

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

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.

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 architectural and implementation details of every technology. 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 to integrate, replace, and reinforce every layer: The language: Scala The engine: Spark (SQL, MLlib, 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 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

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.

The quick growth of computer technology and development of software caused it to be in a constant state of change and advancement. This advancement in software development meant that there would be many types of software developed in order to excel in usability and efficiency. Among these different types of software was open source software, one that grants permission for users to use, study, change, and distribute it freely. Due to its availability, open source software has quickly become a valuable asset to the world of computer technology and across various disciplines including education, business, and library science. The Research Anthology on Usage and Development of Open Source Software presents comprehensive research on the design and development of open source software as well as the ways in which it is used. The text discusses in depth the way in which this computer software has been made into a collaborative effort for the advancement of software technology. Discussing topics such as ISO standards, big data, fault prediction, open collaboration, and software development, this anthology is essential for computer engineers, software developers, IT specialists and consultants, instructors, librarians, managers, executives, professionals, academicians, researchers, and students.

The Good Girl’s Guide to Getting Lost
Service-Oriented Computing – ICSC 2018 Workshops

A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka
Fast Data Processing Systems with SMACK Stack

Proceedings of the 2020 Computing Conference, Volume 2
Real-time apps and microservices with the Kafka Streams API

Spark: The Definitive Guide

Follow this handbook to build, configure, tune, and secure Apache Cassandra databases. Start with the installation of Cassandra and move on to the creation of a single instance, and then a cluster of Cassandra databases. Cassandra is increasingly a key player in many big data environments, and this book shows you how to use Cassandra with Apache Spark, a popular big data processing framework. Also covered are day-to-day topics of importance such as the backup and recovery of Cassandra databases, using the right compression and compaction strategies, and loading and unloading data. Expert Apache Cassandra Administration provides numerous step-by-step examples starting with the basics of a Cassandra database, and going all the way through backup and recovery, performance optimization, and monitoring and securing the data. The book serves as an authoritative and comprehensive guide to the building and management of simple to complex Cassandra databases. The book takes you through building a Cassandra database from installation of the software and creation of a single database, through to complex clusters and data centers Provides numerous examples of actual commands in a real-life Cassandra environment that show how to confidently configure, manage, troubleshoot, and tune Cassandra databases Shows how to use the Cassandra configuration properties to build a highly stable, available, and secure Cassandra database that always operates at peak efficiency What You’ll Learn Install the Cassandra software and create your first database Understand the Cassandra data model, and the internal architecture of a Cassandra database Create your own Cassandra cluster, step-by-step Run a Cassandra cluster on Docker Work with Apache Spark by connecting to a Cassandra database Deploy Cassandra clusters in your data center, or on Amazon EC2 instances Back up and restore mission-critical Cassandra databases Monitor, troubleshoot, and tune production Cassandra databases, and cut your spending on resources such as memory, servers, and storage Who This Book Is For Database administrators, developers, and architects who are looking for an authoritative and comprehensive single volume for all their Cassandra administration needs. Also for administrators who are tasked with setting up and maintaining highly reliable and high-performing Cassandra databases. An excellent choice for big data administrators, database administrators, architects, and developers who use Cassandra as their key data store, to support high volume online transactions, or as a decentralized, elastic data store.

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 MET 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.

A step-by-step guide for turning information into advantage This book describes a ten-step method that empowers companies to transform their information into knowledge, helping managers develop and maintain a balanced knowledge plan, solve information shortfalls, and take advantage of the information at their fingertips.

Work with the Data: A Handbook on How 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 MLlib, 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 componentsIntegrate 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.

Handbook of Research on Big Data Storage and Visualization Techniques

Using the Scala API
Practical Apache Spark

The Last Competitive Advantage

There Is a Tribe of Kids

Recognizing Threats, Defending Your Rights, and Protecting Your Family

A Memoir of Three Continents, Two Friends, and One Unexpected Adventure

How far would you go to find something that might not even exist? All her life, Cricket’s mama has told her stories about a secret room painted by a mysterious artist. Now Mama’s run off, and Cricket thinks the room might be the answer to getting her to come back. If it exists. And if she can find it. Cricket’s first clue is a coin from a grown-over ghost town in the woods. So with her daddy’s old guidebook and a coat full of snacks stolen from the Cash ‘n Carry, Cricket runs away to find the room. Surviving in the woods isn’t easy. While Cricket camps out in an old tree house and looks for clues, she meets the last resident of the ghost town, encounters a poetry-loving dog (who just might hold a key to part of the puzzle), and discovers that sometimes you have to get a little lost . . . to really find your way.

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.

A practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario *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 also expand 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 DRY style, with plenty of practical use cases, well-explained code examples, and relevant screenshots and diagrams.*

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.

A Girl’s Guide to Skateboarding

Next-Generation Big Data

Pro Spark Streaming

Distributed Computing and Event Processing using Apache Spark, Flink, Storm, and Kafka

ADMS, ASOCA, ISYCC, CloTS, DBBS, and NLS4IoT, Hangzhou, China, November 12–15, 2018, Revised Selected Papers

White Space Is Not Your Enemy

A Practitioner’s Guide to Using Spark for Large Scale Data Analysis

Written by the creators of MySQL, and edited by one of the most highly respected MySQL authors, the MySQL Administrator’s Guide and Language Reference is the official guide to installing MySQL, to setting up and administering MySQL databases, and to storing and retrieving data in these databases. This new edition combines into one book the MySQL Language Reference (on CD) with the practical information of the MySQL Administrator’s Guide book.

Colorful introduction to skateboarding for girls.

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.

There are more than one billion Android devices in use today, each one a potential target. Unfortunately, many fundamental Android security features have been little more than a black box to all but the most elite security professionals—until now. In Android Security Internals, top Android security expert Nikolay Elenkov takes us under the hood of the Android security system. Elenkov describes Android security architecture from the bottom up, delving into the implementation of major security-related components and subsystems, like Binder IPC, permissions, cryptographic providers, and device administration. You’ll learn: –How Android permissions are declared, used, and enforced –How Android manages application packages and employs code signing to verify their authenticity –How Android implements the Java Cryptography Architecture (JCA) and Java Secure Socket Extension (JSSE) frameworks –About Android’s credential storage system and APIs, which let applications store cryptographic keys securely –About the online account management framework and how Google accounts integrate with Android –About the implementation of verified boot, disk encryption, lockscreen, and other device security features –How Android’s bootloader and recovery OS are used to perform full system updates, and how to obtain root access With its unprecedented level of depth and detail, Android Security Internals is a must-have for any security-minded Android developer.

Research Anthology on Usage and Development of Open Source Software

Lightning-Fast Big Data Analysis

Big Data Analytics with Spark

The UNIX-halers Handbook

Big Data Processing Made Simple

MySQL Administrator’s Guide and Language Reference

The Definitive Guide to Exchange Disaster Recovery and Availability

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

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, Cermer, 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 Damaeet for data wrangling and interactive data processing Turbocharge Spark with Aluxio, 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 how to manage 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, Cermer, 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

Summary Kafka Streams in Action teaches you everything you need to know to implement stream processing on 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 StreamsAPPENDIX Appendix A - Additional configuration information Appendix B - Exactly once semantics

Explores the alphabet and animals through their collective nouns, including a covey of quails, a string of ponies, and a murder of crows.

Out of My Mind

Big Data SMACK

Practical Real-time Data Processing and Analytics

Open Source Software for Statistical Analysis of Big Data: Emerging Research and Opportunities

A Slap in the Face

Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 2

Kafka Streams in Action

Process large volumes of data in real-time while building high performance and robust data stream processing pipeline using the latest Apache Kafka 2.0 Key FeaturesSolve practical large data and processing challenges with KafkaTackle data processing challenges like late events, windowing, and watermarkingUnderstand real-time streaming applications processing using Schema registry, Kafka connect, Kafka streams, and KSQLBook Description Apache Kafka is a great open source platform for handling your real-time data pipeline to ensure high-speed filtering and pattern matching on the fly. In this book, you will learn how to use Apache Kafka for efficient processing of distributed applications and will get familiar with solving everyday problems in fast data and processing pipelines. This book focuses on programming rather than the configuration management of Kafka clusters or DevOps. It starts off with the installation and setting up the development environment, before quickly moving on to performing functional 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 as avro, XML, JSON, and AvRO. Next, you will learn how to expose the schemas contained in Kafka with the Schema Registry. You will then learn how to work with all relevant connectors with Kafka Connect. While working with Kafka Streams, you will perform various interesting operations on streams, such as windowing, joins, and aggregations. Finally, through KSQL, you will learn how to retrieve, insert, modify, and delete data streams, and how to manipulate watermarks and windows. What you will learnHow to validate data with KafkaAdd information to existing data flowsGenerate new information through message compositionPerform data validation and versioning with the Schema RegistryHow to perform message Serialization and DeserializationHow to perform message Serialization and DeserializationProcess data streams with Kafka StreamsUnderstand the duality between tables and streams with KSQLWho this book is for This book is for developers who want to quickly master the practical concepts behind Apache Kafka. The audience need not have come across Apache Kafka previously; however, a familiarity of Java or any JVM language will be helpful in understanding the code in this book.

Difficult and time-consuming is not an option when you need 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 – 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.

Considered by many to be mentally retarded, a brilliant, impatient fifth-grader with cerebral palsy discovers a technological device that will allow her to speak for the first time. 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

Executive’s Guide to Knowledge Management

What Every SQL Professional Needs to Know about Non-Relational Databases

Apache Kafka Quick Start Guide

The Zen of Real-Time Analytics Using Apache Spark

An Alphabet Book

An In-Depth Guide to Android’s Security Architecture

AETA 2017 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application

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 dive deep 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 efficient 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. Summary This book introduces Apache Kafka 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—showing 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: Broklyn, Mesos, Kafka, and Zeppelin See how Singley can be used to install Mule ESB on a cluster and Cassandra in the cloud Install and use CDOS 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.

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.

This book is about 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 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 architectural and implementation details of every technology. 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 to integrate, replace, and reinforce every layer: The language: Scala The engine: Spark (SQL, MLlib, 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. 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.

Why Insults Hurt—And Why They Shouldn’t

A Beginner’s Guide to Communicating Visually Through Graphic, Web & Multimedia Design

A Strategic Guide to the Network Economy

Have You Ever Seen a Smack of Jellyfish?

A Complete Hand Book for the Use of the Lady in Polite Society

Android Security Internals

Intelligent Systems and Applications

Winner of the Kate Greenaway Medal When a young boy embarks on a journey alone . . . he trails a colony of penguins, undulates in a smack of jellyfish, clasps hands with a constellation of stars, naps for a night in a bed of clams, and follows a trail of shells, home to his tribe of friends. If Lane Smith’s Caldecott Honor Book Grandpa Green was an homage to aging and the end of life, There Is a Tribe of Kids is a meditation on childhood and life’s beginning. Smith’s vibrant sponge-paint illustrations and use of unusual collective nouns such as smack and unkindness bring the book to life. Whimsical, expressive, and perfectly paced, this story plays with language as much as it embodies imagination, and was awarded the 2017 Kate Greenaway Medal. This title has Common Core connections.

Insults are part of the fabric of daily life. But why do we insult each other? Why do insults cause us such pain? Can we do anything to prevent or lessen this pain? Most importantly, how can we overcome our inclination to insult others? In A Slap in the Face, William Irvine undertakes a wide-ranging investigation of insults, their history, the role they play in social relationships, and the science behind them. He examines not just the obvious, such as Elizabeth Bowen’s description of Aldous Huxley as “The stupid person’s idea of a clever person,” but subtle insults as well, such as when someone insults us by reporting the insulting things others have said about us: “I never read bad reviews about myself.” wrote entertainer Oscar Levant, “because my best friends invariably tell me about them.” Irvine also considers the role insults play in our society: they can be used to cement relations, as when a woman playfully teases her husband, or to enforce a social hierarchy, as when a boss publicly berates an employee. He goes on to investigate the many ways society has tried to deal with insults: by adopting codes of politeness, for example, and outlawing hate speech—but concludes that the best way to deal with insults is to immunize ourselves against them: We need to transform ourselves in the manner recommended by Stoic philosophers. We should, more precisely, become insult pacifists, trying hard not to insult others and laughing off their attempts to insult us. A rousing follow-up to A Guide to the Good Life, A Slap in the Face will interest anyone who’s ever delivered an insult or felt the sting of one—in other words, everyone.

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.

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, ICSC 2018, held in Hangzhou, China, in November 2018. The ICSC 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 (DBBS)First International Workshop on Networked Learning Systems for Secured IoT Services and Its Applications (NLS4IoT)8th International Workshop on Context-Aware and IoT Services (CloTS)Third International Workshop on Adaptive Service-oriented and Cloud Applications (ASOCA2018)Third International Workshop on IoT Systems for Context-aware Computing (ISCC)First International Workshop on AI and Data Mining for Services (ADMS)

Learning Spark
Expert Apache Cassandra Administration
Leverage Apache Kafka 2.0 to simplify real-time data processing for distributed applications
Joe Celko's Complete Guide to NoSQL
Smack Dab in the Middle of Maybe
Information Rules
Research Anthology on Privatizing and Securing Data

As one of the first books to distill the economics of information and networks into practical business strategies, this is a guide to the winning moves that can help business leaders--from writers, lawyers and finance professional to executives in the entertainment, publishing and hardware and software industries-- navigate successfully through the information economy. 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

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

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 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.

Intelligent Computing
Complete Guide to Open Source Big Data Stack
Privacy in the Age of Big Data
Emerging Research and Opportunities
Skater Girl

A Practical Guide to Apache Kudu, Impala, and Spark
The Ladies' Book of Etiquette, and Manual of Politeness

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.

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.