Apache Kafka Apache Mesos

Right here, we have countless ebook **Apache Kafka Apache Mesos** and collections to check out. We additionally offer variant types and as a consequence type of the books to browse. The standard book, fiction, history, novel, scientific research, as capably as various extra sorts of books are readily approachable here.

As this Apache Kafka Apache Mesos , it ends up instinctive one of the favored book Apache Kafka Apache Mesos collections that we have. This is why you remain in the best website to look the incredible books to have.

Designing Big Data Platforms - Yusuf Aytas 2021-07-27 DESIGNING BIG DATA PLATFORMS Provides expert guidance and valuable insights on getting the most out of Big Data systems An array of tools are currently available for managing and processing data—some are ready-to-go solutions that can be immediately deployed, while others require complex and time-intensive setups. With such a vast range of options, choosing the right tool to build a solution can be complicated, as can determining which tools work well with each other. Designing Big Data Platforms provides clear and authoritative guidance on the critical decisions necessary for successfully deploying, operating, and maintaining Big Data systems. This highly practical guide helps readers understand how to process large amounts of data with well-known Linux tools and database solutions, use effective techniques to collect and manage data from multiple sources, transform data into meaningful business insights, and much more. Author Yusuf Aytas, a software engineer with a vast amount of big data experience, discusses the design of the ideal Big Data platform: one that meets the needs of data analysts, data engineers, data scientists, software engineers, and a spectrum of other stakeholders across an organization. Detailed yet accessible chapters cover key topics such as stream data processing, data analytics, data science, data discovery, and data security. This real-world manual for Big Data technologies: Provides up-to-date coverage of the tools currently used in Big Data processing and management Offers step-by-step guidance on building a data pipeline, from basic scripting to distributed systems Highlights and explains how data is processed at scale Includes an introduction to the foundation of a modern data platform Designing Big Data Platforms: How to Use, Deploy, and Maintain Big Data Systems is a must-have for all professionals working with Big Data, as well researchers and students in computer science and related fields.

Handbook of Research on Big Data Storage and Visualization Techniques - Segall, Richard S. 2018-01-05
The digital age has presented an exponential growth in the amount of data available to individuals looking to draw conclusions based on given or collected information across industries. Challenges associated with the analysis, security, sharing, storage, and visualization of large and complex data sets continue to plague data scientists and analysts alike as traditional data processing applications struggle to adequately manage big data. The Handbook of Research on Big Data Storage and Visualization Techniques is a critical scholarly resource that explores big data analytics and technologies and their role in developing a broad understanding of issues pertaining to the use of big data in multidisciplinary fields. Featuring coverage on a broad range of topics, such as architecture patterns, programing systems, and computational

energy, this publication is geared towards professionals, researchers, and students seeking current research and application topics on the subject.

Deep Learning and Parallel Computing Environment for Bioengineering Systems - Arun Kumar Sangaiah 2019-07-26
Deep Learning and Parallel Computing Environment for Bioengineering Systems delivers a significant forum for the technical advancement of deep learning in parallel computing environment across bio-engineering diversified domains and its applications. Pursuing an interdisciplinary approach, it focuses on methods used to identify and acquire valid, potentially useful knowledge sources. Managing the gathered knowledge and applying it to multiple domains including health care, social networks, mining, recommendation systems, image processing, pattern recognition and predictions using deep learning paradigms is the major strength of this book. This book integrates the core ideas of deep learning and its applications in bio engineering application domains, to be accessible to all scholars and academicians. The proposed techniques and concepts in this book can be extended in future to accommodate changing business organizations' needs as well as practitioners' innovative ideas. Presents novel, in-depth research contributions from a methodological/application perspective in understanding the fusion of deep machine learning paradigms and their capabilities in solving a diverse range of problems Illustrates the state-of-the-art and recent developments in the new theories and applications of deep learning approaches applied to parallel computing environment in bioengineering systems Provides concepts and technologies that are successfully used in the implementation of today's intelligent data-centric critical systems and multi-media Cloud-Big data

The Cloud-Based Demand-Driven Supply Chain - Vinit Sharma 2018-11-08
It's time to get your head in the cloud! In today's business environment, more and more people are requesting cloud-based solutions to help solve their business challenges. So how can you not only anticipate your clients' needs but also keep ahead of the curve to ensure their goals stay on track? With the help of this accessible book, you'll get a clear sense of cloud computing and understand how to communicate the benefits, drawbacks, and options to your clients so they can make the best choices for their unique needs. Plus, case studies give you the opportunity to relate real-life examples of how the latest technologies are giving organizations worldwide the opportunity to thrive as supply chain solutions in the cloud. Demonstrates how improvements in forecasting, collaboration, and inventory optimization can lead to cost savings Explores why cloud computing is becoming increasingly important Takes a close look at the types of cloud computing Makes sense of demand-driven forecasting using Amazon's cloud Whether you work in management, business, or IT, this is the dog-eared reference you'll want to keep close by as you continue making sense of the cloud.

Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning - Nur Zincir-Heywood 2021-09-03
COMMUNICATION NETWORKS AND SERVICE MANAGEMENT IN THE ERA OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING Discover the impact that new technologies are having on communication systems with this up-to-date and one-stop resource Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning delivers a comprehensive overview of the impact of artificial intelligence (AI) and machine learning (ML) on service and network management. Beginning with a fulsome description of ML and AI, the book moves on to discuss management models, architectures, and frameworks. The authors also explore how AI and ML can be used in service management functions like the generation of workload profiles, service provisioning, and more. The book includes a handpicked selection of applications and case studies, as well as a treatment of emerging technologies the authors predict could have a significant impact on network and service management in the future. Statistical analysis and data mining are also discussed, particularly with respect to how they allow for an improvement of the management and security of IT systems and networks. Readers will also enjoy topics like: A thorough

introduction to network and service management, machine learning, and artificial intelligence An exploration of artificial intelligence and machine learning for management models, including autonomic management, policy-based management, intent based management, and network virtualization-based management Discussions of AI and ML for architectures and frameworks, including cloud systems, software defined networks, 5G and 6G networks, and Edge/Fog networks An examination of AI and ML for service management, including the automatic generation of workload profiles using unsupervised learning Perfect for information and communications technology educators, Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning will also earn a place in the libraries of engineers and professionals who seek a structured reference on how the emergence of artificial intelligence and machine learning techniques is affecting service and network management.

Mastering Mesos - Dipa Dubhashi 2016-05-26
The ultimate guide to managing, building, and deploying large-scale clusters with Apache Mesos About This Book Master the architecture of Mesos and intelligently distribute your task across clusters of machines Explore a wide range of tools and platforms that Mesos works with This real-world comprehensive and robust tutorial will help you become an expert Who This Book Is For The book aims to serve DevOps engineers and system administrators who are familiar with the basics of managing a Linux system and its tools What You Will Learn Understand the Mesos architecture Manually spin up a Mesos cluster on a distributed infrastructure Deploy a multi-node Mesos cluster using your favorite DevOps See the nuts and bolts of scheduling, service discovery, failure handling, security, monitoring, and debugging in an enterprise-grade, production cluster deployment Use Mesos to deploy big data frameworks, containerized applications, or even custom build your own applications effortlessly In Detail Apache Mesos is open source cluster management software that provides efficient resource isolations and resource sharing distributed applications or frameworks. This book will take you on a journey to enhance your knowledge from amateur to master level, showing you how to improve the efficiency, management, and development of Mesos clusters. The architecture is quite complex and this book will explore the difficulties and complexities of working with Mesos. We begin by introducing Mesos, explaining its architecture and functionality. Next, we provide a comprehensive overview of Mesos features and advanced topics such as high availability, fault tolerance, scaling, and efficiency. Furthermore, you will learn to set up multi-node Mesos clusters on private and public clouds. We will also introduce several Mesos-based scheduling and management frameworks or applications to enable the easy deployment, discovery, load balancing, and failure handling of long-running services. Next, you will find out how a Mesos cluster can be easily set up and monitored using the standard deployment and configuration management tools. This advanced guide will show you how to deploy important big data processing frameworks such as Hadoop, Spark, and Storm on Mesos and big data storage frameworks such as Cassandra, Elasticsearch, and Kafka. Style and approach This advanced guide provides a detailed step-by-step account of deploying a Mesos cluster. It will demystify the concepts behind Mesos.

Storage Systems - Alexander Thomasian 2021-10-13
Storage Systems: Organization, Performance, Coding, Reliability and Their Data Processing was motivated by the 1988 Redundant Array of Inexpensive/Independent Disks proposal to replace large form factor mainframe disks with an array of commodity disks. Disk loads are balanced by striping data into strips—with one strip per disk— and storage reliability is enhanced via replication or erasure coding, which at best dedicates k strips per stripe to tolerate k disk failures. Flash memories have resulted in a paradigm shift with Solid State Drives (SSDs) replacing Hard Disk Drives (HDDs) for high performance applications. RAID and Flash have resulted in the emergence of new storage companies, namely EMC, NetApp, SanDisk, and Purestorage, and a multibillion-dollar storage market. Key new conferences and publications are reviewed in this book. The goal of the book is to expose students, researchers, and IT professionals to the more important

developments in storage systems, while covering the evolution of storage technologies, traditional and novel databases, and novel sources of data. We describe several prototypes: FAWN at CMU, RAMCloud at Stanford, and Lightstore at MIT; Oracle's Exadata, AWS' Aurora, Alibaba's PolarDB, Fungible Data Center; and author's paper designs for cloud storage, namely heterogeneous disk arrays and hierarchical RAID. • Surveys storage technologies and lists sources of data: measurements, text, audio, images, and video • Familiarizes with paradigms to improve performance: caching, prefetching, log-structured file systems, and merge-trees (LSMs) • Describes RAID organizations and analyzes their performance and reliability • Conserves storage via data compression, deduplication, compaction, and secures data via encryption • Specifies implications of storage technologies on performance and power consumption • Exemplifies database parallelism for big data, analytics, deep learning via multicore CPUs, GPUs, FPGAs, and ASICs, e.g., Google's Tensor Processing Units

Stream Processing with Apache Flink - Fabian Hueske 2019-04-11
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

Expert Apache Cassandra Administration - Sam R. Alapati 2017-12-09
Follow this handbook to build, configure, tune, and secure Apache Cassandra databases. Start with the installation of Cassandra and move on to the creation of a single instance, and then a cluster of Cassandra databases. Cassandra is increasingly a key player in many big data environments, and this book shows you how to use Cassandra with Apache Spark, a popular big data processing framework. Also covered are day-to-day topics of importance such as the backup and recovery of Cassandra databases, using the right compression and compaction strategies, and loading and unloading data. Expert Apache Cassandra Administration provides numerous step-by-step examples starting with the basics of a Cassandra database, and going all the way through backup and recovery, performance optimization, and monitoring and securing the data. The book serves as an authoritative and comprehensive guide to the building and management of simple to complex Cassandra databases. The book: Takes you through building a Cassandra database from installation of the software and creation of a single database, through to complex clusters and data centers Provides numerous examples of actual commands in a real-life Cassandra environment that show how to confidently configure, manage, troubleshoot, and tune Cassandra databases Shows how to use the Cassandra configuration properties to build a highly stable, available, and secure Cassandra database that always operates at peak efficiency What You'll Learn Install the Cassandra software and create your first database Understand the Cassandra data model, and the internal architecture of a Cassandra database Create your own Cassandra cluster, step-by-step Run a Cassandra cluster on Docker Work with Apache Spark by connecting to a Cassandra database Deploy Cassandra clusters in your data center, or on Amazon EC2 instances Back up and restore mission-critical Cassandra databases Monitor, troubleshoot, and tune production Cassandra databases, and cut your spending on

resources such as memory, servers, and storage Who This Book Is For Database administrators, developers, and architects who are looking for an authoritative and comprehensive single volume for all their Cassandra administration needs. Also for administrators who are tasked with setting up and maintaining highly reliable and high-performing Cassandra databases. An excellent choice for big data administrators, database administrators, architects, and developers who use Cassandra as their key data store, to support high volume online transactions, or as a decentralized, elastic data store.

PySpark Recipes - Raju Kumar Mishra 2017-12-09
Quickly find solutions to common programming problems encountered while processing big data. Content is presented in the popular problem-solution format. Look up the programming problem that you want to solve. Read the solution. Apply the solution directly in your own code. Problem solved! PySpark Recipes covers Hadoop and its shortcomings. The architecture of Spark, PySpark, and RDD are presented. You will learn to apply RDD to solve day-to-day big data problems. Python and NumPy are included and make it easy for new learners of PySpark to understand and adopt the model. What You Will Learn Understand the advanced features of PySpark2 and SparkSQL Optimize your code Program SparkSQL with Python Use Spark Streaming and Spark MLlib with Python Perform graph analysis with GraphFrames Who This Book Is For Data analysts, Python programmers, big data enthusiasts
High-Performance Modelling and Simulation for Big Data Applications - Joanna Kołodziej 2019-03-25
This open access book was prepared as a Final Publication of the COST Action IC1406 "High-Performance Modelling and Simulation for Big Data Applications (cHiPSet)" project. Long considered important pillars of the scientific method, Modelling and Simulation have evolved from traditional discrete numerical methods to complex data-intensive continuous analytical optimisations. Resolution, scale, and accuracy have become essential to predict and analyse natural and complex systems in science and engineering. When their level of abstraction raises to have a better discernment of the domain at hand, their representation gets increasingly demanding for computational and data resources. On the other hand, High Performance Computing typically entails the effective use of parallel and distributed processing units coupled with efficient storage, communication and visualisation systems to underpin complex data-intensive applications in distinct scientific and technical domains. It is then arguably required to have a seamless interaction of High Performance Computing with Modelling and Simulation in order to store, compute, analyse, and visualise large data sets in science and engineering. Funded by the European Commission, cHiPSet has provided a dynamic trans-European forum for their members and distinguished guests to openly discuss novel perspectives and topics of interests for these two communities. This cHiPSet compendium presents a set of selected case studies related to healthcare, biological data, computational advertising, multimedia, finance, bioinformatics, and telecommunications.
Ambient Intelligence – Software and Applications –,10th International Symposium on Ambient Intelligence - Paulo Novais 2019-06-22
This book presents the latest research on Ambient Intelligence including software and applications. Ambient Intelligence (AmI) is a paradigm emerging from Artificial Intelligence, in which computers are used as proactive tools for assisting people with their day-to-day activities, making everyone's lives more comfortable. Another main concern of AmI originates from the human–computer interaction domain and focuses on offering ways to interact with systems in a more natural way by means of user-friendly interfaces. This field is evolving rapidly, as can be seen in emerging natural language and gesture-based types of interaction. This symposium was jointly organized by the Universidade do Minho, Technical University of Valencia, Hiroshima University, and University of Salamanca. The latest installment was held in Ávila, Spain, from 26th to 28th June 2019. The authors wish to thank the sponsors: IEEE Systems Man and Cybernetics Society, Spain Section Chapter and the IEEE Spain Section (Technical Co-Sponsor), IBM, Indra, Viewnext, Global Exchange, AEPIA, APPIA and AIR Institute.
Complete Guide to Open Source Big Data Stack - Michael Frampton

2018-01-18

See a Mesos-based big data stack created and the components used. You will use currently available Apache full and incubating systems. The components are introduced by example and you learn how they work together. In the Complete Guide to Open Source Big Data Stack, the author begins by creating a private cloud and then installs and examines Apache Brooklyn. After that, he uses each chapter to introduce one piece of the big data stack—sharing how to source the software and how to install it. You learn by simple example, step by step and chapter by chapter, as a real big data stack is created. The book concentrates on Apache-based systems and shares detailed examples of cloud storage, release management, resource management, processing, queuing, frameworks, data visualization, and more. What You'll Learn Install a private cloud onto the local cluster using Apache cloud stack Source, install, and configure Apache: Brooklyn, Mesos, Kafka, and Zeppelin See how Brooklyn can be used to install Mule ESB on a cluster and Cassandra in the cloud Install and use DCOS for big data processing Use Apache Spark for big data stack data processing Who This Book Is For Developers, architects, IT project managers, database administrators, and others charged with developing or supporting a big data system. It is also for anyone interested in Hadoop or big data, and those experiencing problems with data size.

Data Analytics with Spark Using Python - Jeffrey Aven 2018-06-18

Solve Data Analytics Problems with Spark, PySpark, and Related Open Source Tools Spark is at the heart of today's Big Data revolution, helping data professionals supercharge efficiency and performance in a wide range of data processing and analytics tasks. In this guide, Big Data expert Jeffrey Aven covers all you need to know to leverage Spark, together with its extensions, subprojects, and wider ecosystem. Aven combines a language-agnostic introduction to foundational Spark concepts with extensive programming examples utilizing the popular and intuitive PySpark development environment. This guide's focus on Python makes it widely accessible to large audiences of data professionals, analysts, and developers—even those with little Hadoop or

Spark experience. Aven's broad coverage ranges from basic to advanced Spark programming, and Spark SQL to machine learning. You'll learn how to efficiently manage all forms of data with Spark: streaming, structured, semi-structured, and unstructured. Throughout, concise topic overviews quickly get you up to speed, and extensive hands-on exercises prepare you to solve real problems. Coverage includes: • Understand Spark's evolving role in the Big Data and Hadoop ecosystems • Create Spark clusters using various deployment modes • Control and optimize the operation of Spark clusters and applications • Master Spark Core RDD API programming techniques • Extend, accelerate, and optimize Spark routines with advanced API platform constructs, including shared variables, RDD storage, and partitioning • Efficiently integrate Spark with both SQL and nonrelational data stores • Perform stream processing and messaging with Spark Streaming and Apache Kafka • Implement predictive modeling with SparkR and Spark MLlib

NoSQL - Ganesh Chandra Deka 2017-05-19

This book discusses the advanced databases for the cloud-based application known as NoSQL. It will explore the recent advancements in NoSQL database technology. Chapters on structured, unstructured and hybrid databases will be included to explore bigdata analytics, bigdata storage and processing. The book is likely to cover a wide range of topics such as cloud computing, social computing, bigdata and advanced databases processing techniques.

Event Streams in Action - Valentin Crettaz 2019-05-10

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

Learning Spark - Holden Karau 2015-01-28
This book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning.--

Apache Kafka Quick Start Guide - Raúl Estrada 2018-12-27
Process large volumes of data in real-time while building high performance and robust data stream processing pipeline using the latest Apache Kafka 2.0 Key FeaturesSolve practical large data and processing challenges with KafkaTackle data processing challenges like late events, windowing, and watermarkingUnderstand real-time streaming applications processing using Schema registry, Kafka connect, Kafka streams, and KSQLBook Description Apache Kafka is a great open source

platform for handling your real-time data pipeline to ensure high-speed filtering and pattern matching on the fly. In this book, you will learn how to use Apache Kafka for efficient processing of distributed applications and will get familiar with solving everyday problems in fast data and processing pipelines. This book focuses on programming rather than the configuration management of Kafka clusters or DevOps. It starts off with the installation and setting up the development environment, before quickly moving on to performing fundamental messaging operations such as validation and enrichment. Here you will learn about message composition with pure Kafka API and Kafka Streams. You will look into the transformation of messages in different formats, such asext, binary, XML, JSON, and AVRO. Next, you will learn how to expose the schemas contained in Kafka with the Schema Registry. You will then learn how to work with all relevant connectors with Kafka Connect. While working with Kafka Streams, you will perform various interesting operations on streams, such as windowing, joins, and aggregations. Finally, through KSQL, you will learn how to retrieve, insert, modify, and delete data streams, and how to manipulate watermarks and windows. What you will learnHow to validate data with KafkaAdd information to existing data flowsGenerate new information through message compositionPerform data validation and versioning with the Schema RegistryHow to perform message Serialization and DeserializationHow to perform message Serialization and DeserializationProcess data streams with Kafka StreamsUnderstand the duality between tables and streams with KSQLWho this book is for This book is for developers who want to quickly master the practical concepts behind Apache Kafka. The audience need not have come across Apache Kafka previously; however, a familiarity of Java or any JVM language will be helpful in understanding the code in this book.

Python for Geeks - Muhammad Asif 2021-10-20
Take your Python skills to the next level to develop scalable, real-world applications for local as well as cloud deployment Key FeaturesAll code examples have been tested with Python 3.7 and Python 3.8 and are expected to work with any future 3.x releaseLearn how to build modular

and object-oriented applications in PythonDiscover how to use advanced Python techniques for the cloud and clustersBook Description Python is a multipurpose language that can be used for multiple use cases. Python for Geeks will teach you how to advance in your career with the help of expert tips and tricks. You'll start by exploring the different ways of using Python optimally, both from the design and implementation point of view. Next, you'll understand the life cycle of a large-scale Python project. As you advance, you'll focus on different ways of creating an elegant design by modularizing a Python project and learn best practices and design patterns for using Python. You'll also discover how to scale out Python beyond a single thread and how to implement multiprocessing and multithreading in Python. In addition to this, you'll understand how you can not only use Python to deploy on a single machine but also use clusters in private as well as in public cloud computing environments. You'll then explore data processing techniques, focus on reusable, scalable data pipelines, and learn how to use these advanced techniques for network automation, serverless functions, and machine learning. Finally, you'll focus on strategizing web development design using the techniques and best practices covered in the book. By the end of this Python book, you'll be able to do some serious Python programming for large-scale complex projects. What you will learnUnderstand how to design and manage complex Python projectsStrategize test-driven development (TDD) in PythonExplore multithreading and multiprogramming in PythonUse Python for data processing with Apache Spark and Google Cloud Platform (GCP)Deploy serverless programs on public clouds such as GCPUse Python to build web applications and application programming interfacesApply Python for network automation and serverless functionsGet to grips with Python for data analysis and machine learningWho this book is for This book is for intermediate-level Python developers in any field who are looking to build their skills to develop and manage large-scale complex projects. Developers who want to create reusable modules and Python libraries and cloud developers building applications for cloud deployment will also find this book useful. Prior experience with Python will help you get the

most out of this book.

Production-Ready Microservices - Susan J. Fowler 2016-11-30 One of the biggest challenges for organizations that have adopted microservice architecture is the lack of architectural, operational, and organizational standardization. After splitting a monolithic application or building a microservice ecosystem from scratch, many engineers are left wondering what's next. In this practical book, author Susan Fowler presents a set of microservice standards in depth, drawing from her experience standardizing over a thousand microservices at Uber. You'll learn how to design microservices that are stable, reliable, scalable, fault tolerant, performant, monitored, documented, and prepared for any catastrophe. Explore production-readiness standards, including: Stability and Reliability: develop, deploy, introduce, and deprecate microservices; protect against dependency failures Scalability and Performance: learn essential components for achieving greater microservice efficiency Fault Tolerance and Catastrophe Preparedness: ensure availability by actively pushing microservices to fail in real time Monitoring: learn how to monitor, log, and display key metrics; establish alerting and on-call procedures Documentation and Understanding: mitigate tradeoffs that come with microservice adoption, including organizational sprawl and technical debt

Open Source Software for Statistical Analysis of Big Data: Emerging Research and Opportunities - Segall, Richard S. 2020-02-21
With the development of computing technologies in today's modernized world, software packages have become easily accessible. Open source software, specifically, is a popular method for solving certain issues in the field of computer science. One key challenge is analyzing big data due to the high amounts that organizations are processing. Researchers and professionals need research on the foundations of open source software programs and how they can successfully analyze statistical data. Open Source Software for Statistical Analysis of Big Data: Emerging Research and Opportunities provides emerging research exploring the theoretical and practical aspects of cost-free software

possibilities for applications within data analysis and statistics with a specific focus on R and Python. Featuring coverage on a broad range of topics such as cluster analysis, time series forecasting, and machine learning, this book is ideally designed for researchers, developers, practitioners, engineers, academicians, scholars, and students who want to more fully understand in a brief and concise format the realm and technologies of open source software for big data and how it has been used to solve large-scale research problems in a multitude of disciplines.

Microservices - Eberhard Wolff 2016-10-03

The Most Complete, Practical, and Actionable Guide to Microservices Going beyond mere theory and marketing hype, Eberhard Wolff presents all the knowledge you need to capture the full benefits of this emerging paradigm. He illuminates microservice concepts, architectures, and scenarios from a technology-neutral standpoint, and demonstrates how to implement them with today's leading technologies such as Docker, Java, Spring Boot, the Netflix stack, and Spring Cloud. The author fully explains the benefits and tradeoffs associated with microservices, and guides you through the entire project lifecycle: development, testing, deployment, operations, and more. You'll find best practices for architecting microservice-based systems, individual microservices, and nanoservices, each illuminated with pragmatic examples. The author supplements opinions based on his experience with concise essays from other experts, enriching your understanding and illuminating areas where experts disagree. Readers are challenged to experiment on their own the concepts explained in the book to gain hands-on experience. Discover what microservices are, and how they differ from other forms of modularization Modernize legacy applications and efficiently build new systems Drive more value from continuous delivery with microservices Learn how microservices differ from SOA Optimize the microservices project lifecycle Plan, visualize, manage, and evolve architecture Integrate and communicate among microservices Apply advanced architectural techniques, including CQRS and Event Sourcing Maximize resilience and stability Operate and monitor microservices in production Build a full implementation with Docker, Java, Spring Boot, the Netflix stack, and Spring Cloud Explore nanoservices with Amazon Lambda, OSGi, Java EE, Vert.x, Erlang, and Seneca Understand microservices' impact on teams, technical leaders, product owners, and stakeholders Managers will discover better ways to support microservices, and learn how adopting the method affects the entire organization. Developers will master the technical skills and concepts they need to be effective. Architects will gain a deep understanding of key issues in creating or migrating toward microservices, and exactly what it will take to transform their plans into reality.

Big Data Analytics in Smart Manufacturing - P Suresh 2022-12-14 The significant objective of this edited book is to bridge the gap between smart manufacturing and big data by exploring the challenges and limitations. Companies employ big data technology in the manufacturing field to acquire data about the products. Manufacturing companies could gain a deep business insight by tracking customer details, monitoring fuel consumption, detecting product defects, and supply chain management. Moreover, the convergence of smart manufacturing and big data analytics currently suffers due to data privacy concern, short of qualified personnel, inadequate investment, long-term storage management of high-quality data. The technological advancement makes the data storage more accessible, cheaper and the convergence of these technologies seems to be more promising in the recent era. This book identified the innovative challenges in the industrial domains by integrating heterogeneous data sources such as structured data, semi-structures data, geo-spatial data, textual information, multimedia data, social networking data, etc. It promotes data-driven business modelling processes by adopting big data technologies in the manufacturing industry. Big data analytics is emerging as a promising discipline in the manufacturing industry to build the rigid industrial data platforms. Moreover, big data facilitates process automation in the complete lifecycle of product design and tracking. This book is an essential guide and reference since it synthesizes interdisciplinary theoretical concepts, definitions, and models, involved in smart manufacturing domain. It also provides real-world scenarios and applications, making it accessible to a

wider interdisciplinary audience. Features The readers will get an overview about the smart manufacturing system which enables optimized manufacturing processes and benefits the users by increasing overall profit. The researchers will get insight about how the big data technology leverages in finding new associations, factors and patterns through data stream observations in real time smart manufacturing systems. The industrialist can get an overview about the detection of defects in design, rapid response to market, innovative products to meet the customer requirement which can benefit their per capita income in better way. Discusses technical viewpoints, concepts, theories, and underlying assumptions that are used in smart manufacturing. Information delivered in a user-friendly manner for students, researchers, industrial experts, and business innovators, as well as for professionals and practitioners.

Designing Production-Grade and Large-Scale IoT Solutions - Mohamed Abdelaziz 2022-05-26
Get to grips with key IoT aspects along with modern trends, architectures, and technologies that support IoT solutions, such as cloud computing, modern app architecture paradigms, and data analytics Key Features • Understand the big picture of designing production-grade IoT solutions from an industry expert • Get up and running with the development and designing aspects of the Internet of Things • Solve business problems specific to your domain using different IoT platforms and technologies Book Description With the rising demand for and recent enhancements in IoT, a developer with sound knowledge of IoT is the need of the hour. This book will help you design, build, and operate large-scale E2E IoT solutions to transform your business and products, increase revenue, and reduce operational costs. Starting with an overview of how IoT technologies can help you solve your business problems, this book will be a useful guide to helping you implement end-to-end IoT solution architecture. You'll learn to select IoT devices; real-time operating systems; IoT Edge covering Edge location, software, and hardware; and the best IoT connectivity for your IoT solution. As you progress, you'll work with IoT device management, IoT data analytics,

IoT platforms, and put these components to work as part of your IoT solution. You'll also be able to build IoT backend cloud from scratch by leveraging the modern app architecture paradigms and cloud-native technologies such as containers and microservices. Finally, you'll discover best practices for different operational excellence pillars, including high availability, resiliency, reliability, security, cost optimization, and high performance, which should be applied for large-scale production-grade IoT solutions. By the end of this IoT book, you'll be confident in designing, building, and operating IoT solutions. What you will learn • Understand the detailed anatomy of IoT solutions and explore their building blocks • Explore IoT connectivity options and protocols used in designing IoT solutions • Understand the value of IoT platforms in building IoT solutions • Explore real-time operating systems used in microcontrollers • Automate device administration tasks with IoT device management • Master different architecture paradigms and decisions in IoT solutions • Build and gain insights from IoT analytics solutions • Get an overview of IoT solution operational excellence pillars Who this book is for This book is for E2E solution architects, systems and technical architects, and IoT developers looking to design, build, and operate E2E IoT applications and solutions. Basic knowledge of cloud computing, software engineering, and distributed system design will help you get the most out of this book.

Big Data SMACK - Raul Estrada 2016-09-29
Learn how to integrate full-stack open source big data architecture and to choose the correct technology—Scala/Spark, Mesos, Akka, Cassandra, and Kafka—in every layer. Big data architecture is becoming a requirement for many different enterprises. So far, however, the focus has largely been on collecting, aggregating, and crunching large data sets in a timely manner. In many cases now, organizations need more than one paradigm to perform efficient analyses. Big Data SMACK explains each of the full-stack technologies and, more importantly, how to best integrate them. It provides detailed coverage of the practical benefits of these technologies and incorporates real-world examples in every situation. This book focuses on the problems and scenarios solved

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

Cloud Native Infrastructure with Azure - Nishant Singh 2022-02-09
The cloud is becoming the de facto home for companies ranging from enterprises to startups. Moving to the cloud means moving your applications from monolith to microservices. But once you do, running and maintaining these services brings its own level of complexity. The answer? Modularity, deployability, observability, and self-healing capacity through cloud native development. With this practical book, Nishant Singh and Michael Kehoe show you how to build a true cloud native infrastructure using Microsoft Azure or another cloud computing solution by following guidelines from the Cloud Native Computing Foundation (CNCF). DevOps and site reliability engineers will learn how adapting applications to cloud native early in the design phase helps you fully utilize the elasticity and distributed nature of the cloud. This book helps you explore: Why go cloud native? How to use infrastructure as code What it takes to containerize an application Why and how Kubernetes is the "grand orchestrator" How to create a Kubernetes cluster on Azure How observability complements monitoring How to use service discovery and a service mesh to find new territories How networking and policy management serve as gatekeepers How distributed databases and storage work

AWS Certified Data Analytics Study Guide - Asif Abbasi 2020-12-01
Move your career forward with AWS certification! Prepare for the AWS Certified Data Analytics Specialty Exam with this thorough study guide This comprehensive study guide will help assess your technical skills and prepare for the updated AWS Certified Data Analytics exam. Earning this AWS certification will confirm your expertise in designing and implementing AWS services to derive value from data. The AWS Certified Data Analytics Study Guide: Specialty (DAS-C01) Exam is designed for business analysts and IT professionals who perform complex Big Data analyses. This AWS Specialty Exam guide gets you ready for certification testing with expert content, real-world knowledge, key exam concepts, and topic reviews. Gain confidence by studying the subject areas and working through the practice questions. Big data concepts covered in the guide include: Collection Storage Processing Analysis Visualization Data security AWS certifications allow professionals to demonstrate skills related to leading Amazon Web Services technology. The AWS Certified Data Analytics Specialty (DAS-C01) Exam specifically evaluates your ability to design and maintain Big Data, leverage tools to automate data analysis, and implement AWS Big Data services according to architectural best practices. An exam study guide can help you feel more prepared about taking an AWS certification test and advancing your professional career. In addition to the guide's content, you'll have access to an online learning environment and test bank that offers practice exams, a glossary, and electronic flashcards.

Fast Data Processing Systems with SMACK Stack - Raul Estrada 2016-12-22
Combine the incredible powers of Spark, Mesos, Akka, Cassandra, and Kafka to build data processing platforms that can take on even the hardest of your data troubles! About This Book This highly practical guide shows you how to use the best of the big data technologies to solve your response-critical problems Learn the art of making cheap-yet-effective big data architecture without using complex Greek-letter architectures Use this easy-to-follow guide to build fast data processing systems for your organization Who This Book Is For If you are a

developer, data architect, or a data scientist looking for information on how to integrate the Big Data stack architecture and how to choose the correct technology in every layer, this book is what you are looking for. What You Will Learn Design and implement a fast data Pipeline architecture Think and solve programming challenges in a functional way with Scala Learn to use Akka, the actors model implementation for the JVM Make on memory processing and data analysis with Spark to solve modern business demands Build a powerful and effective cluster infrastructure with Mesos and Docker Manage and consume unstructured and No-SQL data sources with Cassandra Consume and produce messages in a massive way with Kafka In Detail SMACK is an open source full stack for big data architecture. It is a combination of Spark, Mesos, Akka, Cassandra, and Kafka. This stack is the newest technique developers have begun to use to tackle critical real-time analytics for big data. This highly practical guide will teach you how to integrate these technologies to create a highly efficient data analysis system for fast data processing. We'll start off with an introduction to SMACK and show you when to use it. First you'll get to grips with functional thinking and problem solving using Scala. Next you'll come to understand the Akka architecture. Then you'll get to know how to improve the data structure architecture and optimize resources using Apache Spark. Moving forward, you'll learn how to perform linear scalability in databases with Apache Cassandra. You'll grasp the high throughput distributed messaging systems using Apache Kafka. We'll show you how to build a cheap but effective cluster infrastructure with Apache Mesos. Finally, you will deep dive into the different aspect of SMACK using a few case studies. By the end of the book, you will be able to integrate all the components of the SMACK stack and use them together to achieve highly effective and fast data processing. Style and approach With the help of various industry examples, you will learn about the full stack of big data architecture, taking the important aspects in every technology. You will learn how to integrate the technologies to build effective systems rather than getting incomplete information on single technologies. You will learn how various open source technologies can be used to build cheap and fast data processing systems with the help of various industry examples

Advances in Network-Based Information Systems - Leonard Barolli 2018-08-27

This book presents the latest research findings and innovative theoretical and practical research methods and development techniques related to the emerging areas of information networking and their applications. Today's networks and information systems are evolving rapidly, and there are several new trends and applications, such as wireless sensor networks, ad hoc networks, peer-to-peer systems, vehicular networks, opportunistic networks, grid and cloud computing, pervasive and ubiquitous computing, multimedia systems, security, multi-agent systems, high-speed networks, and web-based systems. These networks have to deal with the increasing number of users, provide support for different services, guarantee the QoS, and optimize the network resources, and as such there are numerous research issues and challenges that need to be considered and addressed.

Mastering Hadoop 3 - Chanchal Singh 2019-02-28

A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key FeaturesGet to grips with the newly introduced features and capabilities of Hadoop 3Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystemSharpen your Hadoop skills with real-world case studies and codeBook Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data processing platform, with a focus on improved scalability and increased efficiency. With this guide, you'll understand advanced concepts of the Hadoop ecosystem tool. You'll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You'll be able to address common challenges like using Kafka efficiently, designing low

latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you'll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you'll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you'll be equipped to tackle a range of real-world problems in data pipelines. What you will learnGain an in-depth understanding of distributed computing using Hadoop 3Develop enterprise-grade applications using Apache Spark, Flink, and moreBuild scalable and high-performance Hadoop data pipelines with security, monitoring, and data governanceExplore batch data processing patterns and how to model data in HadoopMaster best practices for enterprises using, or planning to use, Hadoop 3 as a data platformUnderstand security aspects of Hadoop, including authorization and authenticationWho this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

Apache Kafka 1.0 Cookbook - Raúl Estrada 2017-12-22
Simplify real-time data processing by leveraging the power of Apache Kafka 1.0 Key Features Use Kafka 1.0 features such as Confluent platforms and Kafka streams to build efficient streaming data applications to handle and process your data Integrate Kafka with other Big Data tools such as Apache Hadoop, Apache Spark, and more Hands-on recipes to help you design, operate, maintain, and secure your Apache Kafka cluster with ease Book Description Apache Kafka provides a unified, high-throughput, low-latency platform to handle real-time data feeds. This book will show you how to use Kafka efficiently, and contains practical solutions to the common problems that developers and administrators usually face while working with it. This practical guide contains easy-to-follow recipes to help you set up, configure, and use Apache Kafka in the best possible manner. You will use Apache Kafka Consumers and Producers to build effective real-time streaming applications. The book covers the recently released Kafka version 1.0, the Confluent Platform and Kafka Streams. The programming aspect covered in the book will teach you how to perform important tasks such as message validation, enrichment and composition.Recipes focusing on optimizing the performance of your Kafka cluster, and integrate Kafka with a variety of third-party tools such as Apache Hadoop, Apache Spark, and Elasticsearch will help ease your day to day collaboration with Kafka greatly. Finally, we cover tasks related to monitoring and securing your Apache Kafka cluster using tools such as Ganglia and Graphite. If you're looking to become the go-to person in your organization when it comes to working with Apache Kafka, this book is the only resource you need to have. What you will learn -Install and configure Apache Kafka 1.0 to get optimal performance -Create and configure Kafka Producers and Consumers -Operate your Kafka clusters efficiently by implementing the mirroring technique -Work with the new Confluent platform and Kafka streams, and achieve high availability with Kafka -Monitor Kafka using tools such as Graphite and Ganglia -Integrate Kafka with third-party tools such as Elasticsearch, Logstash, Apache Hadoop, Apache Spark, and more Who this book is for This book is for developers and Kafka administrators who are looking for quick, practical solutions to problems encountered while operating, managing or monitoring Apache Kafka. If you are a developer, some knowledge of Scala or Java will help, while for administrators, some working knowledge of Kafka will be useful.

Real-Time Big Data Analytics - Sumit Gupta 2016-02-26
Design, process, and analyze large sets of complex data in real time About This Book Get acquainted with transformations and database-level interactions, and ensure the reliability of messages processed using Storm Implement strategies to solve the challenges of real-time data processing Load datasets, build queries, and make recommendations using Spark SQL Who This Book Is For If you are a Big Data architect, developer, or a programmer who wants to develop

applications/frameworks to implement real-time analytics using open source technologies, then this book is for you. What You Will Learn Explore big data technologies and frameworks Work through practical challenges and use cases of real-time analytics versus batch analytics Develop real-word use cases for processing and analyzing data in real-time using the programming paradigm of Apache Storm Handle and process real-time transactional data Optimize and tune Apache Storm for varied workloads and production deployments Process and stream data with Amazon Kinesis and Elastic MapReduce Perform interactive and exploratory data analytics using Spark SQL Develop common enterprise architectures/applications for real-time and batch analytics In Detail Enterprise has been striving hard to deal with the challenges of data arriving in real time or near real time. Although there are technologies such as Storm and Spark (and many more) that solve the challenges of real-time data, using the appropriate technology/framework for the right business use case is the key to success. This book provides you with the skills required to quickly design, implement and deploy your real-time analytics using real-world examples of big data use cases. From the beginning of the book, we will cover the basics of varied real-time data processing frameworks and technologies. We will discuss and explain the differences between batch and real-time processing in detail, and will also explore the techniques and programming concepts using Apache Storm. Moving on, we'll familiarize you with "Amazon Kinesis" for real-time data processing on cloud. We will further develop your understanding of real-time analytics through a comprehensive review of Apache Spark along with the high-level architecture and the building blocks of a Spark program. You will learn how to transform your data, get an output from transformations, and persist your results using Spark RDDs, using an interface called Spark SQL to work with Spark. At the end of this book, we will introduce Spark Streaming, the streaming library of Spark, and will walk you through the emerging Lambda Architecture (LA), which provides a hybrid platform for big data processing by combining real-time and precomputed batch data to provide a near real-time view of incoming data. Style and approach This

step-by-step is an easy-to-follow, detailed tutorial, filled with practical examples of basic and advanced features. Each topic is explained sequentially and supported by real-world examples and executable code snippets.

Research Anthology on Usage and Development of Open Source Software - Management Association, Information Resources 2021-06-25 The quick growth of computer technology and development of software caused it to be in a constant state of change and advancement. This advancement in software development meant that there would be many types of software developed in order to excel in usability and efficiency. Among these different types of software was open source software, one that grants permission for users to use, study, change, and distribute it freely. Due to its availability, open source software has quickly become a valuable asset to the world of computer technology and across various disciplines including education, business, and library science. The Research Anthology on Usage and Development of Open Source Software presents comprehensive research on the design and development of open source software as well as the ways in which it is used. The text discusses in depth the way in which this computer software has been made into a collaborative effort for the advancement of software technology. Discussing topics such as ISO standards, big data, fault prediction, open collaboration, and software development, this anthology is essential for computer engineers, software developers, IT specialists and consultants, instructors, librarians, managers, executives, professionals, academicians, researchers, and students.
Mastering Spark with R - Javier Luraschi 2019-10-07 If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster

computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

Information Systems - Marinos Themistocleous 2017-08-14
This book constitutes selected papers from the 14th European, Mediterranean, and Middle Eastern Conference, EMCIS 2017, held in Coimbra, Portugal, in September 2017. EMCIS is focusing on approaches that facilitate the identification of innovative research of significant relevance to the IS discipline following sound research methodologies that lead to results of measurable impact. The 37 full and 16 short papers presented in this volume were carefully reviewed and selected from a total of 106 submissions. They are organized in sections on big data and Semantic Web; digital services, social media and digital collaboration; e-government; healthcare information systems; information systems security and information privacy protection; IT governance; and management and organizational issues in information systems.

Mesos in Action - Roger Ignazio 2016-05-02
Summary Mesos in Action introduces readers to the Apache Mesos cluster manager and the concept of application-centric infrastructure. Filled with helpful figures and hands-on instructions, this book guides you from your first steps creating a highly-available Mesos cluster through deploying applications in production and writing native Mesos frameworks. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Modern datacenters are complex environments, and when

you throw Docker and other container-based systems into the mix, there's a great need to simplify. Mesos is an open source cluster management platform that transforms the whole datacenter into a single pool of compute, memory, and storage resources that you can allocate, automate, and scale as if you're working with a single supercomputer. About the Book Mesos in Action introduces readers to the Apache Mesos cluster manager and the concept of application-centric infrastructure. Filled with helpful figures and hands-on instructions, this book guides you from your first steps creating a highly-available Mesos cluster through deploying applications in production and writing native Mesos frameworks. You'll learn how to scale to thousands of nodes, while providing resource isolation between processes using Linux and Docker containers. You'll also learn practical techniques for deploying applications using popular key frameworks. What's Inside Spinning up your first Mesos cluster Scheduling, resource administration, and logging Deploying containerized applications with Marathon, Chronos, and Aurora Writing Mesos frameworks using Python About the Reader Readers need to be familiar with the core ideas of datacenter administration and need a basic knowledge of Python or a similar programming language. About the Author Roger Ignazio is an experienced systems engineer with a focus on distributed, fault-tolerant, and scalable infrastructure. He is currently a technical lead at Mesosphere. Table of Contents PART 1 HELLO, MESOS Introducing Mesos Managing datacenter resources with Mesos PART 2 CORE MESOS Setting up Mesos Mesos fundamentals Logging and debugging Mesos in production PART 3 RUNNING ON MESOS Deploying applications with MarathoN Managing scheduled tasks with Chronos Deploying applications and managing scheduled tasks with Aurora Developing a framework

PySpark SQL Recipes - Raju Kumar Mishra 2019-03-18
Carry out data analysis with PySpark SQL, graphframes, and graph data processing using a problem-solution approach. This book provides solutions to problems related to dataframes, data manipulation summarization, and exploratory analysis. You will improve your skills in

graph data analysis using graphframes and see how to optimize your PySpark SQL code. PySpark SQL Recipes starts with recipes on creating dataframes from different types of data source, data aggregation and summarization, and exploratory data analysis using PySpark SQL. You'll also discover how to solve problems in graph analysis using graphframes. On completing this book, you'll have ready-made code for all your PySpark SQL tasks, including creating dataframes using data from different file formats as well as from SQL or NoSQL databases. What You Will Learn Understand PySpark SQL and its advanced features Use SQL and HiveQL with PySpark SQL Work with structured streaming Optimize PySpark SQL Master graphframes and graph processing Who This Book Is ForData scientists, Python programmers, and SQL programmers.

Learning Spark - Jules S. Damji 2020-07-16
Data is bigger, arrives faster, and comes in a variety of formatsâ??and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, youâ??ll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Distributed Computing and Artificial Intelligence, 14th International Conference - Sigeru Omatu 2017-06-19
The 14th International Symposium on Distributed Computing and Artificial Intelligence 2017 (DCAI 2017) provided a forum for presenting the application of innovative techniques to study and solve complex problems. The exchange of ideas between scientists and technicians from both the academic and industrial sector is essential to advancing the development of systems that can meet the ever-growing demands of today's society. The book brings together past experience, current work and promising future trends in distributed computing, artificial intelligence and their applications to efficiently solve real-world problems. It combines contributions in well-established and evolving areas of research, including the content of the DCAI 17 Special Sessions, which focused on multi-disciplinary and transversal aspects, such as AI-driven methods for multimodal networks and processes modeling, and secure management towards smart buildings and smart grids. The symposium was jointly organized by the Polytechnic of Porto, the Osaka Institute of Technology and the University of Salamanca. The latest event was held in Porto, Portugal, from 21st to 23rd June 2017.

AWS Certified Data Analytics Study Guide with Online Labs - Asif Abbasi 2021-04-13
Virtual, hands-on learning labs allow you to apply your technical skills in realistic environments. So Sybex has bundled AWS labs from XtremeLabs with our popular AWS Certified Data Analytics Study Guide to give you the same experience working in these labs as you prepare for the Certified Data Analytics Exam that you would face in a real-life application. These labs in addition to the book are a proven way to prepare for the certification and for work as an AWS Data Analyst. AWS Certified Data Analytics Study Guide: Specialty (DAS-C01) Exam is intended for individuals who perform in a data analytics-focused role. This UPDATED exam validates an examinee's comprehensive understanding of using AWS services to design, build, secure, and maintain analytics solutions that provide insight from data. It assesses an examinee's ability to define AWS data analytics services and understand how they integrate with each other; and explain how AWS data analytics services fit in the data lifecycle of collection, storage, processing, and visualization. The book focuses on the following domains: • Collection • Storage and Data Management • Processing • Analysis and Visualization • Data Security This is your opportunity to take the next step in your

career by expanding and validating your skills on the AWS cloud. AWS is the frontrunner in cloud computing products and services, and the AWS Certified Data Analytics Study Guide: Specialty exam will get you fully prepared through expert content, and real-world knowledge, key exam essentials, chapter review questions, and much more. Written by an AWS subject-matter expert, this study guide covers exam concepts, and provides key review on exam topics. Readers will also have access to Sybex's superior online interactive learning environment and test bank, including chapter tests, practice exams, a glossary of key terms, and electronic flashcards. And included with this version of the book, XtremeLabs virtual labs that run from your browser. The registration code is included with the book and gives you 6 months of unlimited access to XtremeLabs AWS Certified Data Analytics Labs with 3 unique lab modules based on the book.