

---

# Spark In Action

---

Data Algorithms  
 Hadoop in Action  
 Learning Apache Spark 2  
 Big Data Processing with Apache Spark  
 Spark and the League of Ursus  
 How to Lead Yourself and Others to Greater Success  
 Scala in Action  
 Mastering Structured Streaming and Spark Streaming  
 Quickly learn the art of writing efficient big data applications with Apache Spark  
 A Novel  
 A Novel  
 A Novel  
 Big Data Processing with Apache Spark  
 Data Analytics with Hadoop  
 Frank Kane's Taming Big Data with Apache Spark and Python  
 With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library  
 Lightning-Fast Big Data Analysis  
 Apache Spark Quick Start Guide  
 Spark: The Definitive Guide  
 Practical Apache Spark  
 Build and deploy distributed deep learning applications on Apache Spark  
 Fahrenheit 451  
 The Complete Guide to Large-Scale Analysis and Modeling  
 Learning PySpark  
 Spark in Action  
 Patterns for Learning from Data at Scale  
 Big Data Processing Made Simple  
 Mastering Apache Spark 2.x  
 Each Tiny Spark  
 From Equity Talk to Equity Walk  
 Instant Action Steps to Spark Your Business  
 How the Science of Acupuncture Explains the Mysteries of Western Medicine  
 Spark  
 Covers Apache Spark 3 with Examples in Java, Python, and Scala  
 Apache Spark in 24 Hours, Sams Teach Yourself  
 Learning Spark  
 The Spark in the Machine  
 Best Practices for Scaling and Optimizing Apache Spark  
 Stream Processing with Apache Spark

*Spark In Action*

Downloaded from [business.itu.edu.tr](https://business.itu.edu.tr)  
 guest

---

## CODY PAOLA

---

*Data Algorithms* Lulu.com

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

[Hadoop in Action](#) Packt Publishing Ltd

An army of toys, a menacing threat, and a thrilling adventure collide in the high-stakes sequel to *Spark and the League of Ursus*. Spark may be a cute and cuddly teddy bear, but she's also a fierce protector. Weeks after rescuing her human owner—a budding young filmmaker named Loretta—from a hideous monster, everything seems to be returning to normal. But then Spark is summoned before the mysterious Grand Sleuth, the high council of teddy bears, who task her with a dangerous mission: locating the portal to the monster's world. During her daring quest, Spark discovers a terrible secret that changes everything. In order to keep Loretta and their whole town safe, she must enlist the help of her loyal toy friends and team up with an unexpected ally. As the menace grows, Spark realizes that Loretta has a hidden power that may be the key to saving them all . . . This dark middle grade fantasy is perfect for fans of the *Nightmares!* series and Holly Black's *Doll Bones*.

[Learning Apache Spark 2](#) Puffin Books

*Spark in Action, Second Edition* Covers Apache Spark 3 with Examples in Java, Python, and Scala Manning Publications

[Big Data Processing with Apache Spark](#) Apress

Unleash the data processing and analytics capability of Apache Spark with the language of choice: Java About This Book Perform big data processing with Spark—without having to learn Scala! Use the Spark Java API to implement efficient enterprise-grade applications for data processing and analytics Go beyond mainstream data processing by adding querying capability, Machine Learning, and graph processing using Spark Who This Book Is For If you are a Java developer interested in learning to use the popular Apache Spark framework, this book is the resource you need to get started. Apache Spark developers who are looking to build enterprise-grade applications in Java will also find this book very useful. What You Will Learn Process data using different file formats such as XML, JSON, CSV, and plain and delimited text, using the Spark core Library. Perform analytics on data from various data sources such as Kafka, and Flume using Spark Streaming Library Learn SQL schema creation and the analysis of structured data using various SQL functions including Windowing functions in the Spark SQL Library Explore Spark MLib APIs while implementing Machine Learning techniques to solve real-world problems Get to know Spark GraphX so you understand various graph-based analytics that can be performed with Spark In Detail Apache Spark is the buzzword in the big data industry right now, especially with the increasing need for real-time streaming and data processing. While Spark is built on Scala, the Spark Java API exposes all the Spark features available in the Scala version for Java developers. This book will show you how you can implement various functionalities of the Apache Spark framework in Java, without stepping out of your comfort zone. The book starts with an introduction to the Apache Spark 2.x ecosystem, followed by explaining how to install and configure Spark, and refreshes the Java concepts that will be useful to you when consuming Apache Spark's APIs. You will explore RDD and its associated common Action and Transformation Java APIs, set up a production-like clustered environment, and work with Spark SQL. Moving on, you will perform near-real-time processing with Spark streaming, Machine Learning analytics with Spark MLib, and graph processing with GraphX, all using various Java packages. By the end of the book,

you will have a solid foundation in implementing components in the Spark framework in Java to build fast, real-time applications. Style and approach This practical guide teaches readers the fundamentals of the Apache Spark framework and how to implement components using the Java language. It is a unique blend of theory and practical examples, and is written in a way that will gradually build your knowledge of Apache Spark.

**Spark and the League of Ursus** Apress

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 MLib, 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 MLib to a variety of problems, including classification or recommendation

*How to Lead Yourself and Others to Greater Success* "O'Reilly Media, Inc."

If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You'll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA Social network analysis (recommendation systems, counting triangles, sentiment analysis)

*Scala in Action* Sams Publishing

Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLib, and R on Spark with the help of practical code snippets for each topic. Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You'll follow a learn-to-do-by-yourself approach to learning – learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you'll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You'll also become

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

### **Mastering Structured Streaming and Spark Streaming**

Packt Publishing Ltd

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

### **Quickly learn the art of writing efficient big data**

**applications with Apache Spark** O'Reilly Media

Advanced analytics on your Big Data with latest Apache Spark 2.x About This Book An advanced guide with a combination of instructions and practical examples to extend the most up-to-date Spark functionalities. Extend your data processing capabilities to process huge chunk of data in minimum time using advanced concepts in Spark. Master the art of real-time processing with the help of Apache Spark 2.x Who This Book Is For If you are a developer with some experience with Spark and want to strengthen your knowledge of how to get around in the world of Spark, then this book is ideal for you. Basic knowledge of Linux, Hadoop and Spark is assumed. Reasonable knowledge of Scala is expected. What You Will Learn Examine Advanced Machine Learning and DeepLearning with MLlib, SparkML, SystemML, H2O and DeepLearning4J Study highly optimised unified batch and real-time data processing using SparkSQL and Structured Streaming Evaluate large-scale Graph Processing and Analysis using GraphX and GraphFrames Apply Apache Spark in Elastic deployments using Jupyter and Zeppelin Notebooks, Docker, Kubernetes and the IBM Cloud Understand internal details of cost based optimizers used in Catalyst, SystemML and GraphFrames Learn how specific parameter settings affect overall performance of an Apache Spark cluster Leverage Scala, R and python for your data science projects In Detail Apache Spark is an in-memory cluster-based parallel processing system that provides a wide range of functionalities such as graph processing, machine learning, stream processing, and SQL. This book aims to take your knowledge of Spark to the next level by teaching you how to expand Spark's functionality and implement your data flows and machine/deep learning programs on top of the platform. The book commences with an overview of the Spark ecosystem. It will introduce you to Project Tungsten and Catalyst, two of the major advancements of Apache Spark 2.x. You will understand how

memory management and binary processing, cache-aware computation, and code generation are used to speed things up dramatically. The book extends to show how to incorporate H2O, SystemML, and Deeplearning4j for machine learning, and Jupyter Notebooks and Kubernetes/Docker for cloud-based Spark. During the course of the book, you will learn about the latest enhancements to Apache Spark 2.x, such as interactive querying of live data and unifying DataFrames and Datasets. You will also learn about the updates on the APIs and how DataFrames and Datasets affect SQL, machine learning, graph processing, and streaming. You will learn to use Spark as a big data operating system, understand how to implement advanced analytics on the new APIs, and explore how easy it is to use Spark in day-to-day tasks. Style and approach This book is an extensive guide to Apache Spark modules and tools and shows how Spark's functionality can be extended for real-time processing and storage with worked examples.

**A Novel** "O'Reilly Media, Inc."

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

**A Novel** "O'Reilly Media, Inc."

Toy Story meets Stranger Things in this epic tale of warrior teddy bears and the children they protect. Spark is not your average teddy bear. She's soft and cuddly, sure, but she's also a fierce warrior. At night she fulfills her sacred duty: to protect the household from monsters. But Spark's owner Loretta is growing up and thinks she doesn't need her old teddy anymore. When a monster unlike any other descends on the quiet home, everything changes. Children are going missing, and the monster wants Loretta next. Only Spark can stop it. She must call upon the ancient League of Ursus—a secret alliance of teddy bears who are pledged to protect their human friends. Together with an Amazon-princess doll and a timid sock monkey, the bears are all that stands between our world and the one that lies beneath. It will be a heroic chapter in the history of the League . . . if the bears live to tell the tale.

**A Novel** John Wiley & Sons

Ready to use statistical and machine-learning techniques across large data sets? This practical guide shows you why the Hadoop ecosystem is perfect for the job. Instead of deployment, operations, or software development usually associated with distributed computing, you'll focus on particular analyses you can build, the data warehousing techniques that Hadoop provides, and higher order data workflows this framework can produce. Data scientists and analysts will learn how to perform a wide range of techniques, from writing MapReduce and Spark applications with Python to using advanced modeling and data management with Spark MLlib, Hive, and HBase. You'll also learn

about the analytical processes and data systems available to build and empower data products that can handle—and actually require—huge amounts of data. Understand core concepts behind Hadoop and cluster computing Use design patterns and parallel analytical algorithms to create distributed data analysis jobs Learn about data management, mining, and warehousing in a distributed context using Apache Hive and HBase Use Sqoop and Apache Flume to ingest data from relational databases Program complex Hadoop and Spark applications with Apache Pig and Spark DataFrames Perform machine learning techniques such as classification, clustering, and collaborative filtering with Spark's MLlib

**Big Data Processing with Apache Spark** Spark in Action, Second Edition Covers Apache Spark 3 with Examples in Java, Python, and Scala

No need to spend hours ploughing through endless data - let Spark, one of the fastest big data processing engines available, do the hard work for you. Key Features Get up and running with Apache Spark and Python Integrate Spark with AWS for real-time analytics Apply processed data streams to machine learning APIs of Apache Spark Book Description Processing big data in real time is challenging due to scalability, information consistency, and fault-tolerance. This book teaches you how to use Spark to make your overall analytical workflow faster and more efficient. You'll explore all core concepts and tools within the Spark ecosystem, such as Spark Streaming, the Spark Streaming API, machine learning extension, and structured streaming. You'll begin by learning data processing fundamentals using Resilient Distributed Datasets (RDDs), SQL, Datasets, and Dataframes APIs. After grasping these fundamentals, you'll move on to using Spark Streaming APIs to consume data in real time from TCP sockets, and integrate Amazon Web Services (AWS) for stream consumption. By the end of this book, you'll not only have understood how to use machine learning extensions and structured streams but you'll also be able to apply Spark in your own upcoming big data projects. What you will learn Write your own Python programs that can interact with Spark Implement data stream consumption using Apache Spark Recognize common operations in Spark to process known data streams Integrate Spark streaming with Amazon Web Services (AWS) Create a collaborative filtering model with the movielens dataset Apply processed data streams to Spark machine learning APIs Who this book is for Data Processing with Apache Spark is for you if you are a software engineer, architect, or IT professional who wants to explore distributed systems and big data analytics. Although you don't need any knowledge of Spark, prior experience of working with Python is recommended.

[Data Analytics with Hadoop](#) Packt Publishing Ltd

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.

[Frank Kane's Taming Big Data with Apache Spark and Python](#) Simon and Schuster

Before you can build analytics tools to gain quick insights, you first need to know how to process data in real time. With this practical guide, developers familiar with Apache Spark will learn how to put this in-memory framework to use for streaming data. You'll discover how Spark enables you to write streaming jobs in almost the same way you write batch jobs. Authors Gerard Maas and François Garillot help you explore the theoretical underpinnings of Apache Spark. This comprehensive guide features two sections that compare and contrast the streaming APIs Spark now supports: the original Spark Streaming library and the newer Structured Streaming API. Learn fundamental stream processing concepts and examine different streaming architectures Explore Structured Streaming through practical examples; learn different aspects of stream processing in detail Create and operate streaming jobs and applications with Spark Streaming; integrate Spark Streaming with other Spark APIs Learn advanced Spark Streaming techniques, including approximation algorithms and machine learning algorithms Compare Apache Spark to other stream processing projects, including Apache Storm, Apache Flink, and Apache Kafka Streams [With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library](#) Houghton Mifflin Harcourt

Life is created by a series of events. If you can find some way to pull 20 extra minutes out of your life, you'd better use them wisely. Where will you discover this time, and when you find this time, how will you use it? Just 20 minutes a day will convert into a powerful 121 hours a year. That is a huge piece of time that you can use to gain new capacities and expand. The advantage is seen when we make a day by day decision and submit those 20 minutes to something beneficial. You can awaken 20 minutes before, take 20 minutes from lunch, after work, or just before bed - the decision is totally up to you. Just do it! Remove all excuses and distractions. Imagine if you just use 20 minutes a day to create a spark for your business. That small spark can turn into a major flame! How do you create this spark for your business? It's through promotion. Promotion isn't something that ought to be done randomly, just when you think you have sufficient energy to do it. Honestly, as an entrepreneur who's always busy, you're never going to have room schedule-wise. Consequently, you need to make time for what's important. This implies doing something intentional each and every day. Before you think you don't have room schedule-wise to do that, reconsider. Showcasing doesn't need to be hours of your day. Instead, it's something you can do in as little time as 20 minutes a day.

[Lightning-Fast Big Data Analysis](#) Singing Dragon

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

*Apache Spark Quick Start Guide* "O'Reilly Media, Inc."

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

Best Sellers - Books :

- [The Creative Act: A Way Of Being](#)
- [Flash Cards: Sight Words By Scholastic Teacher Resources](#)
- [The Light We Carry: Overcoming In Uncertain Times](#)
- [The Untethered Soul: The Journey Beyond Yourself By Michael A. Singer](#)
- [The Alchemist, 25th Anniversary: A Fable About Following Your Dream](#)
- [Young Forever: The Secrets To Living Your Longest, Healthiest Life \(the Dr. Hyman Library, 11\)](#)
- [Outlive: The Science And Art Of Longevity](#)
- [Things We Hide From The Light \(knockemout Series, 2\)](#)

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 Cassandra) • 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.

[Spark: The Definitive Guide](#) 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

[Practical Apache Spark](#) Simon and Schuster

Named one of the best 25 space opera books by BookRiot! The first book in a scifi retelling of the Mahabrahata. When Esmee wins a contest of skill, she sets off events that trigger an inevitable and unwinnable war that pits her against the family she would give anything to return to. In a universe of capricious gods, dark moons, and kingdoms built on the backs of spaceships, a cursed queen sends her infant daughter away, a jealous uncle steals the throne of Kali from his nephew, and an exiled prince vows to take his crown back. Raised alone and far away from her home on Kali, Esmee longs to return to her family. When the King of Wychstar offers to gift the unbeatable, sentient warship Titania to a warrior that can win his competition, she sees her way home: she'll enter the competition, reveal her true identity to the world, and help her famous brother win back the crown of Kali. It's a great plan. Until it falls apart. Inspired by the Mahabharata and other ancient Indian stories, *A Spark of White Fire* is a lush, sweeping space opera about family, curses, and the endless battle between jealousy and love.

- [The Complete Summer I Turned Pretty Trilogy \(boxed Set\): The Summer I Turned Pretty; It's Not Summer Without You; We'll Always Have Summer By Jenny Han](#)
- [Twisted Hate \(twisted, 3\)](#)