Read PDF Pro Spark Streaming The Zen Of Real Time Analytics Using Apache Spark

Getting the books **Pro Spark Streaming The Zen Of Real Time Analytics Using Apache Spark** now is not type of challenging means. You could not forlorn going gone book accrual or library or borrowing from your connections to entry them. This is an extremely simple means to specifically acquire lead by on-line. This online revelation Pro Spark Streaming The Zen Of Real Time Analytics Using Apache Spark can be one of the options to accompany you similar to having supplementary time.

It will not waste your time. allow me, the e-book will categorically proclaim you other issue to read. Just invest little grow old to door this on-line revelation **Pro Spark Streaming The Zen Of Real Time Analytics Using Apache Spark** as competently as review them wherever you are now.

429 - MACK WEBER

One day Sophie comes home from school to find two questions in her mail: "Who are you?" and "Where does the world come from?" Before she knows it she is enrolled in a correspondence course with a mysterious philosopher. Thus begins Jostein Gaarder's unique novel, which is not only a mystery, but also a complete and entertaining history of philosophy.

Society is now completely driven by data with many industries relying on data to conduct business or basic functions within the organization. With the efficiencies that big data bring to all institutions, data is continuously being collected and analyzed. However, data sets may be too complex for traditional data-processing, and therefore, different strategies must evolve to solve the issue. The field of big data works as a valuable tool for many different industries. The Research Anthology on Big Data Analytics, Architectures, and Applications is a complete reference source on big data analytics that offers the latest, innovative architectures and frameworks and explores a variety of applications within various industries. Offering an international perspective, the applications discussed within this anthology feature global representation. Covering topics such as advertising curricula, driven supply chain, and smart cities, this research anthology is ideal for data scientists, data analysts, computer engineers, software engineers, technologists, government officials, managers, CEOs, professors, graduate students, researchers, and academicians.

In this practical book, four Cloudera data scientists present a set of self-contained patterns for performing large-scale data analysis with Spark. The authors bring Spark, statistical methods, and real-world data sets together to teach you how to approach analytics problems by example. You'll start with an introduction to Spark and its ecosystem,

and then dive into patterns that apply common techniques—classification, collaborative filtering, and anomaly detection among others—to fields such as genomics, security, and finance. If you have an entry-level understanding of machine learning and statistics, and you program in Java, Python, or Scala, you'll find these patterns useful for working on your own data applications. Patterns include: Recommending music and the Audioscrobbler data set Predicting forest cover with decision trees Anomaly detection in network traffic with K-means clustering Understanding Wikipedia with Latent Semantic Analysis Analyzing co-occurrence networks with GraphX Geospatial and temporal data analysis on the New York City Taxi Trips data Estimating financial risk through Monte Carlo simulation Analyzing genomics data and the BDG project Analyzing neuroimaging data with PySpark and Thunder The professional programmer's Deitel® guide to Python® with introductory artificial intelligence case studies Written for programmers with a background in another high-level language, Python for Programmers uses hands-on instruction to teach today's most compelling, leading-edge computing technologies and programming in Python–one of the world's most popular and fastest-growing languages. Please read the Table of Contents diagram inside the front cover and the Preface for more details. In the context of 500+, real-world examples ranging from individual snippets to 40 large scripts and full implementation case studies, you'll use the interactive IPython interpreter with code in Jupyter Notebooks to quickly master the latest Python coding idioms. After covering Python Chapters 1-5 and a few key parts of Chapters 6-7, you'll be able to handle significant portions of the hands-on introductory AI case studies in Chapters 11-16, which are loaded with cool, powerful, contemporary examples. These include natural language processing, data mining Twitter® for sentiment analysis, cognitive computing with IBM® Watson™, supervised machine learning with classification and regression, unsupervised machine learning with clustering, computer vision through deep learning and convolutional neural networks, deep learning with recurrent neural networks, big data with Hadoop®, Spark™ and NoSQL databases, the Internet of Things and more. You'll also work directly or indirectly with cloud-based services, including Twitter, Google Translate™, IBM Watson, Microsoft® Azure®, OpenMapQuest, PubNub and more. Features 500+ hands-on, real-world, live--code examples from snippets to case studies IPython + code in Jupyter® Notebooks Library-focused: Uses Python Standard Library and data science libraries to accomplish significant tasks with minimal code Rich Python coverage: Control statements, functions, strings, files, JSON serialization, CSV, exceptions Procedural, functional-style and object-oriented programming Collections: Lists, tuples, dictionaries, sets, NumPy arrays, pandas Series & DataFrames Static, dynamic and interactive visualizations Data experiences with real-world datasets and data sources Intro to Data Science sections: AI, basic stats, simulation, animation, random variables, data wrangling, regression AI, big data and cloud data science case studies: NLP, data mining Twitter®, IBM® Watson™, machine learning, deep learning, computer vision, Ha-

doop®, Spark™, NoSQL, IoT Open-source libraries: NumPy, pandas, Matplotlib, Seaborn, Folium, SciPy, NLTK, TextBlob, spaCy, Textatistic, Tweepy, scikit-learn®, Keras and more Accompanying code examples are available here: http://ptgmedia.pearsoncmg.com/imprint_downloads/informit/-bookreg/9780135224335/9780135224335_examples.zip. Register your product for convenient access to downloads, updates, and/or corrections as they become available. See inside book for more information.

Winner of the International Lannan Literary Award for Nonfiction Animal tracks, word magic, the speech of stones, the power of letters, and the taste of the wind all figure prominently in this intellectual tour de force that returns us to our senses and to the sensuous terrain that sustains us. This major work of ecological philosophy startles the senses out of habitual ways of perception. For a thousand generations, human beings viewed themselves as part of the wider community of nature, and they carried on active relationships not only with other people with other an-imals, plants, and natural objects (including mountains, rivers, winds, and weather patters) that we have only lately come to think of as "inanimate." How, then, did humans come to sever their ancient reciprocity with the natural world? What will it take for us to recover a sustaining relation with the breathing earth? In The Spell of the Sensuous David Abram draws on sources as diverse as the philosophy of Merleau-Ponty, Balinese shamanism, Apache storytelling, and his own experience as an accomplished sleight-of--hand of magician to reveal the subtle dependence of human cognition on the natural environment. He explores the character of perception and excavates the sensual foundations of language, which--even at its most abstract--echoes the calls and cries of the earth. On every page of this lyrical work, Abram weaves his arguments with a passion, a precision, and an intellectual daring that recall such writers as Loren Eisleley, Annie Dillard, and Barry Lopez.

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.

Pulitzer Prize winner Tracy Kidder memorably records the drama, comedy, and excitement of one company's efforts to bring a new microcomputer to market. Computers have changed since 1981, when The Soul of a New Machine first examined the culture of the computer revolution. What has not changed is the feverish pace of the high-tech industry, the go-for-broke approach to business that has caused so many computer companies to win big (or go belly up), and the cult of pursuing mind-bending technological innovations. The Soul of a New Machine is an essential chapter in the history of the machine that revolutionized the world in the twentieth century.

Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applica-

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

Summary CORS in Action introduces Cross-Origin Resource Sharing (CORS) from both the server and the client perspective. It starts with the basics: how to make CORS requests and how to implement CORS on the server. It then explores key details such as performance, debugging, and security.

API authors will learn how CORS opens their APIs to a wider range of users. JavaScript developers will find valuable techniques for building rich web apps that can take advantage of APIs hosted anywhere. The techniques described in this book are especially applicable to mobile environments, where browsers are guaranteed to support CORS. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book Suppose you need to share some JSON data with another application or service. If everything is hosted on one domain, it's a snap. But if the data is on another domain, the browser's "same-origin" policy stops you cold. CORS is a new web standard that enables safe cross-domain access without complex server-side code. Mastering CORS makes it possible for web and mobile applications to share data simply and securely. CORS in Action introduces CORS from both the server and the client perspective. It starts with making and enabling CORS requests and then explores performance, debugging, and security. You'll learn to build apps that can take advantage

of APIs hosted anywhere and how to write APIs that expand your products to a wider range of users. For web developers comfortable with JavaScript. No experience with CORS is assumed. What's Inside CORS from the ground up Serving and consuming cross-domain data Best practices for building CORS APIs When to use CORS alternatives like JSON-P and proxies About the Author Monsur Hossain is an engineer at Google who has worked on API-related projects such as the Google JavaScript Client, the APIs Discovery Service, and CORS support for Google APIs. Table of Contents PART 1 INTRODUCING CORS The Core of CORS Making CORS requests PART 2 CORS ON THE SERVER Handling CORS requests Handling preflight requests Cookies and response headers Best practices PART 3 DEBUGGING CORS REQUESTS Debugging CORS requests APPENDIXES CORS reference Configuring your environment What is CSRF? Other cross-origin techniques An electrifying story of the sensational murder trial that divided a city and ignited the civil rights struggle In 1925, Detroit was a smoky swirl of jazz and speakeasies, assembly lines and fistfights. The advent of automobiles had brought workers from around the globe to compete for manufacturing jobs, and tensions often flared with the KKK in ascendance and violence rising. Ossian Sweet, a proud Negro doctor-grandson of a slave-had made the long climb from the ghetto to a home of his own in a previously all-white neighborhood. Yet just after his arrival, a mob gathered outside his house; suddenly, shots rang out: Sweet, or one of his defenders, had accidentally killed one of the whites threatening their lives and homes. And so it began-a chain of events that brought America's greatest attorney, Clarence Darrow, into the fray and transformed Sweet into a controversial symbol of equality. Historian Kevin Boyle weaves the police investigation and courtroom drama of Sweet's murder trial into an unforgettable tapestry of narrative history that documents the volatile America of the 1920s and movingly re-creates the Sweet family's journey from slavery through the Great Migration to the middle class. Ossian Sweet's story, so richly and poignantly captured here, is an epic tale of one man trapped by the battles of his era's changing times. Arc of Justice is the winner of the 2004 National Book Award for Nonfiction. Due to the growing use of web applications and communication devices, the use of data has increased throughout various industries, including business and healthcare. It is necessary to develop specific software programs that can analyze and interpret large amounts of data quickly in order to ensure adequate usage and predictive results. Cognitive Analytics: Concepts, Methodologies, Tools, and Applications provides emerging perspectives on the theoretical and practical aspects of data analysis tools and techniques. It also examines the incorporation of pattern management as well as decision-making and prediction processes through the use of data management and analysis. Highlighting a range of topics such as natural language processing, big data, and pattern recognition, this multi-volume book is ideally designed for information technology professionals, software developers, data analysts, graduate-level students, researchers, computer engineers, software engineers, IT specialists, and aca-

demicians.

There's growing interest in learning how to analyze streaming data in large-scale systems such as web traffic, financial transactions, machine logs, industrial sensors, and many others. But analyzing data streams at scale has been difficult to do well—until now. This practical book delivers a deep introduction to Apache Flink, a highly innovative open source stream processor with a surprising range of capabilities. Authors Ellen Friedman and Kostas Tzoumas show technical and nontechnical readers alike how Flink is engineered to overcome significant tradeoffs that have limited the effectiveness of other approaches to stream processing. You'll also learn how Flink has the ability to handle both stream and batch data processing with one technology. Learn the consequences of not doing streaming well—in retail and marketing, IoT, telecom, and banking and finance Explore how to design data architecture to gain the best advantage from stream processing Get an overview of Flink's capabilities and features, along with examples of how companies use Flink, including in production Take a technical dive into

Flink, and learn how it handles time and stateful computation Examine how Flink processes both streaming (unbounded) and batch (bounded) data without sacrificing performance

Construct a robust end-to-end solution for analyzing andvisualizing streaming data Real-time analytics is the hottest topic in data analyticstoday. In Real-Time Analytics: Techniques to Analyze andVisualize Streaming Data, expert Byron Ellis teaches dataanalysts technologies to build an effective real-time analyticsplatform. This platform can then be used to make sense of theconstantly changing data that is beginning to outpace traditionalbatch-based analysis platforms. The author is among a very few leading experts in the field. Hehas a prestigious background in research, development, analytics,real-time visualization, and Big Data streaming and is uniquelyqualified to help you explore this revolutionary field. Moving froma description of the overall analytic architecture of real-timeanalytics to using specific tools to obtain targeted results,Real-Time Analytics leverages open source and moderncom-

mercial tools to construct robust, efficient systems that canprovide real-time analysis in a cost-effective manner. The bookincludes: A deep discussion of streaming data systems andarchitectures Instructions for analyzing, storing, and delivering streamingdata Tips on aggregating data and working with sets Information on data warehousing options and techniques Real-Time Analytics includes in-depth case studies forwebsite analytics, Big Data, visualizing streaming and mobile data,and mining and visualizing operational data flows. The book's"recipe" layout lets readers quickly learn and implement differenttechniques. All of the code examples presented in the book, alongwith their related data sets, are available on the companionwebsite. 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.

By his early thirties, Paul Allen was a world-famous billionaire-and that was just the beginning. In 2007 and 2008, Time named Paul Allen, the co-founder of Microsoft, one of the hundred most influential people in the world. Since he made his fortune, his impact has been felt in science, technology, business, medicine, sports, music, and philanthropy. His passion, curiosity, and intellectual rigor-combined with the resources to launch and support new initiatives-have literally changed the world. In 2009 Allen discovered that he had lymphoma, lending urgency to his desire to share his story for the first time. In this classic memoir, Allen explains how he has solved problems, what he's learned from his many endeavors-both the triumphs and the failures-and his compelling vision for the future. He reflects candidly on an ex-

traordinary life. The book also features previously untold stories about everything from the true origins of Microsoft to Allen's role in the dawn of private space travel (with Space-ShipOne) and in discoveries at the frontiers of brain science. With honesty, humor, and insight, Allen tells the story of a life of ideas made real.

This is the definitive work on the first and greatest of Japan's twentieth-century philosophers, Nishida Kitaro (1870-1945). Interspersed throughout the narrative of Nishida's life and thought is a generous selection of the philosopher's own essays, letters, and short presentations, newly translated into English.

Apache Spark is a fast, scalable, and flexible open source distributed processing engine for big data systems and is one of the most active open source big data projects to date. In just 24 lessons of one hour or less, Sams Teach Yourself Apache Spark in 24 Hours helps you build practical Big Data solutions that leverage Spark's amazing speed, scalability, simplicity, and versatility. This book's straightforward, step-by-step approach shows you how to deploy, program,

optimize, manage, integrate, and extend Spark–now, and for years to come. You'll discover how to create powerful solutions encompassing cloud computing, real-time stream processing, machine learning, and more. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Whether you are a data analyst, data engineer, data scientist, or data steward, learning Spark will help you to advance your career or embark on a new career in the booming area of Big Data. Learn how to • Discover what Apache Spark does and how it fits into the Big Data landscape • Deploy and run Spark locally or in the cloud • Interact with Spark from the shell • Make the most of the Spark Cluster Architecture • Develop Spark applications with Scala and functional Python • Program with the Spark API, including transformations and actions • Apply practical data engineering/analysis approaches designed for Spark • Use Resilient Distributed Datasets (RDDs) for caching, persistence, and output • Optimize Spark solution performance • Use Spark with SQL (via Spark SQL) and with NoSQL (via Cassan-

dra) • Leverage cutting-edge functional programming techniques • Extend Spark with streaming, R, and Sparkling Water • Start building Spark-based machine learning and graph-processing applications • Explore advanced messaging technologies, including Kafka • Preview and prepare for Spark's next generation of innovations Instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Spark to solve a wide spectrum of Big Data problems.

Since he began posting in 2003, Dempsey has used his blog to explore nearly every important facet of library technology, from the emergence of Web 2.0 as a concept to open source ILS tools and the push to web-scale library management systems. "Designing a large software system is an extremely complicated undertaking that requires juggling differing perspectives and differing goals, and evaluating differing

options. Applied Software Architecture is the best book yet that gives guidance as to how to sort out and organize the conflicting pressures and produce a successful design." -- Len Bass, author of Software Architecture in Practice. Quality software architecture design has always been important, but in today's fast-paced, rapidly changing, and complex development environment, it is essential. A solid, well-thought-out design helps to manage complexity, to resolve trade-offs among conflicting requirements, and, in general, to bring quality software to market in a more timely fashion. Applied Software Architecture provides practical guidelines and techniques for producing quality software designs. It gives an overview of software architecture basics and a detailed guide to architecture design tasks, focusing on four fundamental views of architecture--conceptual, module, execution, and code. Through four real-life case studies, this book reveals the insights and best practices of the most skilled software architects in designing software architecture. These case studies, written with the masters who created them, demonstrate how the book's concepts and techniques are embodied in state-of-the-art architecture design. You will learn how to: create designs flexible enough to incorporate tomorrow's technology; use architecture as the basis for meeting performance, modifiability, reliability, and safety requirements; determine priorities among conflicting requirements and arrive at a successful solution; and use software architecture to help integrate system components. Anyone involved in software architecture will find this book a valuable compendium of best practices and an insightful look at the critical role of architecture in software development. 0201325713B07092001

Fifth-grade language arts teacher Mrs. Granger comes to regret her lesson that language is controlled by human beings when Nick Allen begins a wildly successful campaign to rename the ballpoint pen. Reprint.

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 deploy-

ments 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.

Production-targeted Spark guidance with real-world use cases Spark: Big Data Cluster Computing in Production goes beyond general Spark overviews to provide targeted guidance toward using lightning-- fast big-data clustering in production. Written by an expert team well-known in the big data community, this book walks you through the challenges in moving from proof-of-concept or demo Spark appli-

cations to live Spark in production. Real use cases provide deep insight into common problems, limitations, challenges, and opportunities, while expert tips and tricks help you get the most out of Spark performance. Coverage includes Spark SQL, Tachyon, Kerberos, ML Lib, YARN, and Mesos, with clear, actionable guidance on resource scheduling, db connectors, streaming, security, and much more. Spark has become the tool of choice for many Big Data problems, with more active contributors than any other Apache Software project. General introductory books abound, but this book is the first to provide deep insight and real-world advice on using Spark in production. Specific guidance, expert tips, and invaluable foresight make this guide an incredibly useful resource for real production settings. Review Spark hardware requirements and estimate cluster size Gain insight from real-world production use cases Tighten security, schedule resources, and fine-tune performance Overcome common problems encountered using Spark in production Spark works with other big data tools including MapReduce and Ha-

doop, and uses languages you already know like Java, Scala, Python, and R. Lightning speed makes Spark too good to pass up, but understanding limitations and challenges in advance goes a long way toward easing actual production implementation. Spark: Big Data Cluster Computing in Production tells you everything you need to know, with real-- world production insight and expert guidance, tips, and tricks.

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 START-ED WITH KAFKA STREAMS Welcome to Kafka Streams Kafka quickly-PART 2 - KAFKA STREAMS DEVELOPMENT Developing Kafka Streams Streams and state The

KTable API The Processor APIPART 3 - ADMINISTER-ING KAFKA STREAMS Monitoring and performance Testing a Kafka Streams applicationPART 4 - AD-VANCED CONCEPTS WITH KAFKA STREAMS Advanced applications with Kafka StreamsAPPENDIX-ES Appendix A - Additional configuration information Appendix B - Exactly once semantics

This extraordinary book explains the engine that has catapulted the Internet from backwater to ubiquity—and reveals that it is sputtering precisely because of its runaway success. With the unwitting help of its users, the generative Internet is on a path to a lockdown, ending its cycle of innovation—and facilitating unsettling new kinds of control. IPods, iPhones, Xboxes, and TiVos represent the first wave of Internet-centered products that can't be easily modified by anyone except their vendors or selected partners. These "tethered appliances" have already been used in remarkable but little-known ways: car GPS systems have been reconfigured at the demand of law enforcement to eavesdrop on the occupants at all times, and digital video recorders have been ordered to self-de-

struct thanks to a lawsuit against the manufacturer thousands of miles away. New Web 2.0 platforms like Google mash-ups and Facebook are rightly touted—but their applications can be similarly monitored and eliminated from a central source. As tethered appliances and applications eclipse the PC, the very nature of the Internet—its "generativity," or innovative character—is at risk. The Internet's current trajectory is one of lost opportunity. Its salvation, Zittrain argues, lies in the hands of its millions of users. Drawing on generative technologies like Wikipedia that have so far survived their own successes, this book shows how to develop new technologies and social structures that allow users to work creatively and collaboratively, participate in solutions, and become true "netizens."

By introducing in-memory persistent storage, Apache Spark eliminates the need to store intermediate data in filesystems, thereby increasing processing speed by up to 100 times. This book will focus on how to analyze large and complex sets of data. Starting with installing and configuring Apache Spark with various

cluster managers, you will cover setting up development environments. You will then cover various recipes to perform interactive queries using Spark SQL and real-time streaming with various sources such as Twitter Stream and Apache Kafka. You will then focus on machine learning, including supervised learning, unsupervised learning, and recommendation engine algorithms. After mastering graph processing using GraphX, you will cover various recipes for cluster optimization and troubleshooting.

Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates. Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. 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. Quickly dive into Spark capabilities such as distributed datasets, in--memory caching, and the interactive shell Leverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLlib Use one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and Storm Learn how to deploy interactive, batch, and streaming applications Connect to data sources including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables

First published in 2000. Routledge is an imprint of Taylor & Francis, an informa company.

Das Internet sorgt für ein rasches Wachstum von verfügbaren Daten. Dem „EMC/IDC Digital Universe" Report nach verdoppelt sich die Menge an Daten alle zwei Jahre. Es werden täglich eine Millionen Uber-Fahrten gebucht, monatlich 10 Milliarden Stunden Netflix-Videos konsumiert und jährlich eine Trillion US--Dollar auf E-Commerce-Webseiten ausgegeben. Dabei spielt der Umgang mit diesen Daten eine große Rolle für den Erfolg dieser Unternehmen. Insbesondere die Echtzeit-Datenanalyse hilft den Unternehmen, Kundenwünsche schnellstmöglich zu erkennen und diese auszuschöpfen. Ein weiterer Faktor des immensen Datenwachstums sind Web 2.0 Plattformen. In diesem Bereich sind die Benutzer der treibende Faktor für die stetig wachsende Datenflut dieser Plattformen, indem sie Beiträge veröffentlichen, Fotos teilen und Videos streamen. Beispielsweise sammelt Twitter 8 Terabyte, Facebook 500 Terabyte und Google 20 Terabyte an benutzergenerierten Daten täglich. Inhalte sind nicht länger statisch, sondern dynamisch. Diese Datenflut gilt es, zu kontrollieren. Unmittelbar nachdem Daten entstehen, haben sie den höchsten Wert. In der heutigen Zeit wird erwartet, Ergebnisse in schnellstmöglicher Zeit bereitzustellen. Analytische Prozesse werden in Minuten, Sekunden oder Bruchteilen von Sekunden durchgeführt, welche früher Monate, Tage oder Stunden gebraucht hätten. Aufgrund dieser Evolution der stetig neu entstehenden Daten müssen Veränderungen in der Art und Weise, wie diese Dat-

en empfangen, transportiert, verarbeitet und analysiert werden, stattfinden. Die Technologie, die Echtzeit-Datenanalysen durchführt, wird in der vorliegenden Arbeit genauer betrachtet.

A neuroscientist and Zen practitioner interweaves the latest research on the brain with his personal narrative of Zen. Aldous Huxley called humankind's basic trend toward spiritual growth the "perennial philosophy." In the view of James Austin, the trend implies a "perennial psychophysiology"—because awakening, or enlightenment, occurs only when the human brain undergoes substantial changes. What are the peak experiences of enlightenment? How could these states profoundly enhance, and yet simplify, the workings of the brain? Zen and the Brain presents the latest evidence. In this book Zen Buddhism becomes the opening wedge for an extraordinarily wide-ranging exploration of consciousness. In order to understand which brain mechanisms produce Zen states, one needs some understanding of the anatomy, physiology, and chemistry of the brain. Austin, both a neurologist and a Zen practitioner, interweaves the most recent brain research with the personal narrative of his Zen experiences. The science is both inclusive and rigorous; the Zen sections are clear and evocative. Along the way, Austin examines such topics as similar states in other disciplines and religions, sleep and dreams, mental illness, consciousness-altering drugs, and the social consequences of the advanced stage of ongoing enlightenment.

Do you know what "quatrefoil" and "impolitic" mean? What about "halcyon" or "narcolepsy"? This book is a handy, easy-to-read reference guide to the proper parlance for any situation. In this book you will find: Words You Absolutely Should Know (covert, exonerate, perimeter); Words You Should Know But Probably Don't (dour, incendiary, scintilla); Words Most People Don't Know (schlimazel, thaumaturgy, epergne); Words You Should Know to Sound Overeducated (ad infinitum, nugatory, garrulity); Words You Probably Shouldn't Know (priapic, damnatory, labia majora); and more. Whether writing an essay, studying for a test, or trying to impress friends, family, and fellow cocktail party guests with their prolixity, you will achieve magniloquence, ebullience, and flights of rhetorical brilliance.

Apache Spark is amazing when everything clicks. But if you haven't seen the performance improvements you expected, or still don't feel confident enough to use Spark in production, this practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources. Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you'll also learn how to make it sing. With this book, you'll explore: How Spark SQL's new interfaces improve performance over SQL's RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark's key/value pair paradigm Writing

high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLlib and Spark ML machine learning libraries Spark's Streaming components and external community packages

Analyze vast amounts of data in record time using Apache Spark with Databricks in the Cloud. Learn the fundamentals, and more, of running analytics on large clusters in Azure and AWS, using Apache Spark with Databricks on top. Discover how to squeeze the most value out of your data at a mere fraction of what classical analytics solutions cost, while at the same time getting the results you need, incrementally faster. This book explains how the confluence of these pivotal technologies gives you enormous power, and cheaply, when it comes to huge datasets. You will begin by learning how cloud infrastructure makes it possible to scale your code to large amounts of processing units, without having to pay for the machinery in advance. From there you will learn how Apache Spark, an open source framework, can enable all those CPUs for data analyt-

ics use. Finally, you will see how services such as Databricks provide the power of Apache Spark, without you having to know anything about configuring hardware or software. By removing the need for expensive experts and hardware, your resources can instead be allocated to actually finding business value in the data. This book guides you through some advanced topics such as analytics in the cloud, data lakes, data ingestion, architecture, machine learning, and tools, including Apache Spark, Apache Hadoop, Apache Hive, Python, and SQL. Valuable exercises help reinforce what you have learned. What You Will Learn Discover the value of big data analytics that leverage the power of the cloudGet started with Databricks using SQL and Python in either Microsoft Azure or AWSUnderstand the underlying technology, and how the cloud and Apache Spark fit into the bigger picture See how these tools are used in the real world Run basic analytics, including machine learning, on billions of rows at a fraction of a cost or free Who This Book Is For Data engineers, data scientists, and cloud architects who want or need to run ad-

vanced analytics in the cloud. It is assumed that the reader has data experience, but perhaps minimal exposure to Apache Spark and Azure Databricks. The book is also recommended for people who want to get started in the analytics field, as it provides a strong foundation.

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 effi-

cient 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 MLlib. 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.

When The Sound of the One Hand came out in Japan in 1916 it caused a scandal. Zen was a secretive practice, its wisdom relayed from master to novice in strictest privacy. That a handbook existed recording not only the riddling koans that are central to Zen teaching but also detailing the answers to them seemed to mark Zen as rote, not revelatory. For all that, The Sound of the One Hand opens the door to Zen like no other book. Including koans that go back to the master who first brought the koan teaching method from China to Japan in the eighteenth century, this book offers, in the words of the translator, editor, and Zen initiate Yoel Hoffmann, the clearest, most detailed, and most correct picture of Zen that can be found. What we have here is an extraordinary introduction to Zen thought as lived thought, a treasury of problems, paradoxes, and performance that will appeal to artists, writers, and philosophers as well as Buddhists and students of religion."

'Wishful Drinking is a touching and incisive account of bipolarity, addiction and motherhood.' Independent 'No motive is pure. No one is good or bad – but a hearty mix of both. And sometimes life actually gives to you by taking away.' Carrie Fisher in Wishful Drinking In Wishful Drinking, Carrie Fisher told the true and intoxicating story of her life with inimitable wit. Born to celebrity parents, she was picked to play a princess in a little movie called Star Wars when only 19 years old. "But it isn't all sweetness and light sabres." Alas, aside from a demanding career and her role as a single mother (not to mention the hyperspace hairdo), Carrie also spends her free time battling addiction and weathering the wild ride of manic depression. It's an incredible tale: from having Elizabeth Taylor as a step-

mother, to marrying (and divorcing) Paul Simon, and from having the father of her daughter leave her for a man, to ultimately waking up one morning and finding a friend dead beside her in bed. Carrie Fisher's star-studded career included roles in numerous films such as The Blues Brothers and When Harry Met Sally. She was the author of four best-selling novels, Surrender in the Pink, Delusions of Grandma, The Best Awful and Postcards from the Edge, which was made into a hit film starring Shirley MacLaine and Meryl Streep. Carrie's experience with addiction and mental illness – and her willingness to talk honestly about them – made her a sought-after speaker and respected advocate. She was truly one of the most magical people to walk among us. Further praise for Carrie Fisher:-[Shockaholic] is the finest, funniest chronicler of the maddest celebrity mores.' Sunday Times 'Fisher has a talent for lacerating insight that masquerades as carefree self-deprecation' Los Angeles Times 'She is one of the rare inhabitants of La-La Land who can actually write' New York Times
Summary Spring Integration in Action is a hand-s-on guide to Spring-based messaging and integration. After addressing the core messaging patterns, such as those used in transformation and routing, the book turns to the adapters that enable integration with external systems. Readers will explore real-world enterprise integration scenarios using JMS, Web Services, file systems, and email. They will also learn about Spring Integration's support for working with XML. The book concludes with a practical guide to advanced topics such as concurrency, performance, system-management, and monitoring. The book features a foreword by Rod Johnson, Founder of the Spring Network. About the Technology Spring Integration extends the Spring Framework to support the patterns described in Gregor Hohpe and Bobby Woolf's Enterprise Integration Patterns. Like the Spring Framework itself, it focuses on developer productivity, making it easier to build, test, and maintain enterprise integration solutions. About the Book Spring Integration in Action is an introduction and guide to enterprise integration and messaging using the Spring Integration framework. The book starts off by reviewing core messaging patterns, such as those used in transformation and routing. It then drills down into real-world enterprise integration scenarios using JMS, Web Services, filesystems, email, and more. You'll find an emphasis on testing, along with practical coverage of topics like concurrency, scheduling, system management, and monitoring. This book is accessible to developers who know Java. Experience with Spring and EIP is helpful but not assumed. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside Realistic examples Expert advice from Spring Integration creators Detailed coverage of Spring Integration 2 features About the Authors Mark Fisher is the Spring Integration founder and project lead. Jonas Partner, Marius Bogoevici, and Iwein Fuld have all been project committers and are recognized experts on Spring and Spring Integration. Table of Contents PART 1 BACKGROUND Introduction to Spring Integration Enterprise integration fundamentals 24 PART 2 MESSAGING Messages and channels Message End-

points Getting down to business Go beyond sequential processing: routing and filtering Splitting and aggregating messages PART 3 INTEGRATING SYSTEMS Handling messages with XML payloads Spring Integration and the Java Message Service Email-based integration Filesystem integration Spring Integration and web services Chatting and tweeting PART 4 ADVANCED TOPICS Monitoring and management Managing scheduling and concurrency Batch applications and enterprise integration Scaling messaging applications with OSGi Testing

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 LearnDiscover 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.

This comprehensive theory and practice of Christian spiritual formation weaves together biblical and theological foundations with interdisciplinary scholarship, real-world examples, personal vignettes, and practical tools to assist readers in becoming whole persons in relationship with God and others.

For this third edition of - Distributed Systems, - the material has been thoroughly revised and extended, integrating principles and paradigms into nine chapters: 1. Introduction 2. Architectures 3. Processes 4. Communication 5. Naming 6. Coordination 7. Replication 8. Fault tolerance 9. Security A separation has been made between basic material and more specific subjects. The latter have been organized into boxed sections, which may be skipped on first reading. To assist in understanding the more algorithmic parts, example programs in Python have been included. The examples in the book leave out many details for readability, but the complete code is available through the book's Website, hosted at www.distributed-systems. net. A personalized digital copy of the book is available for free, as well as a printed version through Amazon.com.

Estamos na era dos dados. Não importa qual seja a sua área de atuação, uma palavra atualmente em evidência é: Big Data. Podemos encontrar casos de uso em que esse conceito permitiu a redução do número de fraudes, redução de custos na produção, eficiência energética, aumento de segurança, entre outros benefí-

cios tão almejados em diversos domínios. Muito embora o interesse esteja em alta, Big Data ainda é um termo incipiente, gerando incertezas sobre sua definição, característica, aplicabilidade e desafios. Como obter dados de diferentes fontes? Como extrair valor a partir dos dados? Qual a infraestrutura necessária para criar uma solução de Big Data? Quais habilidades são necessárias para atuar com Big Data em seu projeto? Neste livro, Rosangela Marquesone apresenta as tecnologias e soluções de Big Data, em uma abordagem conceitual que detalha as características e capacidades de cada uma delas. Você verá as principais fases de um projeto de Big Data, desde a captura, o armazenamento, o processamento, análise, até a visualização de dados.