Streaming Architecture New Designs Using Apache Kafka And Mapr Streams

Tips, techniques, and trends on how to use dashboard technology tooptimize business performance Business performance management is a hot new management disciplinethat delivers tremendous value when supported by informationtechnology. Through case studies and industry research, this bookshows how leading companies are using performance dashboards toexecute strategy, optimize business processes, and improveperformance. Wayne W. Eckerson (Hingham, MA) is the Director of Research for TheData Warehousing Institute (TDWI), the leading association ofbusiness intelligence and data warehousing professionals worldwidethat provide high-quality, in-depth education, training, andresearch. He is a columnist for SearchCIO.com, DM Review,Application Development Trends, the Business Intelligence Journal,and TDWI Case Studies & Solution.

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

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 Adam Bellemare takes you through the process of building an 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 business value The role of microservices in supporting 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

Software development today is embracing events and streaming data, which optimizes not only how technology interacts but also how businesses integrate with one another to meet customer needs. This phenomenon, called flow, consists of patterns and standards that determine which activity and related data is communicated between parties over the internet. This book explores critical implications of that evolution: What happens when events and data streams help you discover new activity 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

This practical guide takes a hands-on approach to implementation and associated methodologies to have you up and running with all that Amazon Kinesis has to offer. You'll work with use cases and practical examples to be able to ingest, process, analyze, and stream real-time data in no time.

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., quantiles, norms, join aggregates, heavy hitters) over streaming data. Part II then examines important techniques for basic stream mining tasks (e.g., clustering, classification, frequent itemsets). Part III discusses a number of advanced topics on stream processing algorithms, and Part IV focuses on system and language aspects of data stream processing 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 new application domains (e.g. cloud computing, 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.

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 Solve 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 offers industry-tested advice to help you design, implement, test, and deploy your microservices-based application. What You Will Learn How (and why!) to use microservices architecture Service decomposition strategies Transaction management and querying patterns Effective testing strategies Deployment patterns This Book Is Written For 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

There's a lot of information about big data technologies, but splicing these technologies into an end-to-end enterprise data platform is a daunting task not widely covered. With this practical book, you'll learn how to build big data infrastructure both on-premises and in the cloud and successfully architect a modern data platform. Ideal for enterprise architects, IT managers, application architects, and data engineers, this book shows you how to overcome the many challenges that emerge during Hadoop projects. You'll explore the vast landscape of tools available in the Hadoop and big data realm in a thorough technical primer before diving into: Infrastructure: Look at all component layers in a modern data platform, from the server to the data center, to establish a solid foundation for data in your enterprise Platform: Understand aspects of deployment, operation, security, high availability, and disaster recovery, along with everything you need to know to integrate your platform with the rest of your enterprise IT Taking Hadoop to the cloud: Learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

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

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

Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relational databases, NoSQL datastores, stream or batch processors, and message brokers. What are the right choices for your application? How do you make sense of all these buzzwords? In this practical and comprehensive guide, author Martin Kleppmann helps you navigate this diverse landscape by examining the pros and cons of various technologies for processing

and storing data. Software keeps changing, but the fundamental principles remain the same. With this book, software engineers and architects will learn how to apply those ideas in practice, and how to make full use of data in modern applications. Peer under the hood of the systems you already use, and learn how to use and operate them more effectively Make informed decisions by identifying the strengths and weaknesses of different tools Navigate the trade-offs around consistency, scalability, fault tolerance, and complexity Understand the distributed systems research upon which modern databases are built Peek behind the scenes of major online services, and learn from their architectures

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

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

Buildings have often been studies whole in space, but never before have they been studied whole in time. How Buildings Learn is a masterful new synthesis that proposes that buildings adapt best when constantly refined and reshaped by their occupants, and that architects can mature from being artists of space to becoming artists of time. From the connected farmhouses of New England to I.M. Pei's Media Lab, from "satisficing" to "form follows funding," from the evolution of bungalows to the invention of Santa Fe Style, from Low Road military surplus buildings to a High Road English classic like Chatsworth—this is a far-ranging survey of unexplored essential territory. More than any other human artifacts, buildings improve with time—if they're allowed to. How Buildings Learn shows how to work with time rather than against it.

This book constitutes the thoroughly refereed post-conference proceedings of the Third International Workshop on New Frontiers in Mining Complex Patterns, NFMCP 2014, held in conjunction with ECML-PKDD 2014 in Nancy, France, in September 2014. The 13 revised full papers presented were carefully reviewed and selected from numerous submissions. They illustrate advanced data mining techniques which preserve the informative richness of complex data and allow for efficient and effective identification of complex information units present in such data. The papers are organized in the following sections: classification and regression; clustering; data streams and sequences; applications.

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.

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

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 Maas 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

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.

Apache Spark is a flexible in-memory framework that allows processing of both batch and real-time data. Its unified engine has made it quite popular for big data use cases. This book will help you to quickly get started with Apache Spark 2.0 and write efficient big data applications for a variety of use cases.

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.

Design, implement, and deliver successful streaming applications, machine learning pipelines and graph applications using Spark SQL API About This Book Learn about the design and implementation of streaming applications, machine learning pipelines, deep learning, and large-scale graph processing applications using Spark SQL APIs and Scala. Learn data exploration, data munging, and how to process structured and semi-structured data using real-world datasets and gain hands-on exposure to the issues and challenges of working with noisy and "dirty" real-world data. Understand design considerations for scalability and performance in web-scale Spark application architectures. Who This Book Is For If you are a developer, engineer, or an architect and want to learn how to use Apache Spark in a web-scale project, then this is the book for you. It is assumed that you have prior knowledge of SQL querying. A basic programming knowledge with Scala, Java, R, or Python is all you need to get started with this book. What You Will Learn Familiarize yourself with Spark SQL programming, including working with DataFrame/Dataset API and SQL Perform a series of hands-on exercises with different types of data sources, including CSV, JSON, Avro, MySQL, and MongoDB Perform data quality checks, data visualization, and basic statistical analysis tasks Perform data munging tasks on publically available datasets Learn how to use Spark SQL and Apache Kafka to build streaming applications Learn key performance-tuning tips and tricks in Spark SQL applications Learn key architectural components and patterns in large-scale Spark SQL applications In Detail In the past year, Apache Spark has been increasingly adopted for the development of distributed applications. Spark SQL APIs provide an optimized interface that helps developers build such applications quickly and easily. However, designing web-scale production applications using Spark SQL APIs can be a complex task. Hence, understanding the design and implementation best practices before you start your project will help you avoid these problems. This book gives an insight into the engineering practices used to design and build real-world, Spark-based applications. The book's

hands-on examples will give you the required confidence to work on any future projects you encounter in Spark SQL. It starts by familiarizing you with data exploration and data munging tasks using Spark SQL and Scala. Extensive code examples will help you understand the methods used to implement typical use-cases for various types of applications. You will get a walkthrough of the key concepts and terms that are common to streaming, machine learning, and graph applications. You will also learn key performance-tuning details including Cost Based Optimization (Spark 2.2) in Spark SQL applications. Finally, you will move on to learning how such systems are architected and deployed for a successful delivery of your project. Style and approach This book is a hands-on guide to designing, building, and deploying Spark SQL-centric production applications at scale.

Develop and manage effective real-time streaming solutions by leveraging the power of Microsoft Azure About This Book Analyze your data from various sources using Microsoft Azure Stream Analytics Develop, manage and automate your stream analytics solution with Microsoft Azure A practical guide to real-time event processing and performing analytics on the cloud Who This Book Is For If you are looking for a resource that teaches you how to process continuous streams of data in real-time, this book is what you need. A basic understanding of the concepts in analytics is all you need to get started with this book What You Will Learn Perform real-time event processing with Azure Stream Analysis Incorporate the features of Big Data Lambda architecture pattern in real-time data processing Design a streaming pipeline for storage and batch analysis Implement data transformation and computation activities over stream of events Automate your streaming pipeline using Powershell and the .NET SDK Integrate your streaming pipeline with popular Machine Learning and Predictive Analytics modelling algorithms Monitor and troubleshoot your Azure Streaming jobs effectively In Detail Microsoft Azure is a very popular cloud computing service used by many organizations around the world. Its latest analytics offering, Stream Analytics, allows you to process and get actionable insights from different kinds of data in real-time. This book is your guide to understanding the basics of how Azure Stream Analytics works, and building your own analytics solution using its capabilities. You will start with understanding what Stream Analytics is, and why it is a popular choice for getting real-time insights from data. Then, you will be introduced to Azure Stream Analytics, and see how you can use the tools and functions in Azure to develop your own Streaming Analytics. Over the course of the book, you will be given comparative analytic guidance on using Azure Streaming with other Microsoft Data Platform resources such as Big Data Lambda Architecture integration for real time data analysis and differences of scenarios for architecture designing with Azure HDInsight Hadoop clusters with Storm or Stream Analytics. The book also shows you how you can manage, monitor, and scale your solution for optimal performance. By the end of this book, you will be well-versed in using Azure Stream Analytics to develop an efficient analytics solution that can work with any type of data. Style and approach A comprehensive guidance on developing real-time event processing with Azure Stream Analysis More and more data-driven companies are looking to adopt stream processing and streaming analytics. With this concise ebook, you{u2019}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.

"This book spans a number of interdependent and emerging topics in streaming media, offering a comprehensive collection of topics including media coding, wireless/mobile video, P2P media streaming, and applications of streaming media"--Provided by publisher.

In a large global health services company, streaming data for processing and sharing comes with its own challenges. Data science and analytics platforms need data fast, from relevant sources, to act on this data quickly and share the insights with consumers with the same speed and urgency. Join Mohammad Quraishi (Cigna) to learn why streaming data architectures are a necessity-Kafka and Hadoop are key. Mohammad outlines architectures centered around the Hadoop Platform and Kafka that were implemented to support a variety of integration and analytics requirements. Topics include: Enabling streaming to and from relational sources and files using custom frameworks that automate and speed up workflows Combining the polyglot techniques with Kafka API to support various streaming solutions Combining data driven techniques to support consumers through a simple streaming architecture and microservices How HBase, Kudu, and Kafka Streams are used to reduce latency between these microservices and frontend application APIs Enabling the consumption and sharing of data sources and results using streams Enabling Spark Structured Streaming, Flink, and Spark ML on these streams Enabling data sync between on-premises data lakes and the cloud Supporting cloud native architectures that enable machine learning in the cloud This session was recorded at the 2019 O'Reilly Strata Data Conference in San Francisco.

This excellent collection of historic decorative ornament contains 3,000 examples ranging from the cultures of the Greeks and Romans through the Victorians: chairs, thrones, crowns, heraldic emblems, altars, armor, architecture, and more.

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.

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 have helped others succeed with Hadoop 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

Summary Streaming 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 customers leave the shop. Recent advances in streaming data technology and techniques make it possible for any developer to build these applications if they have the right mindset. This book will let you join them. About the Book Streaming Data is an idea-rich tutorial that teaches you to think about efficiently interacting with fast-flowing data. Through relevant examples and illustrated use cases, you'll explore designs 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 thinking 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 Written 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

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

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

Multimedia data are used more and more widely in human being's life, e.g., videoconferencing, visual telephone, IPTV, etc. Nearly most of the applications need multimedia transmission techniques that send multimedia data from one side to another side and keep the properties of efficiency, robustness and security. Here, the efficiency denotes the time cost of transmission operations, the robustness denotes the ability to survive transmission errors or noises, and the security denotes the protection of the transmitted media content. Recently, various intelligent or innovative techniques are invented, which bring vast performance improvements to practical applications. For example, such content transmission techniques as p2p, sensor network and ad hoc network are constructed, which adaptively use the peers' properties to improve the network's resources. Multimedia adaptation techniques can adjust the multimedia data rate in order to compliant with the network's bandwidth. Scalable encryption techniques can generate the data stream that can be correctly decrypted after bit rate conversion. Ubiquitous multimedia services make the user share any kind of content anywhere. The book includes fourteen chapters highlighting current concepts, issues and emerging technologies. Distinguished scholars from many prominent research institutions around the world contribute to the book. The book covers various aspects, including not only some fundamental knowledge and the latest key techniques, but also typical applications and open issues. For example, the covered topics include the present and future video coding standards, stereo and multiview coding techniques, free-viewpoint TV techniques, wireless broadcasting techniques, media streaming techniques, wireless media transmission techniques and systems, and User-Generated Content sharing.

Get expert guidance on architecting end-to-end data management solutions with Apache Hadoop. While many sources explain how to use various components in the Hadoop ecosystem, this practical book takes you through architectural considerations necessary to tie those components together into a complete tailored application, based on your particular use case. To reinforce those lessons, the book's second section provides detailed examples of architectures used in some of the most commonly found Hadoop applications. Whether you're designing a new Hadoop application, or planning to integrate Hadoop into your existing data infrastructure, Hadoop Application Architectures will skillfully guide you through the process. This book covers: Factors to consider when using Hadoop to store and model data Best practices for moving data in and out of the system Data processing frameworks, including MapReduce, Spark, and Hive Common Hadoop processing patterns, such as removing duplicate records and using windowing analytics Giraph, GraphX, and other tools for large graph processing on Hadoop Using workflow orchestration and scheduling tools such as Apache Oozie Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume Architecture examples for clickstream analysis, fraud detection, and data warehousing

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 key aspects of privacy, security and trust. Additional aspects covered include: creating and supporting IoT ecosystems; edge computing; data mining of sensor datasets; and crowd-sourcing, amongst others. 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 driven aspects from data modelling 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 verticals 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.

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

Copyright: 20dff49ee031f458874e5102e87f9854