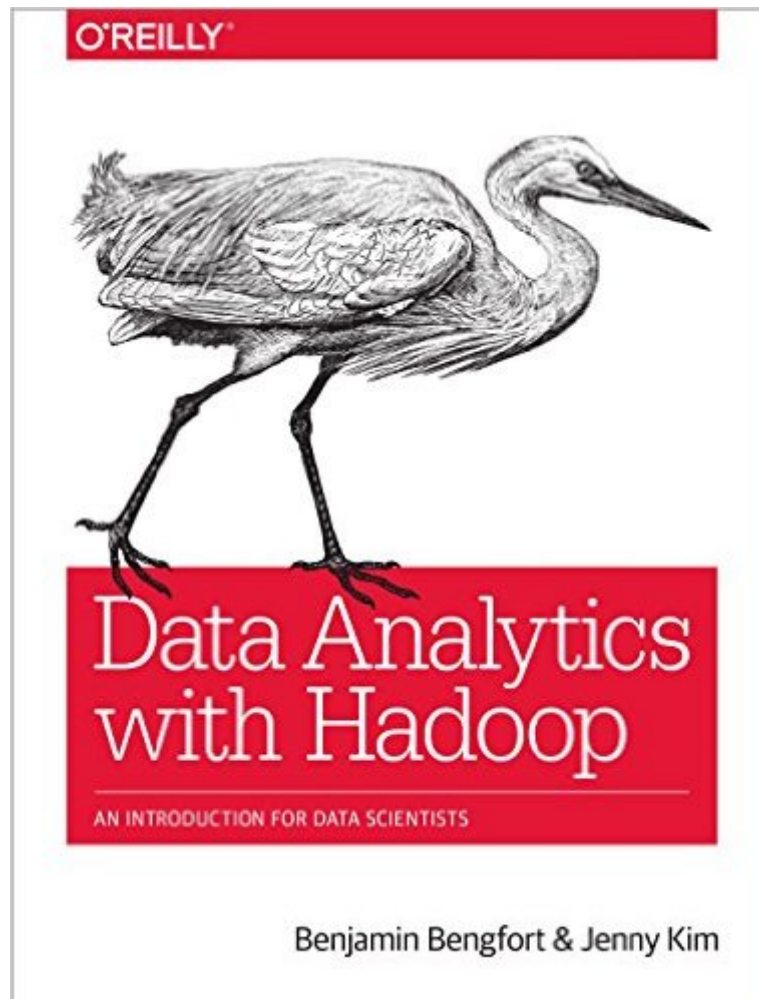


The book was found

Data Analytics With Hadoop: An Introduction For Data Scientists



Synopsis

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.

Book Information

Paperback: 288 pages

Publisher: O'Reilly Media; 1 edition (June 18, 2016)

Language: English

ISBN-10: 1491913703

ISBN-13: 978-1491913703

Product Dimensions: 7 x 0.5 x 9.1 inches

Shipping Weight: 1 pounds (View shipping rates and policies)

Average Customer Review: 5.0 out of 5 stars — See all reviews (1 customer review)

Best Sellers Rank: #42,188 in Books (See Top 100 in Books) #8 in Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Structured Design #22 in Books > Computers & Technology > Databases & Big Data > Data Mining #29 in Books > Computers & Technology > Databases & Big Data > Data Modeling & Design

Customer Reviews

I really like this book. It is a great overview of a plethora of topics around doing scalable data analytics and data science. It is extremely up-to date, going through techniques that have existed

for many years now like MapReduce, but also newer systems like Spark, all in the context of the Hadoop eco-system. They go into machine learning techniques, data management, and overall paint a nice picture around what data science is, and why data products are important, while teaching you how to make them! Every single concept is explained in a clear and concise manner, and wherever details are omitted there is always a citation to a source where the reader can continue reading more about it, which I think is great. Although I wouldn't classify myself as a beginner, I believe it is friendly to both professionals and beginners, as it is centered around python which makes most examples (that are conveniently uploaded in a nice github repository) really easy to simply run and play around with. After describing something, whether that would be a technique for data analysis, or just the in-and out workings of some analysis platform like HBase, Hive etc, the authors provide examples so that while you're reading about this stuff you can also run it, play around with it and really explore how these systems function; I believe this is a crucial part of familiarizing ones self with new platforms. Another thing I enjoyed a lot was the ending of this book. After you really dive into all of these systems and get your feet wet with each one of them, the authors wrap it all up in a nice bow by taking a step back and describing the entire end-to-end process of how you would go about productively using the knowledge you've gotten from this book to build data analytics workflows! I highly recommend this to anyone who both knows that they want to learn how to deploy scalable analytics workflows in 2016, but also to readers who are simply just curious about data science; this book will suck you in!

[Download to continue reading...](#)

Data Analytics with Hadoop: An Introduction for Data Scientists Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business. Leveraging the Power of Data Analytics, Data ... (Hacking Freedom and Data Driven) (Volume 2) Agile Data Science: Building Data Analytics Applications with Hadoop Big Data Analytics with R and Hadoop Analytics: Data Science, Data Analysis and Predictive Analytics for Business MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Practical Hive: A Guide to Hadoop's Data Warehouse System The Octopus Scientists (Scientists in the Field Series) The Bat Scientists (Scientists in the Field Series) Physics for Scientists and Engineers with Modern Physics: Volume II (3rd Edition) (Physics for Scientists &

Engineers) Physics for Scientists and Engineers, Vol. 1: Mechanics, Oscillations and Waves, Thermodynamics (Physics for Scientists & Engineers, Chapters 1-21) Even You Can Learn Statistics and Analytics: An Easy to Understand Guide to Statistics and Analytics (3rd Edition) People Analytics: How Social Sensing Technology Will Transform Business and What It Tells Us about the Future of Work (FT Press Analytics) Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results Healthcare Data Analytics (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) From Big Data to Big Profits: Success with Data and Analytics RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)

[Dmca](#)