
Spark In Action Free Pdf

Data Algorithms with Spark
Spark in Action
Spark in Action, Second Edition
The Book of R
High Performance Spark
Spark
Spring Batch in Action
Spark
The Spark of Learning
Event Streams in Action
Spark in Action
Relevant Search
Rx.NET in Action
The Big Book of Conflict Resolution Games: Quick, Effective Activities to Improve Communication, Trust and Collaboration
Learning Spark SQL
Jazz
A Wrinkle in Time
Mastering Spark with R
How To Win Friends And Influence People
Spark Cookbook
Spark GraphX in Action
Nicomachean Ethics
Learning Spark
Spark: The Definitive Guide
Spark in Action
Hadoop in Action
Learning Spark
Practical Apache Spark
Beginning Apache Spark 2
Python for Geeks
Learning PySpark
Graph Algorithms
Apache Spark for the Enterprise: Setting the Business Free
Frank Kane's Taming Big Data with Apache Spark and Python
A Long Walk to Water
Node.js in Action
The Book Thief
Data Engineering with Apache Spark, Delta Lake, and Lakehouse
Apache Spark Deep Learning Cookbook

Spark In Action Free Pdf

Downloaded from ftp.wtvq.com by guest

ADRIENNE ROTH

Data Algorithms with Spark WaterBrook

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 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 Learn Discover the functional programming features of Scala Understand the complete architecture of Spark and its components Integrate Apache Spark with Hive and Kafka Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries Work with different machine learning concepts and libraries using Spark's MLlib packages Who This Book Is For Developers and professionals who deal with batch and stream data processing.

Spark in Action Packt Publishing Ltd

Build data-intensive applications locally and deploy at scale using the combined powers of Python and Spark 2.0 About This Book Learn why and how

you can efficiently use Python to process data and build machine learning models in Apache Spark 2.0 Develop and deploy efficient, scalable real-time Spark solutions Take your understanding of using Spark with Python to the next level with this jump start guide Who This Book Is For If you are a Python developer who wants to learn about the Apache Spark 2.0 ecosystem, this book is for you. A firm understanding of Python is expected to get the best out of the book. Familiarity with Spark would be useful, but is not mandatory. What You Will Learn Learn about Apache Spark and the Spark 2.0 architecture Build and interact with Spark DataFrames using Spark SQL Learn how to solve graph and deep learning problems using GraphFrames and TensorFrames respectively Read, transform, and understand data and use it to train machine learning models Build machine learning models with MLlib and ML Learn how to submit your applications programmatically using spark-submit Deploy locally built applications to a cluster In Detail Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. This book will show you how to leverage the power of Python and put it to use in the Spark ecosystem. You will start by getting a firm understanding of the Spark 2.0 architecture and how to set up a Python environment for Spark. You will get familiar with the modules available in PySpark. You will learn how to abstract data with RDDs and DataFrames and understand the streaming capabilities of PySpark. Also, you will get a thorough overview of machine learning capabilities of PySpark using ML and MLlib, graph processing using GraphFrames, and polyglot persistence using Blaze. Finally, you will learn how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will have established a firm understanding of the Spark Python API and how it can be used to build data-intensive applications. Style and approach This book takes a very comprehensive, step-by-step approach so you understand how the Spark ecosystem can be used with Python to develop efficient, scalable solutions.

Every chapter is standalone and written in a very easy-to-understand manner, with a focus on both the hows and the whys of each concept.

[Spark in Action, Second Edition](#) Spark in Action

Summary Spark in Action teaches you the theory and skills you need to effectively handle batch and streaming data using Spark. Fully updated for Spark 2.0. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Big data systems distribute datasets across clusters of machines, making it a challenge to efficiently query, stream, and interpret them. Spark can help. It is a processing system designed specifically for distributed data. It provides easy-to-use interfaces, along with the performance you need for production-quality analytics and machine learning. Spark 2 also adds improved programming APIs, better performance, and countless other upgrades. About the Book Spark in Action teaches you the theory and skills you need to effectively handle batch and streaming data using Spark. You'll get comfortable with the Spark CLI as you work through a few introductory examples. Then, you'll start programming Spark using its core APIs. Along the way, you'll work with structured data using Spark SQL, process near-real-time streaming data, apply machine learning algorithms, and munge graph data using Spark GraphX. For a zero-effort startup, you can download the preconfigured virtual machine ready for you to try the book's code. What's Inside Updated for Spark 2.0 Real-life case studies Spark DevOps with Docker Examples in Scala, and online in Java and Python About the Reader Written for experienced programmers with some background in big data or machine learning. About the Authors Petar Zečević and Marko Bonačić are seasoned developers heavily involved in the Spark community. Table of Contents PART 1 - FIRST STEPS Introduction to Apache Spark Spark fundamentals Writing Spark applications The Spark API in depth PART 2 - MEET THE SPARK FAMILY Sparkling queries with Spark SQL Ingesting data with Spark Streaming Getting smart with MLlib ML: classification and clustering Connecting the dots with GraphX PART 3 - SPARK OPS Running Spark Running on a Spark standalone cluster Running on YARN and Mesos PART 4 - BRINGING IT TOGETHER Case study: real-time dashboard Deep learning on Spark with H2O

The Book of R IBM Redbooks

Working with big data can be complex and challenging, in part because of the multiple analysis frameworks and tools required. Apache Spark is a big data processing framework perfect for analyzing near-real-time streams and discovering historical patterns in batched data sets. But Spark goes much further than other frameworks. By including machine learning and graph processing capabilities, it makes many specialized data processing platforms obsolete. Spark's unified framework and programming model significantly lowers the initial infrastructure investment, and Spark's core abstractions are intuitive for most Scala, Java, and Python developers. Spark in Action teaches readers to use Spark for stream and batch data processing. It starts with an introduction to the Spark architecture and ecosystem followed by a taste of Spark's command line interface. Readers then discover the most fundamental concepts and abstractions of Spark, particularly Resilient Distributed Datasets (RDDs) and the basic data transformations that RDDs provide. The first part of the book covers writing Spark applications using the the core APIs. Readers also learn how to work with structured data using Spark SQL, how to process near-real time data with Spark Streaming, how to apply machine learning algorithms with Spark MLlib, how to apply graph algorithms on graph-shaped data using Spark GraphX, and an introduction to Spark clustering. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

High Performance Spark McGraw Hill Professional

Analytics is increasingly an integral part of day-to-day operations at today's leading businesses, and transformation is also occurring through huge growth in mobile and digital channels. Enterprise organizations are attempting to leverage analytics in new ways and transition existing analytics capabilities to respond with more flexibility while making the most efficient use of highly valuable data science skills. The recent growth and adoption of Apache Spark as an analytics framework and platform is very timely and helps meet these challenging demands. The Apache Spark environment on IBM z/OS® and Linux on IBM z Systems™ platforms allows this analytics framework to run on the same enterprise platform as the originating sources of data and transactions that feed it. If most of the data that will be used for Apache Spark analytics, or the most sensitive or quickly changing data is originating on z/OS, then an Apache Spark z/OS based environment will be the optimal choice for performance, security, and governance. This IBM® Redpaper™ publication explores the enterprise analytics market, use of Apache Spark on IBM z Systems™ platforms, integration between Apache Spark and other enterprise data sources, and case studies and examples of what can be achieved with Apache Spark in enterprise environments. It is of interest to data scientists, data engineers, enterprise architects, or anybody looking to better understand how to combine an analytics framework and platform on enterprise systems.

Spark Simon and Schuster

Summary Relevant Search demystifies relevance work. Using Elasticsearch, it teaches you how to return engaging search results to your users, helping you understand and leverage the internals of Lucene-based search engines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Users are accustomed to and expect instant, relevant search results. To achieve this, you must master the search engine. Yet for many developers, relevance ranking is mysterious or confusing. About the Book Relevant Search demystifies the subject and shows you that a search engine is a programmable relevance framework. You'll learn how to apply Elasticsearch or Solr to your business's unique ranking problems. The book demonstrates how to program relevance and how to incorporate secondary data sources, taxonomies, text analytics, and personalization. In practice, a relevance framework requires softer skills as well, such as collaborating with stakeholders to discover the right relevance requirements for your business. By the end, you'll be able to achieve a virtuous cycle of provable, measurable relevance improvements over a search product's lifetime. What's Inside Techniques for debugging relevance? Applying search engine features to real problems? Using the user interface to guide searchers? A systematic approach to relevance? A business culture focused on improving search About the Reader For developers trying to build smarter search with Elasticsearch or Solr. About the Authors Doug Turnbull is lead relevance consultant at OpenSource Connections, where he frequently speaks and blogs. John Berryman is a data engineer at Eventbrite, where he specializes in recommendations and search. Foreword author, Trey Grainger, is a director of engineering at CareerBuilder and author of Solr in Action. Table of Contents The search relevance problem Search under the hood Debugging your first relevance problem Taming tokens Basic multifield search Term-centric search Shaping the relevance function Providing relevance feedback Designing a relevance-focused search application The relevance-centered

enterprise Semantic and personalized search

Spring Batch in Action GENERAL PRESS

The New York Times–bestselling, non-nonsense guide to becoming a better leader through 7 key behaviors, based on a mix of military and corporate training. Leadership is not about job titles—it's about action and behavior. “Sparks” are the doers, thinkers, innovators, and key influencers who are catalysts for personal and organizational change. But these extraordinary individuals aren't defined by the place they hold on an organizational chart—they are defined by their actions, commitment, and will. Leadership experts Angie Morgan, Courtney Lynch, and Sean Lynch show how you can become a Spark by cultivating seven key leadership behaviors. Grounded in the latest research on leadership development, this fresh, accessible road map is packed with real-world stories from inside companies like Facebook, Google, and Boston Scientific, and from the authors' own high-stakes, challenging experiences serving in the U.S. Armed Forces. With SPARK as a blueprint, anyone can become a catalyst for change, and any organization can identify and develop Sparks. “A myth-destroying book that will make you rethink both the theory and practice of leadership.”—Daniel H. Pink, #1 New York Times–bestselling author of *Drive* “If you truly want to become a Spark in your organization and in your life, I urge you to read this book now.”—Mike “Coach K” Krzyzewski, head coach, Duke University Men's Basketball “These authors are not only great leadership thinkers, but they have all led people in challenging circumstances.... Trust them to take you to a new level.”—Brigadier General Thomas A. Kolditz, U.S. Army (Ret.), director of the Ann and John Doerr Institute for New Leaders at Rice University

Spark Packt Publishing Ltd

Design, implement, and deliver successful streaming applications, machine learning pipelines and graph applications using Spark SQL API About This Book Learn about the design and implementation of streaming applications, machine learning pipelines, deep learning, and large-scale graph processing applications using Spark SQL APIs and Scala. Learn data exploration, data munging, and how to process structured and semi-structured data using real-world datasets and gain hands-on exposure to the issues and challenges of working with noisy and "dirty" real-world data. Understand design considerations for scalability and performance in web-scale Spark application architectures. Who This Book Is For If you are a developer, engineer, or an architect and want to learn how to use Apache Spark in a web-scale project, then this is the book for you. It is assumed that you have prior knowledge of SQL querying. A basic programming knowledge with Scala, Java, R, or Python is all you need to get started with this book. What You Will Learn Familiarize yourself with Spark SQL programming, including working with DataFrame/Dataset API and SQL Perform a series of hands-on exercises with different types of data sources, including CSV, JSON, Avro, MySQL, and MongoDB Perform data quality checks, data visualization, and basic statistical analysis tasks Perform data munging tasks on publicly available datasets Learn how to use Spark SQL and Apache Kafka to build streaming applications Learn key performance-tuning tips and tricks in Spark SQL applications Learn key architectural components and patterns in large-scale Spark SQL applications In Detail In the past year, Apache Spark has been increasingly adopted for the development of distributed applications. Spark SQL APIs provide an optimized interface that helps developers build such applications quickly and easily. However, designing web-scale production applications using Spark SQL APIs can be a complex task. Hence, understanding the design and implementation best practices before you start your project will help you avoid these problems. This book gives an insight into the engineering practices used to design and build real-world, Spark-based applications. The book's hands-on examples will give you the required confidence to work on any future projects you encounter in Spark SQL. It starts by familiarizing you with data exploration and data munging tasks using Spark SQL and Scala. Extensive code examples will help you understand the methods used to implement typical use-cases for various types of applications. You will get a walkthrough of the key concepts and terms that are common to streaming, machine learning, and graph applications. You will also learn key performance-tuning details including Cost Based Optimization (Spark 2.2) in Spark SQL applications. Finally, you will move on to learning how such systems are architected and deployed for a successful delivery of your project. Style and approach This book is a hands-on guide to designing, building, and deploying Spark SQL-centric production applications at scale.

The Spark of Learning "O'Reilly Media, Inc."

A Wrinkle in Time is the winner of the 1963 Newbery Medal. It was a dark and stormy night—Meg Murry, her small brother Charles Wallace, and her mother had come down to the kitchen for a midnight snack when they were upset by the arrival of a most disturbing stranger. "Wild nights are my glory," the unearthly stranger told them. "I just got caught in a downdraft and blown off course. Let me sit down for a moment, and then I'll be on my way. Speaking of ways, by the way, there is such a thing as a tesseract." A tesseract (in case the reader doesn't know) is a wrinkle in time. To tell more would rob the reader of the enjoyment of Miss L'Engle's unusual book. *A Wrinkle in Time*, winner of the Newbery Medal in 1963, is the story of the adventures in space and time of Meg, Charles Wallace, and Calvin O'Keefe (athlete, student, and one of the most popular boys in high school). They are in search of Meg's father, a scientist who disappeared while engaged in secret work for the government on the tesseract problem.

Event Streams in Action Simon and Schuster

"How to Win Friends and Influence People" is one of the first best-selling self-help books ever published. It can enable you to make friends quickly and easily, help you to win people to your way of thinking, increase your influence, your prestige, your ability to get things done, as well as enable you to win new clients, new customers._x000D_ Twelve Things This Book Will Do For You: _x000D_ Get you out of a mental rut, give you new thoughts, new visions, new ambitions._x000D_ Enable you to make friends quickly and easily._x000D_ Increase your popularity._x000D_ Help you to win people to your way of thinking._x000D_ Increase your influence, your prestige, your ability to get things done._x000D_ Enable you to win new clients, new customers._x000D_ Increase your earning power._x000D_ Make you a better salesman, a better executive._x000D_ Help you to handle complaints, avoid arguments, keep your human contacts smooth and pleasant._x000D_ Make you a better speaker, a more entertaining conversationalist._x000D_ Make the principles of psychology easy for you to apply in your daily contacts._x000D_ Help you to arouse enthusiasm among your associates._x000D_ Dale Carnegie (1888-1955) was an American writer and lecturer and the developer of famous courses in self-improvement, salesmanship, corporate training, public speaking, and interpersonal skills. Born into poverty on a farm in Missouri, he was the author of *How to Win Friends and Influence People* (1936), a massive bestseller that remains popular today._x000D_

[Spark in Action](#) Simon and Schuster

Summary The Spark distributed data processing platform provides an easy-to-implement tool for ingesting, streaming, and processing data from any source. In Spark in Action, Second Edition, you'll learn to take advantage of Spark's core features and incredible processing speed, with applications including real-time computation, delayed evaluation, and machine learning. Spark skills are a hot commodity in enterprises worldwide, and with Spark's powerful and flexible Java APIs, you can reap all the benefits without first learning Scala or Hadoop. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Analyzing enterprise data starts by reading, filtering, and merging files and streams from many sources. The Spark data processing engine handles this varied volume like a champ, delivering speeds 100 times faster than Hadoop systems. Thanks to SQL support, an intuitive interface, and a straightforward multilanguage API, you can use Spark without learning a complex new ecosystem. About the book Spark in Action, Second Edition, teaches you to create end-to-end analytics applications. In this entirely new book, you'll learn from interesting Java-based examples, including a complete data pipeline for processing NASA satellite data. And you'll discover Java, Python, and Scala code samples hosted on GitHub that you can explore and adapt, plus appendixes that give you a cheat sheet for installing tools and understanding Spark-specific terms. What's inside Writing Spark applications in Java Spark application architecture Ingestion through files, databases, streaming, and Elasticsearch Querying distributed datasets with Spark SQL About the reader This book does not assume previous experience with Spark, Scala, or Hadoop. About the author Jean-Georges Perrin is an experienced data and software architect. He is France's first IBM Champion and has been honored for 12 consecutive years. Table of Contents PART 1 - THE THEORY CRIPPLED BY AWESOME EXAMPLES 1 So, what is Spark, anyway? 2 Architecture and flow 3 The majestic role of the dataframe 4 Fundamentally lazy 5 Building a simple app for deployment 6 Deploying your simple app PART 2 - INGESTION 7 Ingestion from files 8 Ingestion from databases 9 Advanced ingestion: finding data sources and building your own 10 Ingestion through structured streaming PART 3 - TRANSFORMING YOUR DATA 11 Working with SQL 12 Transforming your data 13 Transforming entire documents 14 Extending transformations with user-defined functions 15 Aggregating your data PART 4 - GOING FURTHER 16 Cache and checkpoint: Enhancing Spark's performances 17 Exporting data and building full data pipelines 18 Exploring deployment

[Relevant Search](#) Apress

Discover how graph algorithms can help you leverage the relationships within your data to develop more intelligent solutions and enhance your machine learning models. You'll learn how graph analytics are uniquely suited to unfold complex structures and reveal difficult-to-find patterns lurking in your data. Whether you are trying to build dynamic network models or forecast real-world behavior, this book illustrates how graph algorithms deliver value—from finding vulnerabilities and bottlenecks to detecting communities and improving machine learning predictions. This practical book walks you through hands-on examples of how to use graph algorithms in Apache Spark and Neo4j—two of the most common choices for graph analytics. Also included: sample code and tips for over 20 practical graph algorithms that cover optimal pathfinding, importance through centrality, and community detection. Learn how graph analytics vary from conventional statistical analysis Understand how classic graph algorithms work, and how they are applied Get guidance on which algorithms to use for different types of questions Explore algorithm examples with working code and sample datasets from Spark and Neo4j See how connected feature extraction can increase machine learning accuracy and precision Walk through creating an ML workflow for link prediction combining Neo4j and Spark

[Rx.NET in Action](#) Houghton Mifflin Harcourt

The Book of R is a comprehensive, beginner-friendly guide to R, the world's most popular programming language for statistical analysis. Even if you have no programming experience and little more than a grounding in the basics of mathematics, you'll find everything you need to begin using R effectively for statistical analysis. You'll start with the basics, like how to handle data and write simple programs, before moving on to more advanced topics, like producing statistical summaries of your data and performing statistical tests and modeling. You'll even learn how to create impressive data visualizations with R's basic graphics tools and contributed packages, like ggplot2 and ggvis, as well as interactive 3D visualizations using the rgl package. Dozens of hands-on exercises (with downloadable solutions) take you from theory to practice, as you learn: -The fundamentals of programming in R, including how to write data frames, create functions, and use variables, statements, and loops -Statistical concepts like exploratory data analysis, probabilities, hypothesis tests, and regression modeling, and how to execute them in R -How to access R's thousands of functions, libraries, and data sets -How to draw valid and useful conclusions from your data -How to create publication-quality graphics of your results Combining detailed explanations with real-world examples and exercises, this book will provide you with a solid understanding of both statistics and the depth of R's functionality. Make The Book of R your doorway into the growing world of data analysis.

The Big Book of Conflict Resolution Games: Quick, Effective Activities to Improve Communication, Trust and Collaboration "O'Reilly Media, Inc."

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.

[Learning Spark SQL](#) Simon and Schuster

Summary Node.js in Action, Second Edition is a thoroughly revised book based on the best-selling first edition. It starts at square one and guides you through all the features, techniques, and concepts you'll need to build production-quality Node applications. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology You already know JavaScript. The trick to mastering Node.js is learning how to build applications that fully exploit its powerful asynchronous event handling and non-blocking I/O features. The Node server radically simplifies event-driven real-time apps like chat, games, and live data analytics, and with its incredibly rich ecosystem of modules, tools, and libraries, it's hard to beat! About the Book Based on the bestselling first edition, Node.js in Action, Second Edition is a completely new book. Packed with practical examples, it teaches you how to create high-performance web servers using JavaScript and Node. You'll master key design concepts such as asynchronous programming, state management, and event-driven programming. And you'll learn to put together MVC servers using Express

and Connect, design web APIs, and set up the perfect production environment to build, lint, and test. What's Inside Mastering non-blocking I/O The Node event loop Testing and deploying Web application templating About the Reader Written for web developers with intermediate JavaScript skills. About the Authors The Second Edition author team includes Node masters Alex Young, Bradley Meck, Mike Cantelon, and Tim Oxley, along with original authors Marc Harter, T.J. Holowaychuk, and Nathan Rajlich. Table of contents PART 1 - WELCOME TO NODE Welcome to Node.js Node programming fundamentals What is a Node web application? PART 2 - WEB DEVELOPMENT WITH NODE Front-end build systems Server-side frameworks Connect and Express in depth Web application templating Storing application data Testing Node applications Deploying Node applications and maintaining uptime PART 3 - BEYOND WEB DEVELOPMENT Writing command-line applications Conquering the desktop with Electron **Jazz** "O'Reilly Media, Inc."

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such as analyzing changes in word frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework. This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g. histogram, correlation) will help the reader appreciate the more advanced data processing examples. 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.

[A Wrinkle in Time](#) Packt Publishing Ltd

Develop the Art of Self-Challenge Tired of dreaming about changing your life? Ready to change it? You came to the right place. Jason Jaggard has developed a simple, direct way to practice the art of self-challenge. It requires a willingness to take healthy risks and is energized by the involvement of a small group of fellow risk-takers. Here are just a few of the Spark insights that lead to lasting change: · More resources are available to you than you realize. · Living a changed life works best when a few friends get involved. · When you relearn what “good” is, you experience more goodness in your life. · Taking a healthy risk is always worth it, even if you fail. When you practice the art of self-challenge, you notice that your life and the world around you start changing for the better. It involves doing the one thing that only you can do: take responsibility for your life. So go ahead and take your first risk: open this book and start reading.

[Mastering Spark with R](#) Simon and Schuster

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

[How To Win Friends And Influence People](#) Apress

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 applications 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.

Spark Cookbook "O'Reilly Media, Inc."

Understand the complexities of modern-day data engineering platforms and explore strategies to deal with them with the help of use case scenarios led by an industry expert in big data Key Features Become well-versed with the core concepts of Apache Spark and Delta Lake for building data platforms Learn how to ingest, process, and analyze data that can be later used for training machine learning models Understand how to operationalize data models in production using curated data Book Description In the world of ever-changing data and schemas, it is important to build data pipelines that can auto-adjust to changes. This book will help you build scalable data platforms that managers, data scientists, and data analysts can rely on. Starting with an introduction to data engineering, along with its key concepts and architectures, this book will show you how to use Microsoft Azure Cloud services effectively for data engineering. You'll cover data lake design patterns and the different stages through which the data needs to flow in a typical data lake. Once you've explored the main features of Delta Lake to build data lakes with fast performance and governance in mind, you'll advance to implementing the lambda architecture using Delta Lake. Packed with practical examples and code snippets, this book takes

you through real-world examples based on production scenarios faced by the author in his 10 years of experience working with big data. Finally, you'll cover data lake deployment strategies that play an important role in provisioning the cloud resources and deploying the data pipelines in a repeatable and continuous way. By the end of this data engineering book, you'll know how to effectively deal with ever-changing data and create scalable data pipelines to streamline data science, ML, and artificial intelligence (AI) tasks. What you will learn Discover the challenges you may face in the data engineering world Add ACID transactions to Apache Spark using Delta Lake Understand effective design strategies to build enterprise-grade data

lakes Explore architectural and design patterns for building efficient data ingestion pipelines Orchestrate a data pipeline for preprocessing data using Apache Spark and Delta Lake APIs Automate deployment and monitoring of data pipelines in production Get to grips with securing, monitoring, and managing data pipelines models efficiently Who this book is for This book is for aspiring data engineers and data analysts who are new to the world of data engineering and are looking for a practical guide to building scalable data platforms. If you already work with PySpark and want to use Delta Lake for data engineering, you'll find this book useful. Basic knowledge of Python, Spark, and SQL is expected.