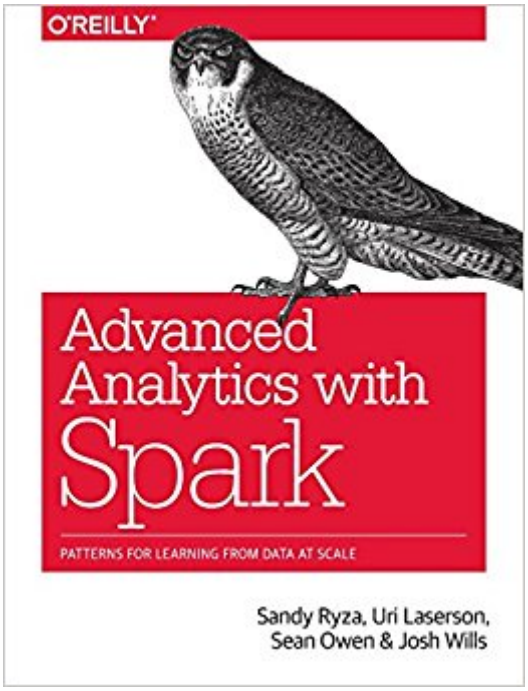
Advanced Analytics With Spark: Patterns For Learning From Data At Scale



Synopsis

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.Youâ ™ll start with an introduction to Spark and its ecosystem, and then dive into patterns that apply common techniquesâ ”classification, collaborative filtering, and anomaly detection among othersâ ”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, youâ ™ll find these patterns useful for working on your own data applications.Patterns include:Recommending music and the Audioscrobbler data setPredicting forest cover with decision treesAnomaly detection in network traffic with K-means clusteringUnderstanding Wikipedia with Latent Semantic AnalysisAnalyzing co-occurrence networks with GraphXGeospatial and temporal data analysis on the New York City Taxi Trips dataEstimating financial risk through Monte Carlo simulationAnalyzing genomics data and the BDG projectAnalyzing neuroimaging data with PySpark and Thunder

Book Information

Paperback: 276 pages

Publisher: O’Reilly Media; 1 edition (April 20, 2015)

Language: English

ISBN-10: 1491912766

ISBN-13: 978-1491912768

Product Dimensions: 7 x 0.6 x 9.2 inches

Shipping Weight: 14.4 ounces (View shipping rates and policies)

Average Customer Review: 4.6 out of 5 starsÂ Â See all reviewsÂ (21 customer reviews)

Best Sellers Rank: #19,153 in Books (See Top 100 in Books) #1 inÂ Books > Computers & Technology > Web Development & Design > Website Analytics #13 inÂ Books > Computers & Technology > Programming > Languages & Tools > Java #13 inÂ Books > Textbooks > Computer Science > Database Storage & Design

Customer Reviews

TL;DR If you are looking for a intro to data science, data analysis and machine learning at scale - this is the right book. Sure, there are others, maybe more popular books from O’Reilly considering these topics, but the authors of those are using R and Python and the books are not focused on the performance and scalability. For closer details regarding Spark you can also take a look at this

introductory Spark book - Learning Spark.This book presents 9 case studies of data analysis applications in various domains. The topics are diverse and the authors always use real world datasets. Beside learning Spark and a data science you will also have the opportunity to gain insight about topics like taxi traffic in NYC, deforestation or neuroscience. Without any previous exposure or contact with machine learning readers might struggle to understand certain chapters, so I think it's good idea to actually try those examples yourself while reading and Google for further details about the used methods. Many of the chapters end only with basic models, which barely outperform the baselines, so if you want to, there is a lot of space for their improvement and further work.Spark itself provides it's users with APIs in three languages - Java, Scala and Python. This books successfully covers each one of these, although you can feel slight preference of a Scala throughout the book. For Scala starters - they always explain some of the special constructs or syntax features which is in fact a nice thing. Introduction and Appendix chapters provides basic information about the Spark core, RDDs (Resilient distributed datasets) or options of running Spark - whether in cluster (Mesos, YARN, Spark's own) or standalone settings.

Neighborhoods: Creating Small-Scale Community in a Large-Scale World Scale Studies for Viola: Based on the Hrimaly Scale Studies for the Violin Rand McNally 2017 Large Scale Road Atlas (Rand Mcnally Large Scale Road Atlas USA) L590 - Progressive Scale Studies - Scale Study and Practical Theory in Major and Minor Keys for the Young Violinist