

Streaming Architecture New Designs Using Apache Kafka And Mapr Streams

This volume focuses on the theory and practice of data stream management, and the novel challenges this emerging domain poses for data-management algorithms, systems, and applications. The collection of chapters, contributed by authorities in the field, offers a comprehensive introduction to both the algorithmic/theoretical foundations of data streams, as well as the streaming systems and applications built in different domains. A short introductory chapter provides a brief summary of some basic data streaming concepts and models, and discusses the key elements of a generic stream query processing architecture. Subsequently, Part I focuses on basic streaming algorithms for some key analytics functions (e.g., queries, norms, join aggregates, heavy hitters) over streaming data. Part II then examines important techniques for basic stream mining tasks (e.g., clustering, classification, frequent items). Part III discusses a number of advanced topics on stream processing algorithms, and Part IV focuses on system and language aspects with surveys of influential system prototypes and language designs. Part V then presents some representative applications of streaming techniques in different domains (e.g., network management, financial analytics). Finally, the volume concludes with an overview of current data streaming products and an analysis of emerging trends in big data analytics, and complex event processing, and a discussion of future directions in this exciting field. The book provides a comprehensive overview of core concepts and technological foundations, as well as various systems and applications, and is of particular interest to students, lecturers and researchers in the area of data stream management.

Before you can build analytics tools to gain quick insights, you first need to know how to process data in real time. With this practical guide, developers familiar with Apache Spark will learn how to put this in-memory framework to use for streaming data. You'll discover how Spark enables you to write streaming jobs in almost the same way you write batch jobs. Authors Gerard Mass and François Garillot help you explore the theoretical underpinnings of Apache Spark. This comprehensive guide features two sections that compare and contrast the streaming APIs Spark now supports: the original Spark Streaming library and the newer Structured Streaming API. Learn fundamental stream processing concepts and examine different streaming architectures Explore Structured Streaming through practical examples; learn different aspects of stream processing in detail Create and operate streaming jobs and applications with Spark Streaming; integrate Spark Streaming with other Spark APIs Learn advanced Spark Streaming techniques, including approximation algorithms and machine learning algorithms Compare Apache Spark to other stream processing projects, including Apache Storm, Apache Flink, and Apache Kafka Streams

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 layer New messaging technologies, including Apache Kafka and MapR Streams, with links to sample code Technology choices for streaming analytics: Apache Spark Streaming, Apache Flink, Apache Storm, and Apache Apex How stream-based architectures are helpful to support microservices Specific 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. Software development is a fast-moving field, and this book is constantly being updated to reflect the latest changes in the industry. What happens when events and data streams help you discover new technology sources to enhance existing businesses or drive new markets? What technologies and architectural patterns can position your company for opportunities enabled by flow? James Urquhart, global field CTO at VMware, guides enterprise architects, software developers, and product managers through the process. Learn the benefits of flow dynamics when businesses, governments, and other institutions integrate via events and data streams Understand the value chain for flow integration through Wardley mapping visualization and promise theory modeling Walk through basic concepts behind today's event-driven systems marketplace Learn how today's integration patterns will influence the real-time events flow in the future Explore why companies should architect and build software today to take advantage of flow in coming years

Mahout in Action
Pro Spark Streaming
Real-World Hadoop
19th International Conference, ICA3PP 2019, Melbourne, VIC, Australia, December 9–11, 2019, Proceedings, Part II
Algorithms and Applications
AWS Certified Solutions Architect – Associate Guide
Streaming Architecture

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 security delivers capabilities make it a perfect source for stream processing systems

"A comprehensive overview of the challenges teams face when moving to microservices, with industry-tested solutions." - Tim Moore, Lightbend 44 reusable patterns to develop and deploy reliable production-quality microservices-based applications, with worked examples in Java Key Features 44 design patterns for building and deploying microservices applications Drawing on decades of unique experience from author and microservice architecture pioneer Chris Richardson A pragmatic approach to the benefits and the drawbacks of microservices architecture Save service decomposition, transaction management, and inter-service communication. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About The Book Microservices Patterns teaches you 44 reusable patterns to reliably develop and deploy production-quality microservices-based applications. This invaluable set of design patterns builds on decades of distributed system experience, adding new patterns for composing services into systems that scale and perform under real-world conditions. More than just a patterns catalog, this practical guide with worked examples of fers industry-tested advice to help you design, implement, test, and deploy your microservices-based application. What You Will Learn How (and why) to use free services in architecture Service decomposition strategies Transaction management and querying patterns Effective testing strategies Deployment patterns This Book Is Written For enterprise developers familiar with standard enterprise application architecture. Examples are in Java. About The Author Chris Richardson is a Java Champion, a JavaOne rock star, author of Manning's POJOs in Action, and creator of the original CloudFoundry.com. Table of Contents Escaping monolithic hell Decomposition strategies Interprocess communication in a microservice architecture Managing transactions with sagas

Designing business logic in a microservice architecture Developing business logic with event sourcing Implementing queries in a microservice architecture External API patterns Testing microservices: part 1 Testing microservices; part 2 Developing production-ready services Deploying microservices Refactoring to microservices Provides comprehensive coverage of the current state of IoT, focusing on data processing infrastructure and techniques Written by experts in the field, this book addresses the IoT technology stack, from connectivity through data platforms to end-user case studies, and considers the tradeoffs between business needs and data security and privacy throughout. There is a particular emphasis on data processing technologies that enable the extraction of actionable insights from data to inform improved decision making. These include artificial intelligence techniques such as stream processing, deep learning and knowledge graphs, as well as data interoperability and the benefits of privacy-preserving data security and supporting IoT ecosystems; creating and supporting data mining of sensor datasets; and crowd-sourcing. The book also presents several sections featuring use cases across a range of application areas such as smart energy, transportation, smart factories, and more. The book concludes with a chapter on key considerations when deploying IoT technologies in the enterprise, followed by a brief review of future research directions and challenges. The Internet of Things: From Data to Insight Provides a comprehensive overview of the Internet of Things technology stack with focus on data processing and processing to presentation. For decision making. Explains how IoT technology is applied in practice and the benefits being delivered. Acquaints readers that are new to the area with concepts, components, technologies, and vertically related to and enabled by IoT Gives IoT specialists a deeper insight into data and decision-making aspects as well as novel technologies and application areas Analyzes and presents important emerging technologies for the IoT arena Shows how different objects and devices can be connected to decision making processes at various levels of abstraction The Internet of Things: From Data to Insight will appeal to a wide audience, including IT and network specialists seeking a broad and complete understanding of IoT, CIOs and CIO teams, researchers in IoT and related fields, final year undergraduates, graduate students, post-graduates, and IT and science media professionals.

Have you heard about the tremendous success Amazon and Netflix have had by switching to a microservice architecture? Are you wondering how this can benefit your company? Or are you skeptical about how it might work? If you've answered yes to any of these questions, this practical book will benefit you. You'll learn how to take advantage of the microservice architectural style for building systems, and learn from the experiences of others to adopt and execute this approach most successfully.

Real-Time Data and Stream Processing at Scale

Measuring, Monitoring, and Managing Your Business

Event Streams in Action

Stream Processing with Apache Spark

I Heart Logs

Data Streams

Kafka: The Definitive Guide

If you're a business team leader, CIO, business analyst, or developer interested in how Apache Hadoop and Apache HBase-related technologies can address problems involving large-scale data in cost-effective ways, this book is for you. Using real-world stories and situations, authors Ted Dunning and Ellen Friedman show Hadoop newcomers and seasoned users alike how NoSQL databases and Hadoop can solve a variety of business and research issues. You'll learn about early decisions and pre-planning that can make the process easier and more productive. If you're already using these technologies, you'll discover ways to gain the full range of benefits possible with Hadoop. While you don't need a deep technical background to get started, this book does provide expert guidance to help managers, architects, and practitioners succeed with their Hadoop projects. Examine a day in the life of big data: India's ambitious Aadhaar project Review tools in the Hadoop ecosystem such as Apache's Spark, Storm, and Drill to learn how they can help you Pick up a collection of technical and strategic tips that help you succeed with your Hadoop project Learn from several prototypical Hadoop use cases, based on how organizations have actually applied the technology Explore real-world stories that reveal how MapR customers combine use cases when putting Hadoop and NoSQL to work, including in production While wireless technologies continue to provide an array of new challenges and multi-domain applications for business processes and solutions, there still remains to be a comprehensive understanding of its various dimensions and environments. Security, Design, and Architecture for Broadband and Wireless Network Technologies provides a discussion on the latest research achievements in wireless networks and broadband technology. Highlighting new trends, applications, developments, and standards, this book is essential for next generation researchers and practitioners in the ICT field.

More useful techniques, tips, and tricks for harnessing the power of the new generation of powerful GPUs.

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 environments. 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 Cretz 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 Static 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

Stream Processing for Real Time and Beyond

Programming Techniques for High-performance Graphics and General-purpose Computation

Design and secure highly available, cost-effective data streaming applications with Amazon Kinesis

A New Architecture Design for P2P Live Streaming in Wireless Local Area Networks

Patterns and Paradigms for Scalable, Reliable Services

Stream Processing with Apache Flink

Scalable Data Streaming with Amazon Kinesis

Explore Kinesis managed services such as Kinesis Data Streams, Kinesis Data Analytics, Kinesis Data Firehose, and Kinesis Video Streams with the help of practical use cases Key FeaturesGet well versed with the capabilities of Amazon KinesisExplore the monitoring, scaling, security, and deployment patterns of various Amazon Kinesis servicesLearn how other Amazon Web Services and third-party applications such as Splunk can be used as destinations for Kinesis dataBook Description durable, and highly available purpose-built data streaming services. This data streaming service provides APIs and client SDKs that enable you to produce and consume data at scale. Scalable Data Streaming with Amazon Kinesis begins with a quick overview of the core concepts of data streams, along with the essentials of the AWS Kinesis landscape. You'll then explore the requirements of the use case shown through the book to help you get started and cover the key pain points that go to grips with the architectural components of Kinesis, understand how they are configured to build data pipelines, and delve into the applications that connect to them for consumption and processing. You'll also build a Kinesis data pipeline from scratch and learn how to implement and apply practical solutions. Moving on, you'll learn how to configure Kinesis on a cloud platform. Finally, you'll learn how other AWS services can be integrated into Kinesis. These services include third-party applications such as Splunk. By the end of this AWS book, you'll be able to build and deploy your own Kinesis data pipelines with Kinesis Data Streams (KDS), Kinesis Data Firehose (KFH), Kinesis Video Streams (KVS), and Kinesis Data Analytics (KDA). What you will learnGet to grips with data streams, decoupled design, and real-time stream processingUnderstand the properties of KFH that differentiate it from other Kinesis servicesMonitor and scale KDS using CloudWatch metrics (IAM)Deploy KVS as infrastructure as code (IaC)Integrate services such as Redshift, Dynamo Database, and Splunk into KinesisWho this book is for This book is for solutions architects, developers, system administrators, data engineers, and data scientists looking to evaluate and choose the most performant, secure, scalable, and cost-effective data streaming technology to overcome their data ingestion and processing challenges on AWS. Prior knowledge of cloud architectures on the data stream scenario, input arrives very rapidly and there is limited memory to store the input. Algorithms have to work with one or few passes over the data, space less than linear in the input size or time significantly less than the input size. In the past few years, a new theory has emerged for reasoning about algorithms that work within these constraints on space, time, and number of passes. Some of the methods rely on metric embeddings, pseudo-random computation applications for this scenario include IP network traffic analysis, mining text message streams and processing massive data sets in general. Researchers in Theoretical Computer Science, Databases, IP Networking and Computer Systems are working on the data stream challenges.

Organizations today often struggle to balance business requirements with ever-increasing volumes of data. Additionally, the demand for leveraging large-scale, real-time data is growing rapidly among the most competitive digital industries. Conventional system architectures may not be up to the task. With this practical guide, you'll learn how to leverage large-scale data usage across the business units in your organization using the principles of event-driven microservices. Author of event-driven microservice-powered organization. You'll reconsider how data is produced, accessed, and propagated across your organization. Learn powerful yet simple patterns for unlocking the value of this data. Incorporate event-driven design and architectural principles into your own systems. And completely rethink how your organization delivers value by unlocking near-real-time access to data at scale. You'll learn: How to leverage event-driven architectures to deliver exceptional event-driven designs Architectural patterns to ensure success both within and between teams in your organization Application patterns for developing powerful event-driven microservices Components and tooling required to get your microservice ecosystem off the ground

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 Malchimp, explains important stream processing concepts against 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 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

Kafka Streams in Action

Image Processing Using FPGAs

Design Patterns for Cloud Native Applications

Quickly Learn the Art of Writing Efficient Big Data Applications with Apache Spark

Aligning Principles, Practices, and Culture

Algorithms and Architectures for Parallel Processing

Kafka Streams - Real-time Stream Processing

With the immense cost savings and scalability the cloud provides, the rationale for building cloud native applications is no longer in question. The real issue is how. With this practical guide, developers will learn about the most commonly used design patterns for building cloud native applications using APIs, data, events, and streams in both greenfield and brownfield development. You'll learn how to incrementally design, develop, and deploy large and effective cloud native applications that you can manage and maintain at scale with minimal cost, time, and effort challenges you might encounter at each step. Learn the fundamentals of cloud native applications Explore key cloud native communication, connectivity, and composition patterns Learn decentralized data management techniques Use event-driven architecture to build distributed and scalable cloud native applications Explore the most commonly used patterns for API management and consumption Examine some of the tools and technologies you'll need for building cloud native systems

Tips, techniques, and trends on how to use dashboard technology to optimize business performance Business performance management is a hot new management discipline that delivers tremendous value when supported by information technology. Through case studies and industry research, this bookshows how leading companies are using performance dashboards to execute strategy, optimize business processes, and improve performance. Wayne W. Eckerson (Hingham, MA) is the Director of Research for TheData Warehousing Institute (TDWI), the leading association for data warehousing, education, training, andresearch. He is a columnist for SearchCIO.com, DM Review/Application Development Trends, the Business Intelligence Journal, and TDWI Case Studies & Solution.

A practical guide for solving complex data processing challenges by applying the best optimizations techniques in Apache Spark. Key Features Learn about the core concepts and the latest developments in Apache Spark. Master writing efficient big data applications with Spark's built-in modules for SQL, Streaming, Machine Learning and Graph analysis Get introduced to a variety of optimizations based on the actual experience Book Description Apache Spark is a flexible framework that allows processing of batch and real-time data. Its unified engine has made it easy for you to build a data pipeline from end-to-end in a single framework. This practical guide provides a quick start to the Spark 2.0 architecture and its components. It teaches you how to set up Spark on your local machine. As we move ahead, you will be introduced to resilient distributed datasets (RDDs) and DataFrame APIs, and their corresponding transform and action APIs to debug slow-running applications. You will also go through Spark's built-in modules for SQL, streaming, machine learning, and graph analysis. Finally, the book will lay out the best practices and optimization techniques that are key for writing efficient Spark applications. By the end of this book, you will have a sound fundamental understanding of the Apache Spark framework and you will be able to write and optimize Spark applications. What you will learn Learn core concepts such as RDDs, DataFrames, transformations, and more Set up a Spark development env of a Spark application Explore built-in modules for SQL, streaming, ML, and graph analysis Optimize your Spark job for better performance Who this book is for If you are a big data enthusiast and love processing huge amount of data, this book is for you. If you are a data engineer and looking for the best optimization techniques for your Spark applications, then you will find this book helpful. This book also helps data scientists who want to implement their machine learning algorithms in Spark. You need to have a basic understanding of any one of the program's design and administrator 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 practical/practical 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 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 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 text messaging system and then provides a thorough introduction to Apache Kafka and its internal details. The second part of the book takes you through 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

Data Stream Management

Security, Design, and Architecture for Broadband and Wireless Network Technologies

Building Event-Driven Microservices

Streaming Data

Designing Distributed Systems

Big Data

Fundamentals, Implementation, and Operation of Streaming Applications

The two-volume set LNCS 11944-11945 contains the proceedings of the 19th International Conference on Algorithms and Architectures for Parallel Processing, ICA3PP 2019, held in Melbourne, Australia, in December 2019. The 73 full and 29 short papers presented were carefully reviewed and selected from 251 submissions. The papers are organized in topical sections on: Parallel and Distributed Architectures, Software Systems and Programming Models, Distributed and Parallel and Network-based Computing, Big Data and its Applications, Distributed and Parallel Algorithms, Applications of Distributed and Parallel Computing, Service Dependability and Security, IoT and CPS Computing, Performance Modelling and Evaluation.

Learn the right cutting-edge skills and knowledge to leverage Spark Streaming to implement a wide array of real-time, streaming applications. This book walks you through end-to-end real-time application development using real-world applications, data, and code. Taking an application-first approach, each chapter introduces use cases from a specific industry and uses publicly available datasets from that domain to unravel the intricacies of production-grade design and implementation. The domains covered in Pro Spark Streaming include social media, the sharing economy, finance, online advertising, telecommunication, and IoT. In the last few years, Spark has become synonymous with big data processing, DStreams enables the underlying Spark processing engine to support streaming analysis with a novel micro-batch processing model. Pro Spark Streaming by Zubair Nahi will enable you to become a specialist of latest sensitive applications by leveraging the key features of DStreams, micro-batch processing, and functional programming. To this end, the book includes ready-to-deploy examples and actual code. Pro Spark Streaming will act as the bible of Spark Streaming. What You'll Learn Discover Spark Streaming application development and best practices Work with the low-level details of discretized streams Optimize production-grade deployments of Spark Streaming via configuration recipes and instrumentation using Graphite, collectd, and Nagios Ingest data from disparate sources including MQTT, Flume, Kafka, Twitter, and a custom HTTP Receiver Ingest and couple with HBase, Cassandra, and Redis Take advantage of design patterns for side-effects and maintaining state across the Spark Streaming micro-batch model Implement real-time and scalable ETL using data frames, SparkSQL, Hive, and SparkR Use streaming machine learning, predictive analytics, and recommendations Mesh batch processing with stream processing via the Lambda architecture Who This Book Is For Data scientists, big data experts, BI analysts, and data architect

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 layer New messaging technologies, including Apache Kafka and MapR Streams, with links to sample code Technology choices for streaming analytics: Apache Spark Streaming, Apache Flink, Apache Storm, and Apache Apex How stream-based architectures are helpful to support microservices Specific 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 Stream Data introduces the concepts and requirements of streaming and real-time data systems. The book is an idea-rich tutorial that teaches you to think about how to efficiently interact with fast-flowing data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology As humans, we're constantly filtering and deciphering the information streaming toward us. In the same way, streaming data applications can accomplish amazing tasks like reading live location data to recommend nearby services, tracking faults with machinery in real time, and sending digital receipts before your customer even gets the receipt. Real-time data is the right way to build new applications that have the right time to use. About the Book Stream Data is an idea-rich tutorial that teaches you to think about how to efficiently interact with fast-flowing data. Through practical examples and illustrations, you'll explore design for applications that read, analyze, share, and store streaming data. Along the way, you'll discover the roles of key technologies like Spark, Storm, Kafka, Flink, RabbitMQ, and more. This book offers the perfect balance between big-picture theory and implementation details. What's Inside The right way to collect real-time data Architecting a streaming pipeline Analyzing the data Which technologies to use and when About the Reader Writing for developers familiar with relational database concepts. No experience with streaming or real-time applications required. About the Author Andrew Psaltis is a software engineer focused on massively scalable real-time analytics. Table of Contents PART 1 - A NEW HOLISTIC APPROACH Introducing streaming data Getting data from clients: data ingestion Transporting the data from collection tier: decoupling the data pipeline Analyzing streaming data Algorithms for data analysis Storing the analyzed or collected data Making the data available Consumer device capabilities and limitations Accessing the data PART 2 - TAKING IT REAL WORLD Analyzing Meetup RSVPs in real time

Team Topologies

With examples in Java

The Internet of Things

The ultimate exam guide to AWS Solutions Architect certification

Building Data Streaming Applications with Apache Kafka

Streaming Architecture - A Clear and Concise Reference

Processing High-Speed Data Streams

In Team Topologies DevOps consultants Matthew Skelton and Manuel Pais share secrets of successful team patterns and interactions to help readers choose and evolve the right team patterns for their organization, making sure to keep the software healthy and optimize value streams. Team Topologies will help readers discover:

- Team patterns used by successful organizations.**
- Common team patterns to avoid with modern software systems.**
- When and why to use different team patterns**
- How to evolve teams effectively.**
- How to split software and align to teams.**

This book presents a selection of papers representing current research on using field programmable gate arrays (FPGAs) for realising image processing algorithms. These papers are reprints of papers selected for a Special Issue of the Journal of Imaging on image processing using FPGAs. A diverse range of topics is covered, including parallel soft processors, memory management, image filters, segmentation, clustering, image analysis, and image compression. Applications include traffic sign recognition for autonomous driving, cell detection for histopathology, and video compression. Collectively, they represent the current state-of-the-art on image processing using FPGAs.

Summary Mahout in Action is a hands-on introduction to machine learning with Apache Mahout. Following real-world examples, the book presents practical use cases and then illustrates how Mahout can be applied to solve them. Includes a free audio- and video-enhanced ebook. About the Technology A computer system that learns and adapts as it collects data can be really powerful. Mahout, Apache's open source machine learning project, captures the core algorithms of recommendation systems, classification, and clustering in ready-to-use, scalable libraries. With Mahout, you can immediately apply to your own projects the machine learning techniques that drive Amazon, Netflix, and others. About this Book This book covers machine learning using Apache Mahout. Based on experience with real-world applications, it introduces practical use cases and illustrates how Mahout can be applied to solve them. It places particular focus on issues of scalability and how to apply these techniques against large data sets using the Apache Hadoop framework. This book is written for developers familiar with Java – no prior experience with Mahout is assumed. Owners of a Manning eBook purchased anywhere in the world can download a free eBook from Manning.com at any time. They can do so multiple times and in any or all formats available (PDF, ePub or Kindle). To do so, customers must register their printed copy on Manning's site by creating a user account and then following instructions printed on the eBook registration insert at the front of the book. What's Inside Use group data to make individual recommendations Find logical clusters within your data Filter and refine with on-the-fly classification Free audio and video extras Table of Contents Meet Apache Mahout PART 1 RECOMMENDATIONS Introducing recommenders Representing recommender data Making recommendations Taking recommenders to production

Distributing recommendation computations PART 2 CLUSTERING Introduction to clustering Representing data Clustering algorithms in Mahout Evaluating and improving clustering quality Taking clustering to production Real-world applications of clustering PART 3 CLASSIFICATION Introduction to classification Training a classifier Evaluating and tuning a classifier Deploying a classifier Case study: Shop It To Me

Learn from the AWS subject matter experts, apply real-world scenarios and clear the AWS Certified Solutions Architect – Associate exam Key FeaturesBuild highly reliable and scalable workloads on the AWS platformPass the exam in less time and with confidenceGet up and running with building and managing applications on the AWS platformBook Description Amazon Web Services (AWS) is currently the leader in the public cloud market. With an increasing global interest in leveraging cloud infrastructure, the AWS Cloud from Amazon remains the cutting-edge platform for architecting, building, and deploying web-scale cloud applications. As more the rate of cloud platform adoption increases, so does the need for cloud certification. The AWS Certified Solution Architect – Associate Guide is your one-stop solution to gaining certification. Once you have grasped what AWS and its prerequisites are, you will get insights into different types of AWS services such as Amazon S3, EC2, VPC, SNS, and more to get you prepared with core Amazon services. You will then move on to understanding how to design and deploy highly scalable applications. Finally, you will study security concepts along with the AWS best practices and mock papers to test your knowledge. By the end of this book, you will not only be fully prepared to pass the AWS Certified Solutions Architect – Associate exam but also capable of building secure and reliable applications. What you will learnExplore AWS terminology and identify and access managementAcquaint yourself with important cloud services and features in categories such as compute, network, storage, and databasesDefine access control to secure AWS resources and set up efficient monitoringBack up your database and ensure high availability by understanding all of the database-related services in the AWS CloudIntegrate AWS with your applications to meet and exceed non-functional requirementsBuild and deploy cost-effective and highly available applicationsWho this book is for The AWS Certified Solutions Architect –Associate Guide is for you if you are an IT professional or Solutions Architect wanting to pass the AWS Certified Solution Architect – Associate 2018 exam. This book is also for developers looking to start building scalable applications on AWS

Microservices Patterns

Analyze and Develop Data Streaming Architecture for Motion Estimation Using FPGA

From Data to Insight

Streaming Systems

Real-time apps and microservices with the Kafka Streams API

Flow Architectures

Principles and Best Practices of Scalable Realtime Data Systems

Why a book about logs? That's easy: the humble log is an abstraction that lies at the heart of many systems, from NoSQL databases to cryptocurrencies. Even though most engineers don't think much about them, this short book shows you why logs are worthy of your attention. Based on his popular blog posts, LinkedIn principal engineer Jay Kreps shows you how logs work in distributed systems, and then delivers practical applications of these concepts in a variety of common uses—data integration, enterprise architecture, real-time stream processing, data system design, and abstract computing models. Go ahead and take the plunge with logs; you're going love them. Learn how logs are used for programmatic access in databases and distributed systems Discover solutions to the huge data integration problem when more data of more varieties meet more systems Understand why logs are at the heart of real-time stream processing Learn the role of a log in the internals of online data systems Explore how Jay Kreps applies these ideas to his own work on data infrastructure systems at LinkedIn

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 Markhede, 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

In the race to compete in today's fast-moving markets, large enterprises are busy adopting new technologies for creating new products, processes, and business models. But one obstacle on the road to digital transformation is placing too much emphasis on technology, and not enough on the types of processes technology enables. What if different lines of business could build their own services and applications—and decision-making was distributed rather than centralized? This report explores the concept of a digital business platform as a way of empowering individual business sectors to act on data in real time. Much innovation in a digital enterprise will increasingly happen at the edge, whether it involves business users (from marketers to data scientists) or IoT devices. To facilitate the process, your core IT team can provide these sectors with the digital tools they need to innovate quickly. This report explores: Key cultural and organizational changes for developing business capabilities through cross-functional product teams A platform for integrating applications, data sources, business partners, clients, mobile apps, social networks, and IoT devices Creating internal API programs for building innovative edge services in low-code or no-code environments Tools including Integration Platform as a Service, Application Platform as a Service, and Integration Software as a Service The challenge of integrating microservices and serverless architectures Event-driven architectures for processing and reacting to events in real time You'll also learn about a complete pervasive integration solution as a core component of a digital business platform to serve every audience in your organization.

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

Real-time event systems with Kafka and Kinesis

Introduction to Apache Flink

The What, Where, When, and How of Large-Scale Data Processing

New Designs Using Apache Kafka and MapR Streams

Understanding the real-time pipeline

Event Data, Stream Processing, and Data Integration

The Zen of Real-Time Analytics Using Apache Spark

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.

How do you select, collect, align, and integrate Streaming Architecture data and information for tracking daily operations and overall organizational performance, including progress relative to strategic objectives and action plans? How do you use Streaming Architecture data and information to support organizational decision making and innovation? Who will be responsible for deciding whether Streaming Architecture goes ahead or not after the initial investigations? Do you monitor the effectiveness of your Streaming Architecture activities? Are there any constraints known that bear on the ability to perform Streaming Architecture work? How is the team addressing them? Defining, designing, creating, and implementing a process to solve a challenge or meet an objective is the most valuable role... In EVERY group, company, organization and department. Unless you are talking a one-time, single-use project, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, "What are we really trying to accomplish here? And is there a different way to look at it?" This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc... - they are the people who rule the future. They are the person who asks the right questions to make Streaming Architecture investments work better. This Streaming Architecture All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Streaming Architecture Self-Assessment. Featuring 676 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Streaming Architecture improvements can be made. In using the questions you will be better able to: - diagnose Streaming Architecture projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Streaming Architecture and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Streaming Architecture Scorecard, you will develop a clear picture of which Streaming Architecture areas need attention. Your purchase includes access details to the Streaming Architecture self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. You will receive the following contents with New and Updated specific criteria: - The latest quick edition of the book in PDF - The latest complete edition of the book in PDF, which criteria correspond to the criteria in... - The Self-Assessment Excel Dashboard, and... - Example pre-filled Self-Assessment Excel Dashboard to get familiar with results generation ...plus an extra, special, resource that helps you with project managing. INCLUDES LIFETIME SELF ASSESSMENT UPDATES Every self assessment comes with Lifetime Updates and Lifetime Free Updated Books. Lifetime Updates is an industry-first feature which allows you to receive verified self assessment updates, ensuring you always have the most accurate information at your fingertips.

Summary Big Data teaches you to build big data systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data. It describes a scalable, easy-to-understand approach to big data systems that can be built and run by a small team. Following a realistic example, this book guides readers through the theory of big data systems, how to implement them in practice, and how to deploy and operate them once they're built. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book Web-scale applications like social networks, real-time analytics, or e-commerce sites deal with a lot of data, whose volume and velocity exceed the limits of traditional database systems. These applications require architectures built around clusters of machines to store and process data of any size, or speed. Fortunately, scale and simplicity are not mutually exclusive. Big Data teaches you to build big data systems using an architecture designed specifically to capture and analyze web-scale data. This book presents the Lambda Architecture, a scalable, easy-to-understand approach that can be built and run by a small team. You'll explore the theory of big data systems and how to implement them in practice. In addition to discovering a general framework for processing big data, you'll learn specific technologies like Hadoop, Storm, and NoSQL databases. This book requires no previous exposure to large-scale data analysis or NoSQL tools. Familiarity with traditional databases is helpful. What's Inside Introduction to big data systems Real-time processing of web-scale data Tools like Hadoop, Cassandra, and Storm Extensions to traditional database skills About the Authors Nathan Marz is the creator of Apache Storm and the originator of the Lambda Architecture for big data systems. James Warren is an analytics architect with a background in machine learning and scientific computing. Table of Contents A new paradigm for Big Data PART 1 BATCH LAYER Data model for Big Data Data model for Big Data: Illustration Data storage on the batch layer Data storage on the batch layer: Illustration Batch layer Batch layer: Illustration An example batch layer: Architecture and algorithms An example batch layer: Implementation PART 2 SERVING LAYER Serving layer: Illustration PART 3

SPEED LAYER Realtime views Realtime views: Illustration Queuing and stream processing Queuing and stream processing: Illustration Micro-batch stream processing Micro-batch stream processing: Illustration Lambda Architecture in depth There's growing interest in learning how to analyze streaming data in large-scale systems such as web traffic, financial transactions, machine logs, industrial sensors, and many others. But analyzing data streams at scale has been difficult to do well—until now. This practical book delivers a deep introduction to Apache Flink, a highly innovative open source stream processor with a surprising range of capabilities. Authors Ellen Friedman and Kostas Tzoumas show technical and nontechnical readers alike how Flink is engineered to overcome significant tradeoffs that have limited the effectiveness of other approaches to stream processing. You'll also learn how Flink has the ability to handle both stream and batch data processing with one technology. Learn the consequences of not doing streaming well—in retail and marketing, IoT, telecom, and banking and finance Explore how to design data architecture to gain the best advantage from stream processing Get an overview of Flink's capabilities and features, along with examples of how companies use Flink, including in production Take a technical dive into Flink, and learn how it handles time and stateful computation Examine how Flink processes both streaming (unbounded) and batch (bounded) data without sacrificing performance

Organizing Business and Technology Teams for Fast Flow

Microservices Architecture

Performance Dashboards

Apache Spark Quick Start Guide

Mastering Kafka Streams and ksqldb

GPU Gems 2

Mastering Structured Streaming and Spark Streaming

Streaming data is a big deal in big data these days. As more and more businesses seek to tame the massive unbounded data sets that pervade our world, streaming systems have finally reached a level of maturity sufficient for mainstream adoption. With this practical guide, data engineers, data scientists, and developers will learn how to work with streaming data in a conceptual and platform-agnostic way. Expanded from Tyler Akidau's popular blog posts "Streaming 101" and "Streaming 102", this book takes you from an introductory level to a nuanced understanding of the what, where, when, and how of processing real-time data streams. You'll also dive deep into watermarks and exactly-once processing with co-authors Slava Chernyak and Reuven Lax. You'll explore: How streaming and batch data processing patterns compare The core principles and concepts behind robust out-of-order data processing How watermarks track progress and completeness in infinite datasets How exactly-once data processing techniques ensure correctness How the concepts of streams and tables form the foundations of both batch and streaming data processing The practical motivations behind a powerful persistent state mechanism, driven by a real-world example How time-varying relations provide a link between stream processing and the world of SQL and relational algebra