Learning Apache Kafka Second Edition Garg Nishant

Fundamentals, Implementation, and Operation of Streaming Applications
Over 100 practical recipes on using distributed enterprise messaging to handle real-time data
Learning Apache Kafka Second Edition
Build Frictionless and Elastic Machine Learning Solutions with Apache Kafka in the Cloud Using Python
A guide for programmers interested in developing cloud native applications using Ballerina Swan Lake
Apache Flume: Distributed Log Collection for Hadoop - Second Edition
Lightning-Fast Big Data Analysis
Mastering Apache Flink
Beginning Apache Spark 2
Effective Kafka
Big Data SMACK
Stream Processing with Apache Flink
Kafka Streams in Action
A Hands-On Guide to Building Robust and Scalable Event-Driven Applications with Code Examples in Java
Learning Apache Drill
Transactional Machine Learning with Data Streams and AutoML
Spark in Action
Cassandra: The Definitive Guide
Query and Analyze Distributed Data Sources with SQL
Hands-On Industrial Internet of Things
Covers Apache Spark 3 with Examples in Java, Python, and Scala
Real-time apps and microservices with the Kafka Streams API
Learning Apache OpenWhisk
Apache Kafka
Building Data Streaming Applications with Apache Kafka
Beginning PHP, Apache, MySQL Web Development
A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka
Mastering Structured Streaming and Spark Streaming
Mastering Kafka Streams and ksqlDB
Apache Hive Essentials
OpenShift for Developers
With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library
Apache Kafka 1.0 Cookbook
Machine Learning
Apache Tomcat 7
Real-Time Data and Stream Processing at Scale
Distributed Data at Web Scale
Learning Apache Cassandra - Second Edition

Create a powerful Industrial IoT infrastructure using Industry 4.0

HEAVEN JUNE

Fundamentals, Implementation, and Operation of Streaming Applications Simon and Schuster
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
Over 100 practical recipes on using distributed enterprise messaging to handle real-time data
"O'Reilly Media, Inc."
Definitive guide to lightning fast data processing for distributed systems with Apache FlinkAbout This Book* Build your experitse in processing realtime data with Apache Flink and its ecosystem* Gain insights into the working of all components of Apache Flink such as FlinkML, Gelly, and Table APIFilled with real world use cases,* Your guide to take advantage of Apache Flink for solving real world problemsWho This Book Is ForBig data developers who are looking to process batch and real-time data on distributed systems. Basic knowledge of Hadoop and big data is assumed. Reasonable knowledge of Java or Scala is expected.What You Will Learn* Learn how to build end to end real time analytics projects* Integrate with existing big data stack and utilize existing infrastructure.* Build predictive analytics applications using FlinkML* Use graph library to perform graph querying and search.In DetailWith the advent of massive computer systems, organizations in different domains generate large amounts of data at a realtime basis. The latest entrant to big data processing, Apache Flink, is designed to process continuous streams of data at a lightning fast pace.This book will be your definitive guide to batch and stream data processing with Apache Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing batch and streaming datasets. Bringing the power of SQL to Flink, this book will then explore the Table API for querying and manipulating data. In the latter half of the book, readers will get to learn the remaining ecosystem of Apache Flink to achieve complex tasks such as event processing, machine learning, and graph processing. The final part of the book would consist of topics such as scaling Flink solutions, performance optimization and integrating Flink with other tools such as ElasticSearch.Whether you want to dive deeper into Apache Flink, or want to investigate how to get more out of this powerful technology, you'll find everything inside
Learning Apache Kafka Second Edition Packt Publishing Ltd
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.
Build Frictionless and Elastic Machine Learning Solutions with Apache Kafka in the Cloud Using Python O'Reilly Media
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

A guide for programmers interested in developing cloud native applications using Ballerina Swan Lake Simon and Schuster

Learning Apache Kafka - Second Edition

Apache Flume: Distributed Log Collection for Hadoop - Second Edition "O'Reilly Media, Inc."

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

Lightning-Fast Big Data Analysis "O'Reilly Media, Inc."

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

Mastering Apache Flink Packt Pub Limited

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.

Beginning Apache Spark 2 O'Reilly Media

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.

Effective Kafka Packt Pub Limited

Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you

how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active in big data, Hadoop, and Java but who are new to the Apache Spark platform.

Big Data SMACK O'Reilly Media

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

Stream Processing with Apache Flink Packt Publishing Ltd

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

Kafka Streams in Action "O'Reilly Media, Inc."

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.

A Hands-On Guide to Building Robust and Scalable Event-Driven Applications with Code Examples in Java Packt Publishing Ltd

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

Learning Apache Drill "O'Reilly Media, Inc."

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.

Transactional Machine Learning with Data Streams and AutoML Apress

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.

Spark in Action "O'Reilly Media, Inc."

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

Cassandra: The Definitive Guide Packt Publishing Ltd

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

Query and Analyze Distributed Data Sources with SQL Packt Publishing Ltd

Get started with Apache Flink, the open source framework that powers some of the world's largest stream processing applications. With this practical book, you'll explore the fundamental concepts of parallel stream processing and discover how this technology differs from traditional batch data processing. Longtime Apache Flink committers Fabian Hueske and Vasia Kalavri show you how to implement scalable streaming applications with Flink's DataStream API and continuously run and maintain these applications in operational environments. Stream processing is ideal for many use cases, including low-latency ETL, streaming analytics, and real-time dashboards as well as fraud detection, anomaly detection, and alerting. You can process continuous data of any kind, including user interactions, financial transactions, and IoT data, as soon as you generate them. Learn concepts and challenges of distributed stateful stream processing Explore Flink's system architecture, including its event-time processing mode and fault-tolerance model Understand the fundamentals and building blocks of the DataStream API, including its time-based and statefuloperators Read data from and write data to external systems with exactly-once consistency Deploy and configure Flink clusters Operate continuously running streaming applications

Hands-On Industrial Internet of Things "O'Reilly Media, Inc."

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 ksqlDB and perform push and pull queries Deploy your Kafka Streams and ksqlDB applications to production

Related with Learning Apache Kafka Second Edition Garg Nishant:
© Learning Apache Kafka Second Edition Garg Nishant Articles About Radiologic Technology
© Learning Apache Kafka Second Edition Garg Nishant As9100d Internal Auditor Training
© Learning Apache Kafka Second Edition Garg Nishant Articles Of Confederation Answer Key