Big Data Smack A Guide To Apache Spark Mesos Akka Cassandra And Kafka

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
FileMaker Pro 9: The Missing Manual is the clear, thorough and accessible guide to the latest version of this popular desktop database program. FileMaker Pro lets you do almost anything with the information you give it. You can print corporate reports, plan your retirement, or run a small country -- if you know what you're doing. This book helps non-technical folks like you get in, get your database built, and get the results you need. Pronto. The new edition gives novices and experienced users the scoop on versions 8.5 and 9. It offers complete coverage of timesaving new features such as the Quick Start screen that lets you open or a create a database in a snap, the handy "save to" buttons for making Excel documents or PDFs, the multiple level Undo and Redo commands let you step backwards through your typing tasks, and much more. With FileMaker Pro 9: The Missing Manual, you can: Get your first database running in minutes and perform basic tasks right away. Catalog people, processes and things with streamlined data entry and sorting tools. Learn to use layout tools to organize the appearance of your database. Use your data to generate reports, correspondence and other documents with ease. Create, connect, and manage multiple tables and set up complex relationships that show you just the data you need. Crunch numbers, search text, or pin down dates and times with dozens of built-in formulas. Automate repetitive tasks with FileMaker Pro 9's easy-to-learn scripting language. Protect your database with passwords and set up privileges to determine what others can do once they gain entry. Outfit your database for the Web and import and export data to other formats. Each chapter in the book contains "living examples" -- downloadable tutorials that help you learn how to build a database by actually doing it. You also get plenty of sound, objective advice that lets you know which features are really useful, and which ones you'll barely touch. To make the most of FileMaker Pro 9, you need the book that should have been in the box.
This book constitutes the revised selected papers of the scientific satellite events that were held in conjunction with the 16th International Conference on Service-Oriented Computing, ICSOC 2018, held in Hangzhou, China, in November 2018. The ICSOC 2018 workshop track consisted of six workshops on a wide range of topics that fall into the general area of service

computing. A special focus this year was on Internet of Things, Data Analytics, and Smart Services: First International Workshop on Data-Driven Business Services (DDBS)First International Workshop on Networked Learning Systems for Secured IoT Services and Its Applications (NLS4IoT)8th International Workshop on Context-Aware and IoT Services (CIoTS)Third International Workshop on Adaptive Service-oriented and Cloud Applications (ASOCA2018)Third International Workshop on IoT Systems for Context-aware Computing (ISyCC)First International Workshop on AI and Data Mining for Services (ADMS) 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 This proceedings book gathers papers presented at the 4th International Conference on Advanced Engineering Theory and Applications 2017 (AETA 2017), held on 7–9 December 2017 at Ton Duc Thang University, Ho Chi Minh City, Vietnam. It presents selected papers on 13 topical areas, including robotics, control systems, telecommunications, computer science and more. All selected papers represent interesting ideas and collectively provide a state-of-the-art overview. Readers will find intriguing papers on the design and implementation of control algorithms for aerial and underwater robots, for mechanical systems, efficient protocols for vehicular ad hoc networks, motor control, image and signal processing, energy saving, optimization methods in various fields of electrical engineering, and others. The book also offers a valuable resource for practitioners who want to apply the content discussed to solve

real-life problems in their challenging applications. It also addresses common and related subjects in modern electric, electronic and related technologies. As such, it will benefit all scientists and engineers working in the above-mentioned fields of application.

Rachel Friedman has always been the consummate good girl who does well in school and plays it safe, so the college grad surprises no one more than herself when, on a whim (and in an effort to escape impending life decisions), she buys a ticket to Ireland, a place she has never visited. There she forms an unlikely bond with a free-spirited Australian girl, a born adventurer who spurs Rachel on to a yearlong odyssey that takes her to three continents, fills her life with newfound friends, and gives birth to a previously unrealized passion for adventure. As her journey takes her to Australia and South America, Rachel discovers and embraces her love of travel and unlocks more truths about herself than she ever realized she was seeking. Along the way, the erstwhile good girl finally learns to do something she's never done before: simply live for the moment.

The premier guide to digital marketing that works, and a solid framework for success The Art of Digital Marketing is the comprehensive guide to cracking the digital marketing 'code,' and reaching, engaging, and serving the empowered consumer. Based on the industry's leading certification from the Digital Marketing Institute (DMI), this book presents an innovative methodology for successful digital marketing: start with the customer and work backwards. A campaign is only effective as it is reflective of the consumer's wants, needs, preferences, and inclinations; the DMI framework provides structured, implementable, iterative direction for getting it right every time. The heart of the framework is a three-step process called the 3i Principles: Initiate, Iterate, and Integrate. This simple idea translates into higher engagement, real customer interaction, and multichannel campaigns that extend even into traditional marketing channels. The evolution of digital marketing isn't really about the brands; it's about consumers exercising more control over their choices. This book demonstrates how using this single realization as a starting point helps you build and implement more effective campaigns. Get inside the customer's head with deep consumer research Constantly improve your campaigns based on feedback and interactions Integrate digital activities across channels, including traditional marketing Build campaigns based on customer choice and control Digital marketing turns traditional marketing models on their heads. Instead of telling the customer what to think, you find out what they already think and go from there. Instead of front-loading resources, you continually adjust your approach based on real interactions with real customers every day. Digital marketing operates within its own paradigm, and The Art of Digital Marketing opens the door for your next campaign.

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

Living in a "perfect" world without social ills, a boy approaches the time when he will receive a life assignment from the Elders, but his selection leads him to a mysterious man known as the

Giver, who reveals the dark secrets behind the utopian facade.

Joe Celko's Complete Guide to NoSQL provides a complete overview of non-relational technologies so that you can become more nimble to meet the needs of your organization. As data continues to explode and grow more complex, SQL is becoming less useful for querying data and extracting meaning. In this new world of bigger and faster data, you will need to leverage non-relational technologies to get the most out of the information you have. Learn where, when, and why the benefits of NoSQL outweigh those of SQL with Joe Celko's Complete Guide to NoSQL. This book covers three areas that make today's new data different from the data of the past: velocity, volume and variety. When information is changing faster than you can collect and query it, it simply cannot be treated the same as static data. Celko will help you understand velocity, to equip you with the tools to drink from a fire hose. Old storage and access models do not work for big data. Celko will help you understand volume, as well as different ways to store and access data such as petabytes and exabytes. Not all data can fit into a relational model, including genetic data, semantic data, and data generated by social networks. Celko will help you understand variety, as well as the alternative storage, query, and management frameworks needed by certain kinds of data. Gain a complete understanding of the situations in which SQL has more drawbacks than benefits so that you can better determine when to utilize NoSQL technologies for maximum benefit Recognize the pros and cons of columnar, streaming, and graph databases Make the transition to NoSQL with the expert guidance of best-selling SQL expert Joe Celko

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

Learn the right cutting-edge skills and knowledge to leverage Spark Streaming to implement a wide array of real-time, streaming applications. This book walks you through end-to-end real-

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

When you're running for local office for the first time, there are plenty of mistakes you can make. The good thing is that you're not the first person ever to face the challenges of running for office - even though sometimes it might seem like it. In this book, Craig W. Turner collects stories, advice, warnings and best practices from local candidates from around the country who have been there and done that. They've already seen what you're about to see, and they've generously shared their experiences for the benefit of you and other first-time candidates for local office. This is no poli-sci class. This is "in the trenches" life experience. Everyone interviewed for this book as part of The Campaign Coach Podcast, has run and won at least one election. Some are now retired, some have moved on to higher office, and some are still serving in the capacity we discussed in their interview. In these pages, you will find insights that no guidebook or online course can give you - they will inspire you, and scare you, and motivate you and smack you in the face with a reality check. But, they will also make you a stronger candidate. Whether it's talking with voters, fundraising, understanding local government policy, promoting your candidacy or even knowing how to best serve your community, learning from these people's experiences - and avoiding mistakes that they've already made for you - will give you a leg up on any opponent.

This book focuses on the core areas of computing and their applications in the real world. Presenting papers from the Computing Conference 2020 covers a diverse range of research areas, describing various detailed techniques that have been developed and implemented. The Computing Conference 2020, which provided a venue for academic and industry practitioners to share new ideas and development experiences, attracted a total of 514 submissions from pioneering academic researchers, scientists, industrial engineers and students from around the globe. Following a double-blind, peer-review process, 160 papers (including 15 poster papers) were selected to be included in these proceedings. Featuring state-of-the-art intelligent methods and techniques for solving real-world problems, the book is a valuable resource and will inspire further research and technological improvements in this important area.

The Regulatory Technology Handbook The transformational potential of RegTech has been confirmed in recent years with US$1.2 billion invested in start-ups (2017) and an expected additional spending of US$100 billion by 2020. Regulatory technology will not only provide

efficiency gains for compliance and reporting functions, it will radically change market structure and supervision. This book, the first of its kind, is providing a comprehensive and invaluable source of information aimed at corporates, regulators, compliance professionals, start-ups and policy makers. The REGTECH Book brings into a single volume the curated industry expertise delivered by subject matter experts. It serves as a single reference point to understand the RegTech eco-system and its impact on the industry. Readers will learn foundational notions such as: • The economic impact of digitization and datafication of regulation • How new technologies (Artificial Intelligence, Blockchain) are applied to compliance • Business use cases of RegTech for cost-reduction and new product origination • The future regulatory landscape affecting financial institutions, technology companies and other industries Edited by world-class academics and written by compliance professionals, regulators, entrepreneurs and business leaders, the RegTech Book represents an invaluable resource that paves the way for 21st century regulatory innovation.

A practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario About This Book Learn about the various challenges in real-time data processing and use the right tools to overcome them This book covers popular tools and frameworks such as Spark, Flink, and Apache Storm to solve all your distributed processing problems A practical guide filled with examples, tips, and tricks to help you perform efficient Big Data processing in real-time Who This Book Is For If you are a Java developer who would like to be equipped with all the tools required to devise an end-to-end practical solution on real-time data streaming, then this book is for you. Basic knowledge of real-time processing would be helpful, and knowing the fundamentals of Maven, Shell, and Eclipse would be great. What You Will Learn Get an introduction to the established real-time stack Understand the key integration of all the components Get a thorough understanding of the basic building blocks for real-time solution designing Garnish the search and visualization aspects for your real-time solution Get conceptually and practically acquainted with real-time analytics Be well equipped to apply the knowledge and create your own solutions In Detail With the rise of Big Data, there is an increasing need to process large amounts of data continuously, with a shorter turnaround time. Real-time data processing involves continuous input, processing and output of data, with the condition that the time required for processing is as short as possible. This book covers the majority of the existing and evolving open source technology stack for real-time processing and analytics. You will get to know about all the real-time solution aspects, from the source to the presentation to persistence. Through this practical book, you'll be equipped with a clear understanding of how to solve challenges on your own. We'll cover topics such as how to set up components, basic executions, integrations, advanced use cases, alerts, and monitoring. You'll be exposed to the popular tools used in real-time processing today such as Apache Spark, Apache Flink, and Storm. Finally, you will put your knowledge to practical use by implementing all of the techniques in the form of a practical, real-world use case. By the end of this book, you will have a solid understanding of all the aspects of real-time data processing and analytics, and will know how to deploy the solutions in production environments in the best possible manner. Style and Approach In this practical guide to real-time analytics, each chapter begins with a basic high-level concept of the topic, followed by a practical, hands-on implementation of each concept, where you can see the working and execution of it. The book is written in a DIY style, with plenty of practical use cases, well-explained code examples, and relevant screenshots and diagrams.

An imaginative story of a woman caught in an alternate world—where she will need to learn the skills of magic to survive Nora Fischer's dissertation is stalled and her boyfriend is about to marry another woman. During a miserable weekend at a friend's wedding, Nora wanders off and walks through a portal into a different world where she's transformed from a drab grad student into a stunning beauty. Before long, she

has a set of glamorous new friends and her romance with gorgeous, masterful Raclin is heating up. It's almost too good to be true. Then the elegant veneer shatters. Nora's new fantasy world turns darker, a fairy tale gone incredibly wrong. Making it here will take skills Nora never learned in graduate school. Her only real ally—and a reluctant one at that—is the magician Aruendiel, a grim, reclusive figure with a biting tongue and a shrouded past. And it will take her becoming Aruendiel's student—and learning magic herself—to survive. When a passage home finally opens, Nora must weigh her "real life" against the dangerous power of love and magic. For lovers of Lev Grossman's The Magicians series (The Magicians and The Magician King) and Deborah Harkness's All Souls Trilogy (A Discovery of Witches and Shadow of Night).

Digital data collection and surveillance is pervasive and no one can protect your privacy without your help. Before you can help yourself, you need to understand the new technologies, what benefits they provide, and what trade-offs they require. Some of those trade-offs – privacy for convenience – could be softened by our own behavior or be reduced by legislation if we fight for it. This book analyzes why privacy is important to all of us, and it describes the technologies that place your privacy most at risk, starting with modern computing and the Internet.

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.

Longlisted for the National Book Award New York Times Bestseller A former Wall Street quant sounds an alarm on the mathematical models that pervade modern life -- and threaten to rip apart our social fabric We live in the age of the algorithm. Increasingly, the decisions that affect our lives--where we go to school, whether we get a car loan, how much we pay for health insurance--are being made not by humans, but by mathematical models. In theory, this should lead to greater fairness: Everyone is judged according to the same rules, and bias is eliminated. But as Cathy O'Neil reveals in this urgent and necessary book, the opposite is true. The models being used today are opaque, unregulated, and uncontestable, even when they're wrong. Most troubling, they reinforce discrimination: If a poor student can't get a loan because a lending model deems him too risky (by virtue of his zip code), he's then cut off from the kind of education that could pull him out of poverty, and a vicious spiral ensues. Models are propping up the lucky and punishing the downtrodden, creating a "toxic cocktail for democracy." Welcome to the dark side of Big Data. Tracing the arc of a person's life, O'Neil exposes the black box models that shape our future, both as individuals and as a society. These "weapons of math destruction" score teachers and students, sort r sum s, grant (or deny) loans, evaluate workers, target voters, set parole, and monitor our

health. O'Neil calls on modelers to take more responsibility for their algorithms and on policy makers to regulate their use. But in the end, it's up to us to become more savvy about the models that govern our lives. This important book empowers us to ask the tough questions, uncover the truth, and demand change. -- Longlist for National Book Award (Non-Fiction) -- Goodreads, semi-finalist for the 2016 Goodreads Choice Awards (Science and Technology) -- Kirkus, Best Books of 2016 -- New York Times, 100 Notable Books of 2016 (Non-Fiction) -- The Guardian, Best Books of 2016 -- WBUR's "On Point," Best Books of 2016: Staff Picks -- Boston Globe, Best Books of 2016, Non-Fiction

#1 NEW YORK TIMES BESTSELLER • ONE OF TIME MAGAZINE'S 100 BEST YA BOOKS OF ALL TIME The extraordinary, beloved novel about the ability of books to feed the soul even in the darkest of times. When Death has a story to tell, you listen. It is 1939. Nazi Germany. The country is holding its breath. Death has never been busier, and will become busier still. Liesel Meminger is a foster girl living outside of Munich, who scratches out a meager existence for herself by stealing when she encounters something she can't resist–books. With the help of her accordion-playing foster father, she learns to read and shares her stolen books with her neighbors during bombing raids as well as with the Jewish man hidden in her basement. In superbly crafted writing that burns with intensity, award-winning author Markus Zusak, author of I Am the Messenger, has given us one of the most enduring stories of our time. "The kind of book that can be life-changing." —The New York Times "Deserves a place on the same shelf with The Diary of a Young Girl by Anne Frank." —USA Today DON'T MISS BRIDGE OF CLAY, MARKUS ZUSAK'S FIRST NOVEL SINCE THE BOOK THIEF. 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 organizationWho This Book Is ForIf 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 KafkaIn DetailSMACK 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 approachWith 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

White Space Is Not Your Enemy is a practical graphic design and layout guide that introduces concepts and practices necessary for producing effective visual communication across a variety of formats—from web to print. Sections on Gestalt theory, color theory, and WET layout are expanded to offer more in-depth content on those topics. This new edition features new covering current trends in web design—Mobile-first, UI/UX design, and web typography—and how they affect a designer's approach to a project. The entire book will receive an update using new examples and images that show a more diverse set of graphics that go beyond print and web and focus on tablet, mobile and advertising designs.

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 Big Data SMACKA Guide to Apache Spark, Mesos, Akka, Cassandra, and KafkaApress This book highlights the different types of data architecture and illustrates the many possibilities hidden behind the term "Big Data", from the usage of No-SQL databases to the deployment of stream analytics architecture, machine learning, and governance. Scalable Big Data Architecture covers real-world, concrete industry use cases that leverage complex distributed applications , which involve web applications, RESTful API, and high throughput of large amount of data stored in highly scalable No-SQL data stores such as Couchbase and Elasticsearch. This book demonstrates how data processing can be done at scale from the usage of NoSQL datastores to the combination of Big Data distribution. When the data processing is too complex and involves different processing topology like long running jobs, stream processing, multiple data sources correlation, and machine learning, it's often necessary to delegate the load to Hadoop or Spark and use the No-SQL to serve processed data in real time. This book shows you how to choose a relevant combination of big data technologies available within the Hadoop ecosystem. It focuses on processing long jobs, architecture, stream data patterns, log analysis, and real time analytics. Every pattern is illustrated with practical examples, which use the different open sourceprojects such as Logstash, Spark, Kafka, and so on. Traditional data infrastructures are built for digesting and rendering data synthesis and analytics from large amount of data. This book helps you to understand why you should consider using machine learning algorithms early on in the project, before being overwhelmed by constraints imposed by dealing with the high throughput of Big data. Scalable Big Data Architecture is for developers, data architects, and data scientists looking for a better understanding of how to choose the most relevant pattern for a Big Data project and which tools to integrate into that pattern.

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. An exploration as well as a literary celebration of the fascinating young adult fantasy series, this companion guide takes readers deep into the rich universe of Cassandra

Clare's New York Times–bestselling Shadowhunter Chronicles franchise. With intelligent yet accessible dissections of each volume of both the Mortal Instruments series and the Infernal Devices series, Liv Spencer delivers the next best thing to a Shadowhunter's codex with commentary on the books as well as their references to folklore, legends, and literature. The guide also recounts Cassandra Clare's publishing story, from journalist and fan fiction writer to bestselling author; explores the cast and crew who brought the first book to life in the film The Mortal Instruments: City of Bones, premiering in August 2013; and delves into the franchise's fans, a passionate community that is anything but mundane. From the Clave to Chairman Meow and demon pox to dastardly ducks, Navigating the Shadow World is both an insightful introduction to the world of Cassandra Clare and a satisfying companion book for fans. An introduction to the Droid X explains how to get the most out of the device, with a hands-on approach to learning the Droid X functions and applications, a review of its features, customization tips and tricks, and instructions to help users master theDroid X.

The book Intelligent Systems and Applications - Proceedings of the 2020 Intelligent Systems Conference is a remarkable collection of chapters covering a wider range of topics in areas of intelligent systems and artificial intelligence and their applications to the real world. The Conference attracted a total of 545 submissions from many academic pioneering researchers, scientists, industrial engineers, students from all around the world. These submissions underwent a double-blind peer review process. Of those 545 submissions, 177 submissions have been selected to be included in these proceedings. As intelligent systems continue to replace and sometimes outperform human intelligence in decision-making processes, they have enabled a larger number of problems to be tackled more effectively.This branching out of computational intelligence in several directions and use of intelligent systems in everyday applications have created the need for such an international conference which serves as a venue to report on up-to-the-minute innovations and developments. This book collects both theory and application based chapters on all aspects of artificial intelligence, from classical to intelligent scope. We hope that readers find the volume interesting and valuable; it provides the state of the art intelligent methods and techniques for solving real world problems along with a vision of the future research.

With the immense amount of data that is now available online, security concerns have been an issue from the start, and have grown as new technologies are increasingly integrated in data collection, storage, and transmission. Online cyber threats, cyber terrorism, hacking, and other cybercrimes have begun to take advantage of this information that can be easily accessed if not properly handled. New privacy and security measures have been developed to address this cause for concern and have become an essential area of research within the past few years and into the foreseeable future. The ways in which data is secured and privatized should be discussed in terms of the technologies being used, the methods and models for security that have been developed, and the ways in which risks can be detected, analyzed, and mitigated. The Research Anthology on Privatizing and Securing Data reveals the latest tools and technologies for privatizing and securing data across different technologies and industries. It takes a deeper dive into both risk detection and mitigation, including an analysis of cybercrimes and cyber threats, along with a sharper focus on the

technologies and methods being actively implemented and utilized to secure data online. Highlighted topics include information governance and privacy, cybersecurity, data protection, challenges in big data, security threats, and more. This book is essential for data analysts, cybersecurity professionals, data scientists, security analysts, IT specialists, practitioners, researchers, academicians, and students interested in the latest trends and technologies for privatizing and securing data. Explores the alphabet and animals through their collective nouns, including a covey of quails, a string of ponies, and a murder of crows.

A comprehensive guide for seasoned business leaders who struggle with where and how to exploit the economics of data and analytics to gain true value from data, accelerate company operations through AI, and guide their digital transformation. Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLib, and R on Spark with the help of practical code snippets for each topic. Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You'll follow a learn-to-do-by-yourself approach to learning – learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you'll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You'll also become familiar with machine learning algorithms with real-time usage. What You Will Learn Discover the functional programming features of Scala Understand the complete architecture of Spark and its components Integrate Apache Spark with Hive and Kafka Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries Work with different machine learning concepts and libraries using Spark's MLlib packages Who This Book Is For Developers and professionals who deal with batch and stream data processing.

Big Data Analytics with Spark is a step-by-step guide for learning Spark, which is an open-source fast and general-purpose cluster computing framework for large-scale data analysis. You will learn how to use Spark for different types of big data analytics projects, including batch, interactive, graph, and stream data analysis as well as machine learning. In addition, this book will help you become a much sought-after Spark expert. Spark is one of the hottest Big Data technologies. The amount of data generated today by devices, applications and users is exploding. Therefore, there is a critical need for tools that can analyze large-scale data and unlock value from it. Spark is a powerful technology that meets that need. You can, for example, use Spark to perform low latency computations through the use of efficient caching and iterative algorithms; leverage the features of its shell for easy and interactive Data analysis; employ its fast batch processing and low latency features to process your real time data streams and so on. As a result, adoption of Spark is rapidly growing and is replacing Hadoop MapReduce as the technology of choice for big data analytics. This book provides an introduction to Spark and related big-data technologies. It covers Spark core and its add-on libraries, including Spark SQL, Spark Streaming, GraphX, and MLib. Big Data Analytics with Spark is therefore written for busy professionals who prefer learning a new technology from a consolidated source instead of spending countless hours on the Internet trying to pick bits and pieces from different sources. The

book also provides a chapter on Scala, the hottest functional programming language, and the program that underlies Spark. You'll learn the basics of functional programming in Scala, so that you can write Spark applications in it. What's more, Big Data Analytics with Spark provides an introduction to other big data technologies that are commonly used along with Spark, like Hive, Avro, Kafka and so on. So the book is self-sufficient; all the technologies that you need to know to use Spark are covered. The only thing that you are expected to know is programming in any language. There is a critical shortage of people with big data expertise, so companies are willing to pay top dollar for people with skills in areas like Spark and Scala. So reading this book and absorbing its principles will provide a boost—possibly a big boost—to your career. This book is for all people who are forced to use UNIX. It is a humorous book--pure entertainment--that maintains that UNIX is a computer virus with a user interface. It features letters from the thousands posted on the Internet's "UNIX-Haters" mailing list. It is not a computer handbook, tutorial, or reference. It is a self-help book that will let readers know they are not alone. When conducting research in developing countries, an ability to negotiate a bewildering array of cultural and logistical obstacles is essential. Overseas Research: A Practical Guide distills essential lessons from scores of students and scholars who have collected data and done fieldwork abroad, including how to prepare for the field, how and where to find funding for one's fieldwork, issues of personal safety and security, and myriad logistical and relational issues. By encouraging researchers to think through the challenges of research before they begin it, Overseas Research will help prepare fieldworkers for the practical, logistical, and psychological considerations of very demanding work, help save valuable time, make the most of scarce financial resources, and enhance the quality of the field research. This third edition contains new material on social media, including representation of research subjects/collaborators, students' digital branding and image, and representing universities abroad when posting publicly. It also covers emerging technologies such as solar panels for power in remote locations, new ways of digitally sending and receiving money, and incorporates more perspectives of women, LGBTQ+ people, and people of color researching abroad. The book will be of interest to overseas fieldworkers, and also to undergraduates in subjects such as anthropology, economics, geography, history, international studies, politics, sociology, and development studies. Utilize this practical and easy-to-follow guide to modernize traditional enterprise data warehouse and business intelligence environments with next-generation big data technologies. Next-Generation Big Data takes a holistic approach, covering the most important aspects of modern enterprise big data. The book covers not only the main technology stack but also the next-generation tools and applications used for big data warehousing, data warehouse optimization, real-time and batch data ingestion and processing, real-time data visualization, big data governance, data wrangling, big data cloud deployments, and distributed in-memory big data computing. Finally, the book has an extensive and detailed coverage of big data case studies from Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard. What You'll Learn Install Apache Kudu, Impala, and Spark to modernize enterprise data warehouse and business intelligence environments, complete with real-world, easy-to-follow examples, and practical advice Integrate HBase, Solr, Oracle, SQL Server, MySQL, Flume, Kafka,

HDFS, and Amazon S3 with Apache Kudu, Impala, and Spark Use StreamSets, Talend, Pentaho, and CDAP for real-time and batch data ingestion and processing Utilize Trifacta, Alteryx, and Datameer for data wrangling and interactive data processing Turbocharge Spark with Alluxio, a distributed in-memory storage platform Deploy big data in the cloud using Cloudera Director Perform real-time data visualization and time series analysis using Zoomdata, Apache Kudu, Impala, and Spark Understand enterprise big data topics such as big data governance, metadata management, data lineage, impact analysis, and policy enforcement, and how to use Cloudera Navigator to perform common data governance tasks Implement big data use cases such as big data warehousing, data warehouse optimization, Internet of Things, real-time data ingestion and analytics, complex event processing, and scalable predictive modeling Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard Who This Book Is For BI and big data warehouse professionals interested in gaining practical and real-world insight into next-generation big data processing and analytics using Apache Kudu, Impala, and Spark; and those who want to learn more about other advanced enterprise topics

Why use qualitative methods? What kinds of questions can qualitative methods help you answer? How do you actually do rigorous and reflective qualitative research in the real world? Written by a team of leading researchers associated with NatCen Social Research (the National Centre for Social Research) this textbook leads students and researchers through the entire process of qualitative research from beginning to end - moving through design, sampling, data collection, analysis and reporting. In this fully revised second edition you will find: A practical account of how to carry out qualitative research which recognises a range of current approaches and applications A brand new chapter on ethics A brand new chapter on observational research Updated advice on using software when analysing your qualitative data New case studies which illustrate issues you may encounter and how problems have been tackled by other researchers. This book is an ideal guide for students, practitioners and researchers faced with the challenges of doing qualitative research in both applied and academic settings in messy real-life contexts.

This book is about how to integrate full-stack open source big data architecture and how 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 datasets 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. The 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'll learn How to make big data architecture without using complex Greek letter architectures. How to build a

cheap but effective cluster infrastructure. How to make queries, reports, and graphs that business demands. How to manage and exploit unstructured and No-SQL data sources. How use tools to monitor the performance of your architecture. How to integrate all technologies and decide which replace and which reinforce. Who This Book Is For This book is for developers, data architects, and data scientists looking for how to integrate the most successful big data open stack architecture and how to choose the correct technology in every layer.

Copyright: db1a0b6b340cc5b92b20596ce848e999