apache Kafka from scratch Use the basic building blocks of a streaming application Design effective streaming applications with Kafka using Spark, Storm &, and Heron Understand the importance of a low -latency , high- throughput, and fault-tolerant messaging system Make effective capacity planning while deploying your Kafka Application Understand and implement the best security practices In Detail Apache Kafka is a popular distributed streaming platform

that acts as a messaging queue or an enterprise messaging system. It lets you publish and subscribe to a stream of records, and process them in a fault-tolerant way as they occur. This book is a comprehensive guide to designing and architecting enterprise-grade streaming applications using Apache Kafka and other big data tools. It includes best practices for building such applications, and tackles some common challenges such as how to use Kafka efficiently and handle high data volumes with ease. This book first takes you through understanding the type messaging system and then provides a thorough introduction to Apache Kafka and its internal details. The second part of the book takes you through designing streaming application using various frameworks and tools such as Apache Spark, Apache Storm, and more. Once you grasp the basics, we will take you through more advanced concepts in Apache Kafka such as capacity planning and security. By the end of this book, you will have all the information you need to be comfortable with using Apache Kafka, and to design efficient streaming data applications with it. Style and approach A step-by –step, comprehensive guide filled with practical and real- world examples

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets—Spark's core APIs—through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

Advanced analytics on your Big Data with latest Apache Spark 2.x About This Book An advanced guide with a combination of instructions and practical examples to extend the most up-to date Spark functionalities. Extend your data processing capabilities to process huge chunk of data in minimum time using advanced concepts in Spark. Master the art of real-time processing with the help of Apache Spark 2.x Who This Book Is For If you are a developer with some experience with Spark and want to strengthen your knowledge of how to get around in the world of Spark, then this book is ideal for you. Basic knowledge of Linux, Hadoop and Spark is assumed. Reasonable knowledge of Scala is expected. What You Will Learn Examine Advanced Machine Learning and DeepLearning with MLlib, SparkML, SystemML, H2O and DeepLearning4J Study highly optimised unified batch and real-time data processing using SparkSQL and Structured Streaming Evaluate large-scale Graph Processing and Analysis using GraphX and GraphFrames Apply Apache Spark in Elastic deployments using Jupyter and Zeppelin Notebooks, Docker, Kubernetes and the IBM Cloud Understand internal details of cost based optimizers used in Catalyst, SystemML and GraphFrames Learn how specific parameter settings affect overall performance of an Apache Spark cluster Leverage Scala, R and python for your data science projects In Detail Apache Spark is an in-memory cluster-based parallel processing system that provides a wide range of functionalities such as graph processing, machine learning, stream processing, and SQL. This book aims to take your knowledge of Spark to the next level by teaching you how to expand Spark's functionality and implement your data flows and machine/deep learning programs on top of the platform. The book commences with an overview of the Spark ecosystem. It will introduce you to Project Tungsten and Catalyst, two of the major advancements of Apache Spark 2.x. You will

understand how memory management and binary processing, cache-aware computation, and code generation are used to speed things up dramatically. The book extends to show how to incorporate H20, SystemML, and Deeplearning4j for machine learning, and Jupyter Notebooks and Kubernetes/Docker for cloud-based Spark. During the course of the book, you will learn about the latest enhancements to Apache Spark 2.x, such as interactive querying of live data and unifying DataFrames and Datasets. You will also learn about the updates on the APIs and how DataFrames and Datasets affect SQL, machine learning, graph processing, and streaming. You will learn to use Spark as a big data operating system, understand how to implement advanced analytics on the new APIs, and explore how easy it is to use Spark in day-to-day tasks. Style and approach This book is an extensive guide to Apache Spark modules and tools and shows how Spark's functionality can be extended for real-time processing and storage with worked examples.

If you are a Hadoop programmer who wants to learn about Flume to be able to move datasets into Hadoop in a timely and replicable manner, then this book is ideal for you. No prior knowledge about Apache Flume is necessary, but a basic knowledge of Hadoop and the Hadoop File System (HDFS) is assumed.

Build a scalable, fault-tolerant and highly available data layer for your applications using Apache Cassandra About This Book Install Cassandra and set up multi-node clusters Design rich schemas that capture the relationships between different data types Master the advanced features available in Cassandra 3.x through a step-by-step tutorial and build a scalable, high performance database layer Who This Book Is For If you are a NoSQL developer and new to Apache Cassandra who wants to learn its common as well as not-so-common features, this book is for you. Alternatively, a developer wanting to enter the world of NoSQL will find this book useful. It does not assume any prior experience in coding or any framework. What You Will Learn Install Cassandra Create keyspaces and tables with multiple clustering columns to organize related data Use secondary indexes and materialized views to avoid denormalization of data Effortlessly handle concurrent updates with collection columns Ensure data integrity with lightweight transactions and logged batches Understand eventual consistency and use the right consistency level for your situation Understand data distribution with Cassandra Develop simple application using Java driver and implement application-level optimizations In Detail Cassandra is a distributed database that stands out thanks to its robust feature set and intuitive interface, while providing high availability and scalability of a distributed data store. This book will introduce you to the rich feature set offered by Cassandra, and empower you to create and manage a highly scalable, performant and fault-tolerant database layer. The book starts by explaining the new features implemented in Cassandra 3.x and get you set up with Cassandra. Then you'll walk through data modeling in Cassandra and the rich feature set available to design a flexible schema. Next you'll learn to create tables with composite partition keys, collections and user-defined types and get to know different methods to avoid denormalization of data. You will then proceed to create user-defined functions and aggregates in Cassandra. Then, you will set up a multi node cluster and see how the dynamics of Cassandra change with it. Finally, you will implement some application-level optimizations using a Java client. By the end of this book, you'll be fully equipped to build powerful, scalable Cassandra database layers for your applications. Style and approach This book takes a step-by- step approach to give you basic to intermediate knowledge of Apache Cassandra. Every concept is explained in depth, and is supplemented with practical examples when required.

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 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 quicklyPART 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 Advanced applications with Kafka StreamsAPPENDIXES Appendix A - Additional configuration information Appendix B - Exactly once semantics

Are you looking to build resilient big data services and applications without compromising on the reliability, stability and the performance of your high-performance, low latency system and have heard that Apache Kafka might be your best bet but have no idea how to use it?And are you looking for a comprehensive guide that will show you everything you need to know about Apache Kafka so you can understand just how it is designed for real-time, high speed data processing as well as how to put it to use?If you've answered YES, Let This Book Introduce You To The World Of Using Apache Kafka To Build World-Class, Low Latency, High Throughput Systems That Have The Ability To Handle High-Volume Real Time Data Feeds Just Like Some Of The World's Biggest Tech Systems Like Twitter, Uber, Netflix, LinkedIn And More!Every successful business nowadays revolves around big data and that's why there is quite a number of platforms, technologies and frameworks that have cropped up to support this over the years.One such solution which is proving to be effective and the best is Apache Kafka, an open source software platform specifically designed for high-speed, real-time data processing, as seen in its ability to support driver and passenger matching on Uber for example and its ability to support many real time services on LinkedIn.The fact that you are reading this means you've probably grown curious about Apache Kafka having heard a lot about it and you are wondering what kinds of systems it can be implemented in and how to implement it.Perhaps you are wondering...What exactly does Apache Kafka do that makes so exceptional that major applications like Cisco, Walmart, JPMC, Bank of America, Uber and LinkedIn would use it?Who are the closest competitors and what makes Apache Kafka different?How does Apache Kafka work?How can you use Apache Kafka in building resilient, durable and stable applications and services for your business?If you have these and other related questions about Apache Kafka, this corporate IT training courseware is for you so keep reading, as it will teach you everything you need to understand the inner workings of Apache Kafka like the back of our hand.More precisely, you will learn: -The basics of big data, including the place of such concepts like Spark, Zookeeper, the Kafka framework and how they all relate-An insider look into Kafka framework and Kafka use cases to help you understand real world applications inside out-The inner workings of Apache Kafka, including Zookeeper watches, Zookeeper's role in cluster membership, the responsibilities and election of the controller broker, replication, partition and the bootstrap server-How to code producer configurations and consumer groups-The ins and outs of Kafka data delivery, including delivery semantics, and

service goals-How to master Kafka administrative functions, including dynamic configurations, handling partitions, consumer group tools and more-And so much moreEven if this is your first encounter with Apache Kafka as a business, this corporate IT training courseware will leave you feeling confident about your ability to start using it to develop and administer fast and reliable IT systems!What's more - you can download all supporting files from Ernesto.Net along with a docker container that has already been staged to help you complete the activities in the book!Scroll up and click Buy Now With 1-Click or Buy Now to get started!

More and more data-driven companies are looking to adopt stream processing and streaming analytics. With this concise ebook, you'll learn best practices for designing a reliable architecture that supports this emerging big-data paradigm. Authors Ted Dunning and Ellen Friedman (Real World Hadoop) help you explore some of the best technologies to handle stream processing and analytics, with a focus on the upstream queuing or message-passing layer. To illustrate the effectiveness of these technologies, this book also includes specific use cases. Ideal for developers and non-technical people alike, this book describes: Key elements in good design for streaming analytics, focusing on the essential characteristics of the messaging layerNew messaging technologies, including Apache Kafka and MapR Streams, with links to sample codeTechnology choices for streaming analytics: Apache Spark Streaming, Apache Flink, Apache Storm, and Apache ApexHow stream-based architectures are helpful to support microservicesSpecific use cases such as fraud detection and geo-distributed data streams Ted Dunning is Chief Applications Architect at MapR Technologies, and active in the open source community. He currently serves as VP for Incubator at the Apache Foundation, as a champion and mentor for a large number of projects, and as committer and PMC member of the Apache ZooKeeper and Drill projects. Ted is on Twitter as @ted_dunning. Ellen Friedman, a committer for the Apache Drill and Apache Mahout projects, is a solutions consultant and well-known speaker and author, currently writing mainly about big data topics. With a PhD in Biochemistry, she has years of experience as a research scientist and has written about a variety of technical topics. Ellen is on Twitter as @Ellen_Friedman.

Summary The Spark distributed data processing platform provides an easy-to-implement tool for ingesting, streaming, and processing data from any source. In Spark in Action, Second Edition, you'll learn to take advantage of Spark's core features and incredible processing speed, with applications including real-time computation, delayed evaluation, and machine learning. Spark skills are a hot commodity in enterprises worldwide, and with Spark's powerful and flexible Java APIs, you can reap all the benefits without first learning Scala or Hadoop. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Analyzing enterprise data starts by reading, filtering, and merging files and streams from many sources. The Spark data processing engine handles this varied volume like a champ, delivering speeds 100 times faster than Hadoop systems. Thanks to SQL support, an intuitive interface, and a straightforward multilanguage API, you can use Spark without learning a complex new ecosystem. About the book Spark in Action, Second Edition, teaches you to create end-to-end analytics applications. In this entirely new book, you'll learn from interesting Java-based examples, including a complete data pipeline for processing NASA satellite data. And you'll discover Java, Python, and Scala code samples hosted on GitHub that you can explore and adapt, plus appendixes that give you a cheat sheet for installing tools and understanding Spark-specific terms. What's inside Writing Spark applications in Java Spark application architecture Ingestion through files, databases, streaming, and Elasticsearch Querying distributed datasets with Spark SQL About the reader This book does not assume previous experience with Spark, Scala, or Hadoop. About the author Jean-Georges Perrin is an experienced data and software architect. He is France's first IBM Champion and has been honored for 12 consecutive years. Table of Contents PART 1 - THE THEORY CRIPPLED BY AWESOME EXAMPLES 1 So, what is Spark, anyway? 2

Architecture and flow 3 The majestic role of the dataframe 4 Fundamentally lazy 5 Building a simple app for deployment 6 Deploying your simple app PART 2 - INGESTION 7 Ingestion from files 8 Ingestion from databases 9 Advanced ingestion: finding data sources and building your own 10 Ingestion through structured streaming PART 3 - TRANSFORMING YOUR DATA 11 Working with SQL 12 Transforming your data 13 Transforming entire documents 14 Extending transformations with user-defined functions 15 Aggregating your data PART 4 - GOING FURTHER 16 Cache and checkpoint: Enhancing Spark's performances 17 Exporting data and building full data pipelines 18 Exploring deployment

This book explains the key feature to develop a complex and stable network that helps to gather the data to optimize the asset performance and maximize the production in the Industries leveraging on the cloud infrastructure and services. By the end, you can design the Industrial IoT network and the architecture for processing its data in the cloud.

Summary Event Streams in Action is a foundational book introducing the ULP paradigm and presenting techniques to use it effectively in data-rich environments. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many high-profile applications, like LinkedIn and Netflix, deliver nimble, responsive performance by reacting to user and system events as they occur. In large-scale systems, this requires efficiently monitoring, managing, and reacting to multiple event streams. Tools like Kafka, along with innovative patterns like unified log processing, help create a coherent data processing architecture for event-based applications. About the Book Event Streams in Action teaches you techniques for aggregating, storing, and processing event streams using the unified log processing pattern. In this hands-on guide, you'll discover important application designs like the lambda architecture, stream aggregation, and event reprocessing. You'll also explore scaling, resiliency, advanced stream patterns, and much more! By the time you're finished, you'll be designing large-scale data-driven applications that are easier to build, deploy, and maintain. What's inside Validating and monitoring event streams Event analytics Methods for event modeling Examples using Apache Kafka and Amazon Kinesis About the Reader For readers with experience coding in Java, Scala, or Python. About the Author Alexander Dean developed Snowplow, an open source event processing and analytics platform. Valentin Crettaz is an independent IT consultant with 25 years of experience. Table of Contents PART 1 - EVENT STREAMS AND UNIFIED LOGS Introducing event streams The unified log 24 Event stream processing with Apache Kafka Event stream processing with Amazon Kinesis Stateful stream processing PART 2- DATA ENGINEERING WITH STREAMS Schemas Archiving events Railway-oriented processing Commands PART 3 - EVENT ANALYTICS Analytics-on-read Analytics-on-write

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

Dig deep into the data with a hands-on guide to machine learning with updated examples and more! Machine Learning: Hands-On for Developers and Technical Professionals provides hands-on instruction and fully-coded working examples for the most common machine learning techniques used by developers and technical professionals. The book contains a breakdown of each ML variant, explaining how it works and how it is used within certain industries, allowing readers to incorporate the presented techniques into their own work as they follow along. A core tenant of machine learning is a strong focus on data preparation, and a full exploration of the various types of learning algorithms illustrates how the proper tools can help any developer extract information and insights from existing data. The book includes a full complement of Instructor's Materials to facilitate use in the classroom, making this resource useful for students and as a professional reference. At its core, machine learning is a mathematical, algorithm-based technology that forms the basis of historical data mining and modern big data science. Scientific analysis of big data requires a working knowledge of machine learning, which forms predictions based on known properties learned from training data. Machine Learning is an accessible, comprehensive guide for the non-mathematician, providing clear guidance that allows readers to: Learn the languages of machine learning including Hadoop, Mahout, and Weka Understand decision trees, Bayesian networks, and artificial neural networks Implement Association Rule, Real Time, and Batch learning Develop a strategic plan for safe, effective, and efficient machine learning By learning to construct a system that can learn from data, readers can increase their utility across industries. Machine learning sits at the core of deep dive data analysis and visualization, which is increasingly in demand as companies discover the goldmine hiding in their existing data. For the tech professional involved in data science, Machine Learning: Hands-On for Developers and Technical Professionals provides the skills and techniques required to dig deeper.

Learn how to integrate full-stack open source big data architecture and to choose the correct technology—Scala/Spark, Mesos, Akka, Cassandra, and Kafka—in every layer. Big data architecture is becoming a requirement for many different enterprises. So far, however, the focus has largely been on collecting, aggregating, and crunching large data sets in a timely manner. In many cases now, organizations need more than one paradigm to perform efficient analyses. Big Data SMACK explains each of the full-stack technologies and, more importantly, how to best integrate them. It provides detailed coverage of the practical benefits of these technologies and incorporates real-world examples in every situation. This book focuses on the problems and scenarios solved by the architecture, as well as the solutions provided by every technology. It covers the six main concepts of big data architecture and how integrate, replace,

and reinforce every layer: The language: Scala The engine: Spark (SQL, MLib, Streaming, GraphX) The container: Mesos, Docker The view: Akka The storage: Cassandra The message broker: Kafka What You Will Learn: Make big data architecture without using complex Greek letter architectures Build a cheap but effective cluster infrastructure Make queries, reports, and graphs that business demands Manage and exploit unstructured and No-SQL data sources Use tools to monitor the performance of your architecture Integrate all technologies and decide which ones replace and which ones reinforce Who This Book Is For: Developers, data architects, and data scientists looking to integrate the most successful big data open stack architecture and to choose the correct technology in every layer

Process large volumes of data in real-time while building high performance and robust data stream processing pipeline using the latest Apache Kafka 2.0 Key Features Solve practical large data and processing challenges with Kafka Tackle data processing challenges like late events, windowing, and watermarking Understand real-time streaming applications processing using Schema registry, Kafka connect, Kafka streams, and KSQL Book Description Apache Kafka is a great open source platform for handling your real-time data pipeline to ensure high-speed filtering and pattern matching on the fly. In this book, you will learn how to use Apache Kafka for efficient processing of distributed applications and will get familiar with solving everyday problems in fast data and processing pipelines. This book focuses on programming rather than the configuration management of Kafka clusters or DevOps. It starts off with the installation and setting up the development environment, before quickly moving on to performing fundamental messaging operations such as validation and enrichment. Here you will learn about message composition with pure Kafka API and Kafka Streams. You will look into the transformation of messages in different formats, such asext, binary, XML, JSON, and AVRO. Next, you will learn how to expose the schemas contained in Kafka with the Schema Registry. You will then learn how to work with all relevant connectors with Kafka Connect. While working with Kafka Streams, you will perform various interesting operations on streams, such as windowing, joins, and aggregations. Finally, through KSQL, you will learn how to retrieve, insert, modify, and delete data streams, and how to manipulate watermarks and windows. What you will learn How to validate data with Kafka Add information to existing data flows Generate new information through message composition Perform data validation and versioning with the Schema Registry How to perform message Serialization and Deserialization How to perform message Serialization and Deserialization Process data streams with Kafka Streams Understand the duality between tables and streams with KSQL Who this book is for This book is for developers who want to quickly master the practical concepts behind Apache Kafka. The audience need not have come across Apache Kafka previously; however, a familiarity of Java or any JVM language will be helpful in understanding the code in this book.

Designing and writing a real-time streaming publication with Apache Apex About This Book Get a clear, practical approach to real-time data processing Program Apache Apex streaming applications This book shows you Apex integration with the open source Big Data ecosystem Who This Book Is For This book assumes knowledge of application development with Java and familiarity with distributed systems. Familiarity with other real-time streaming frameworks is not required, but some practical experience with other big data processing utilities might be helpful. What You Will Learn Put together a functioning Apex application from scratch Scale an Apex application and configure it for optimal performance Understand how to deal with failures via the fault tolerance features of the platform Use Apex via other frameworks such as Beam Understand the DevOps implications of deploying Apex In Detail Apache Apex is a next-generation stream processing framework designed to operate on data at large scale, with minimum latency, maximum reliability, and strict correctness guarantees. Half of the book consists of Apex applications, showing you key aspects of data processing pipelines such as

connectors for sources and sinks, and common data transformations. The other half of the book is evenly split into explaining the Apex framework, and tuning, testing, and scaling Apex applications. Much of our economic world depends on growing streams of data, such as social media feeds, financial records, data from mobile devices, sensors and machines (the Internet of Things - IoT). The projects in the book show how to process such streams to gain valuable, timely, and actionable insights. Traditional use cases, such as ETL, that currently consume a significant chunk of data engineering resources are also covered. The final chapter shows you future possibilities emerging in the streaming space, and how Apache Apex can contribute to it. Style and approach This book is divided into two major parts: first it explains what Apex is, what its relevant parts are, and how to write well-built Apex applications. The second part is entirely application-driven, walking you through Apex applications of increasing complexity.

Understand how to apply auto machine learning to data streams and create transactional machine learning (TML) solutions that are frictionless (require minimal to no human intervention) and elastic (machine learning solutions that can scale up or down by controlling the number of data streams, algorithms, and users of the insights). This book will strengthen your knowledge of the inner workings of TML solutions using data streams with auto machine learning integrated with Apache Kafka. Transactional Machine Learning with Data Streams and AutoML introduces the industry challenges with applying machine learning to data streams. You will learn the framework that will help you in choosing business problems that are best suited for TML. You will also see how to measure the business value of TML solutions. You will then learn the technical components of TML solutions, including the reference and technical architecture of a TML solution. This book also presents a TML solution template that will make it easy for you to quickly start building your own TML solutions. Specifically, you are given access to a TML Python library and integration technologies for download. You will also learn how TML will evolve in the future, and the growing need by organizations for deeper insights from data streams. By the end of the book, you will have a solid understanding of TML. You will know how to build TML solutions with all the necessary details, and all the resources at your fingertips. What You Will Learn Discover transactional machine learning Measure the business value of TML Choose TML use cases Design technical architecture of TML solutions with Apache Kafka Work with the technologies used to build TML solutions Build transactional machine learning solutions with hands-on code together with Apache Kafka in the cloud Who This Book Is For Data scientists, machine learning engineers and architects, and AI and machine learning business leaders.

Learning Apache Kafka - Second Edition

What is this book about? PHP, Apache, and MySQL are the three key open source technologies that form the basis for most active Web servers. This book takes you step-by-step through understanding each — using it and combining it with the other two on both Linux and Windows servers. This book guides you through creating your own sites using the open source AMP model. You discover how to install PHP, Apache, and MySQL. Then you create PHP Web pages,

including database management and security. Finally, you discover how to integrate your work with e-commerce and other technologies. By building different types of Web sites, you progress from setting up simple database tables to tapping the full potential of PHP, Apache, and MySQL. When you're finished, you will be able to create well-designed, dynamic Web sites using open source tools. What does this book cover? Here's what you will learn from this book: How PHP server-side scripting language works for connecting HTML-based Web pages to a backend database Syntax, functions, and commands for PHP, Apache, and MySQL Methods and techniques for building user-friendly forms How to easily store, update, and access information using MySQL Ways to allow the user to edit a database E-commerce applications using these three technologies How to set up user logins, profiles, and personalizations Proper protocols for error handling Who is this book for? This book is for beginners who are new to PHP and who need to learn quickly how to create Web sites using open source tools. Some basic HTML knowledge is helpful but not essential. Serverless computing greatly simplifies software development. Your team can focus solely on your application while the cloud provider manages the servers you need. This practical guide shows you step-by-step how to build and deploy complex applications in a flexible multicloud, multilanguage environment using Apache OpenWhisk. You'll learn how this platform enables you to pursue a vendor-independent approach using preconfigured containers, microservices, and Kubernetes as your cloud operating system. Michele Sciabarrà demonstrates how to build a serverless application using classical design patterns and the programming language or languages that best fit your task. You'll start by building a simple serverless application hands-on before diving into the more complex aspects of the OpenWhisk platform. Examine how OpenWhisk's serverless architecture works, including the use of packages, actions, sequences, triggers, rules, and feeds Learn how OpenWhisk compares to existing architectures, such as Java Enterprise Edition Manipulate OpenWhisk features using the command-line interface or a JavaScript API Design applications using common Gang of Four design patterns Use architectural design patterns such as model-view-controller to combine several OpenWhisk actions Learn how to test and debug your code in a serverless environment Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second

Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a programming language like Java and have basic familiarity with Hadoop. What's Inside Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents PART 1 BACKGROUND AND FUNDAMENTALS Hadoop in a heartbeat Introduction to YARN PART 2 DATA LOGISTICS Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop PART 3 BIG DATA PATTERNS Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning, debugging, and testing PART 4 BEYOND MAPREDUCE SQL on Hadoop Writing a YARN application

Build a scalable, fault-tolerant and highly available data layer for your applications using Apache CassandraAbout This Book* Install Cassandra and use it to set up multi-node clusters* Design rich schemas that capture the relationships between different data types* Master the advanced features available in Cassandra 3.x through a step-by-step tutorial and build a scalable, high performance database layerWho This Book Is ForIf you are a first-time user of Apache Cassandra who wants to learn the basic of it, as well as some not-so-basic features, this book is for you. It does not assume any prior experience in coding or any framework.What you will learn* Install Cassandra and create your first keyspace* Create tables with multiple clustering columns to organize related data* Use secondary indexes and materialized views to avoid denormalization of data* Effortlessly handle concurrent updates with collection columns* Ensure data integrity with lightweight transactions and logged batches* Understand eventual consistency and use the right consistency level for your situation* Understand data distribution with Cassandra and get to know ways to implement application-level optimizationsIn DetailCassandra is a distributed database that stands out thanks to its robust feature set and intuitive interface, while still providing the high availability and scalability of a distributed store. This book will introduce you to the rich features offered by Cassandra, and empower you to create and manage a highly performant, fault-tolerant database layer.The book starts by explaining the new features implemented in Cassandra 3.x, you'll see

how to install Cassandra, and you'll understand Lightweight Transactions. Next you'll learn to create tables with composite partition keys, and get to know different methods to avoid denormalization of data. You will then proceed to create user-defined functions and data distribution in Cassandra. Finally, you will set up a multi node cluster and implement application-level optimization using a Java client.By the end of this book, you'll be fully equipped to build powerful, scalable Cassandra database layers for your applications.

Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

This book is for readers who want to know more about Apache Kafka at a hands-on level; the key audience is those with software development experience but no prior exposure to Apache Kafka or similar technologies. It is also useful for enterprise application developers and big data enthusiasts who have worked with other publisher-subscriber-based systems and want to explore Apache Kafka as a futuristic solution.

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

Get up to speed on Scala, the JVM language that offers all the benefits of a modern

object model, functional programming, and an advanced type system. Packed with code examples, this comprehensive book shows you how to be productive with the language and ecosystem right away, and explains why Scala is ideal for today's highly scalable, data-centric applications that support concurrency and distribution. This second edition covers recent language features, with new chapters on pattern matching, comprehensions, and advanced functional programming. You'll also learn about Scala's command-line tools, third-party tools, libraries, and language-aware plugins for editors and IDEs. This book is ideal for beginning and advanced Scala developers alike. Program faster with Scala's succinct and flexible syntax Dive into basic and advanced functional programming (FP) techniques Build killer big-data apps, using Scala's functional combinators Use traits for mixin composition and pattern matching for data extraction Learn the sophisticated type system that combines FP and object-oriented programming concepts Explore Scala-specific concurrency tools, including Akka Understand how to develop rich domain-specific languages Learn good design techniques for building scalable and robust Scala applications

The software architecture landscape has evolved dramatically over the past decade. Microservices have displaced monoliths. Data and applications are increasingly becoming distributed and decentralised. But composing disparate systems is a hard problem. More recently, software practitioners have been rapidly converging on event-driven architecture as a sustainable way of dealing with complexity - integrating systems without increasing their coupling.In Effective Kafka, Emil Koutanov explores the fundamentals of Event-Driven Architecture - using Apache Kafka - the world's most popular and supported open-source event streaming platform.You'll learn: - The fundamentals of event-driven architecture and event streaming platforms- The background and rationale behind Apache Kafka, its numerous potential uses and applications- The architecture and core concepts - the underlying software components, partitioning and parallelism, load-balancing, record ordering and consistency modes- Installation of Kafka and related tooling - using standalone deployments, clusters, and containerised deployments with Docker- Using CLI tools to interact with and administer Kafka classes, as well as publishing data and browsing topics- Using third-party web-based tools for monitoring a cluster and gaining insights into the event streams- Building stream processing applications in Java 11 using off-the-shelf client libraries- Patterns and best-practice for organising the application architecture, with emphasis on maintainability and testability of the resulting code- The numerous gotchas that lurk in Kafka's client and broker configuration, and how to counter them- Theoretical background on distributed and concurrent computing, exploring factors affecting their liveness and safety- Best-practices for running multi-tenanted clusters across diverse engineering teams, how teams collaborate to build complex systems at scale and equitably share the cluster with the aid of quotas- Operational aspects of running Kafka clusters at scale, performance tuning and methods for optimising network and storage utilisation- All aspects of Kafka security -including network segregation, encryption, certificates, authentication and authorization.The coverage is progressively delivered and carefully aimed at giving you a journey-like experience into becoming proficient with Apache Kafka and Event-Driven Architecture. The goal is to get you designing and building applications. And by the conclusion of this book, you will be a confident practitioner and a Kafka evangelist within your organisation - wielding the knowledge

necessary to teach others.
Kafka in Action is a practical, hands-on guide to building Kafka-based data pipelines. Filled with real-world use cases and scenarios, this book probes Kafka's most common use cases, ranging from simple logging through managing streaming data systems for message routing, analytics, and more. In systems that handle big data, streaming data, or fast data, it's important to get your data pipelines right. Apache Kafka is a wicked-fast distributed streaming platform that operates as more than just a persistent log or a flexible message queue. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.
Every enterprise application creates data, whether it’s log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you’re an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you’ll learn Kafka’s design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka’s operational measurements Explore how Kafka’s stream delivery capabilities make it a perfect source for stream processing systems
The book will follow a step-by-step tutorial approach which will show the readers how to use Apache Kafka for messaging from scratch.Apache Kafka is for readers with software development experience, but no prior exposure to Apache Kafka or similar technologies is assumed. This book is also for enterprise application developers and big data enthusiasts who have worked with other publisher-subscriber based systems and now want to explore Apache Kafka as a futuristic scalable solution.
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.

Learn how to build scalable cloud native applications with the new-generation Ballerina language using expert tips and best practices Key Features Work with code samples based on the Ballerina Swan Lake Beta1 version Explore the in-built networking protocol support in Ballerina to develop secure distributed apps Build a Ballerina app with an automated CI/CD pipeline with observability to simplify maintenance and deployment Book Description The Ballerina programming language was created by WSO2 for the modern needs of developers where cloud native development techniques have become ubiquitous. Ballerina simplifies how programmers develop and deploy cloud native distributed apps and microservices. Cloud Native Applications with Ballerina will guide you through Ballerina essentials, including variables, types, functions, flow control, security, and more. You'll explore networking as an in-built feature in Ballerina, which makes it a first-class language for distributed computing. With this app development book, you'll learn about different networking protocols as well as different architectural patterns that you can use to implement services on the cloud. As you advance, you'll explore multiple design patterns used in microservice architecture and use serverless in Amazon Web Services (AWS) and Microsoft Azure platforms. You will also get to grips with Docker, Kubernetes, and serverless platforms to simplify maintenance and the deployment process. Later, you'll focus on the Ballerina testing framework along with deployment tools and monitoring tools to build fully automated observable cloud applications. By the end of this book, you will have learned how to apply the Ballerina language for building scalable, resilient, secured, and easy-to-maintain cloud native Ballerina projects and applications. What you will learn Understand the concepts and models in cloud native architecture Get to grips with the high-level concepts of building applications with the Ballerina language Use cloud native architectural design patterns to develop cloud native Ballerina applications Discover how to automate, maintain, and observe cloud native Ballerina applications Use a container to deploy and maintain a Ballerina application with Docker and Kubernetes Explore serverless architecture and use Microsoft Azure and the AWS platform to build serverless applications Who this book is for This Ballerina Swan Lake book is for cloud developers, integration developers, and microservices developers who are facing challenges with legacy tooling and are looking for the latest tools and technologies to solve them. Beginner-level programming knowledge is required before getting started with this Ballerina book.

The book Kafka Streams - Real-time Stream Processing helps you understand the stream processing in general and apply that skill to Kafka streams programming. This book is focusing mainly on the new generation of the Kafka Streams library available in

the Apache Kafka 2.x. The primary focus of this book is on Kafka Streams. However, the book also touches on the other Apache Kafka capabilities and concepts that are necessary to grasp the Kafka Streams programming. Who should read this book? Kafka Streams: Real-time Stream Processing is written for software engineers willing to develop a stream processing application using Kafka Streams library. I am also writing this book for data architects and data engineers who are responsible for designing and building the organization's data-centric infrastructure. Another group of people is the managers and architects who do not directly work with Kafka implementation, but they work with the people who implement Kafka Streams at the ground level. What should you already know? This book assumes that the reader is familiar with the basics of Java programming language. The source code and examples in this book are using Java 8, and I will be using Java 8 lambda syntax, so experience with lambda will be helpful. Kafka Streams is a library that runs on Kafka. Having a good fundamental knowledge of Kafka is essential to get the most out of Kafka Streams. I will touch base on the mandatory Kafka concepts for those who are new to Kafka. The book also assumes that you have some familiarity and experience in running and working on the Linux operating system.

This book takes you on a fantastic journey to discover the attributes of big data using Apache Hive. Key Features Grasp the skills needed to write efficient Hive queries to analyze the Big Data Discover how Hive can coexist and work with other tools within the Hadoop ecosystem Uses practical, example-oriented scenarios to cover all the newly released features of Apache Hive 2.3.3 Book Description In this book, we prepare you for your journey into big data by frstly introducing you to backgrounds in the big data domain, alongwith the process of setting up and getting familiar with your Hive working environment. Next, the book guides you through discovering and transforming the values of big data with the help of examples. It also hones your skills in using the Hive language in an effcient manner. Toward the end, the book focuses on advanced topics, such as performance, security, and extensions in Hive, which will guide you on exciting adventures on this worthwhile big data journey. By the end of the book, you will be familiar with Hive and able to work effeciently to find solutions to big data problems What you will learn Create and set up the Hive environment Discover how to use Hive's definition language to describe data Discover interesting data by joining and filtering datasets in Hive Transform data by using Hive sorting, ordering, and functions Aggregate and sample data in different ways Boost Hive query performance and enhance data security in Hive Customize Hive to your needs by using user-defined functions and integrate it with other tools Who this book is for If you are a data analyst, developer, or simply someone who wants to quickly get started with Hive to explore and analyze Big Data in Hadoop, this is the book for you. Since Hive is an SQL-like language, some previous experience with SQL will be useful to get the most out of this book.

What could you do with data if scalability wasn't a problem? With this hands-on guide, you'll learn how Apache Cassandra handles hundreds of terabytes of data while remaining highly available across multiple data centers -- capabilities that have attracted Facebook, Twitter, and other data-intensive companies. Cassandra: The Definitive Guide provides the technical details and practical examples you need to assess this database management system and put it to work in a production environment. Author Eben Hewitt demonstrates the advantages of Cassandra's nonrelational design, and pays special attention to data modeling. If you're a

developer, DBA, application architect, or manager looking to solve a database scaling issue or future-proof your application, this guide shows you how to harness Cassandra's speed and flexibility. Understand the tenets of Cassandra's column-oriented structure Learn how to write, update, and read Cassandra data Discover how to add or remove nodes from the cluster as your application requires Examine a working application that translates from a relational model to Cassandra's data model Use examples for writing clients in Java, Python, and C# Use the JMX interface to monitor a cluster's usage, memory patterns, and more Tune memory settings, data storage, and caching for better performance

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

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.

Get up to speed with Apache Drill, an extensible distributed SQL query engine that reads massive datasets in many popular file formats such as Parquet, JSON, and CSV. Drill reads data in HDFS or in cloud-native storage such as S3 and works with Hive metastores along with distributed databases such as HBase, MongoDB, and relational databases. Drill works everywhere: on your laptop or in your largest cluster. In this practical book, Drill committers Charles Givre and Paul Rogers show analysts and data scientists how to query and analyze raw data using this powerful tool. Data scientists today spend about 80% of their time just gathering and cleaning data. With this book, you'll learn how Drill helps you analyze data more effectively to drive down time to insight. Use Drill to clean, prepare, and summarize delimited data for further analysis Query file types including logfiles, Parquet, JSON, and other complex formats Query Hadoop, relational databases, MongoDB, and Kafka with standard SQL Connect to Drill programmatically using a variety of languages Use Drill even with challenging

or ambiguous file formats Perform sophisticated analysis by extending Drill’s functionality with user-defined functions Facilitate data analysis for network security, image metadata, and machine learning

Get a practical introduction to React Native, the JavaScript framework for writing and deploying fully featured mobile apps that render natively. The second edition of this hands-on guide shows you how to build applications that target iOS, Android, and other mobile platforms instead of browsers—apps that can access platform features such as the camera, user location, and local storage. Through code examples and step-by-step instructions, web developers and frontend engineers familiar with React will learn how to build and style interfaces, use mobile components, and debug and deploy apps. You’ll learn how to extend React Native using third-party libraries or your own Java and Objective-C libraries. Understand how React Native works under the hood with native UI components Examine how React Native’s mobile-based components compare to basic HTML elements Create and style your own React Native components and applications Take advantage of platform-specific APIs, as well as modules from the framework’s community Incorporate platform-specific components into cross-platform apps Learn common pitfalls of React Native development, and tools for dealing with them Combine a large application’s many screens into a cohesive UX Handle state management in a large app with the Redux library