



Data Analytics with Hadoop: An Introduction for Data Scientists

By Benjamin Bengfort, Jenny Kim

 Download

 Read Online

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim

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

 [Download Data Analytics with Hadoop: An Introduction for Da ...pdf](#)

 [Read Online Data Analytics with Hadoop: An Introduction for ...pdf](#)

Data Analytics with Hadoop: An Introduction for Data Scientists

By Benjamin Bengfort, Jenny Kim

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim

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

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim **Bibliography**

- Sales Rank: #92934 in Books
- Brand: imusti
- Published on: 2016-06-18
- Original language: English
- Number of items: 1
- Dimensions: 9.10" h x .50" w x 7.00" l, .0 pounds
- Binding: Paperback
- 288 pages

 [Download Data Analytics with Hadoop: An Introduction for Da ...pdf](#)

 [Read Online Data Analytics with Hadoop: An Introduction for ...pdf](#)



Download and Read Free Online Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim

Editorial Review

About the Author

Benjamin Bengfort is a Data Scientist who lives inside the beltway but ignores politics (the normal business of DC) favoring technology instead. He is currently working to finish his PhD at the University of Maryland where he studies machine learning and distributed computing. His lab does have robots (though this field of study is not one he favors) and, much to his chagrin, they seem to constantly arm said robots with knives and tools; presumably to pursue culinary accolades. Having seen a robot attempt to slice a tomato, Benjamin prefers his own adventures in the kitchen where he specializes in fusion French and Guyanese cuisine as well as BBQ of all types. A professional programmer by trade, a Data Scientist by vocation, Benjamin's writing pursues a diverse range of subjects from Natural Language Processing, to Data Science with Python to analytics with Hadoop and Spark.

Jenny Kim is an experienced big data engineer who works in both commercial software efforts as well as in academia. She has significant experience in working with large scale data, machine learning, and Hadoop implementations in production and research environments. Jenny (with Benjamin Bengfort) previously built a large scale recommender system that used a web crawler to gather ontological information about apparel products and produce recommendations from transactions. Currently, she is working with the Hue team at Cloudera, to help build intuitive interfaces for analyzing big data with Hadoop.

Users Review

From reader reviews:

Gary Glover:

This book untitled Data Analytics with Hadoop: An Introduction for Data Scientists to be one of several books that will best seller in this year, here is because when you read this guide you can get a lot of benefit into it. You will easily to buy this book in the book retailer or you can order it through online. The publisher in this book sells the e-book too. It makes you more readily to read this book, as you can read this book in your Cell phone. So there is no reason for you to past this book from your list.

Amelia Brown:

This Data Analytics with Hadoop: An Introduction for Data Scientists is great guide for you because the content that is certainly full of information for you who have always deal with world and get to make decision every minute. This book reveal it info accurately using great plan word or we can state no rambling sentences in it. So if you are read it hurriedly you can have whole details in it. Doesn't mean it only will give you straight forward sentences but challenging core information with lovely delivering sentences. Having Data Analytics with Hadoop: An Introduction for Data Scientists in your hand like obtaining the world in your arm, facts in it is not ridiculous 1. We can say that no book that offer you world throughout ten or

fifteen moment right but this reserve already do that. So , this really is good reading book. Hey Mr. and Mrs. hectic do you still doubt that?

Richard Daniels:

Don't be worry when you are afraid that this book will probably filled the space in your house, you might have it in e-book method, more simple and reachable. That Data Analytics with Hadoop: An Introduction for Data Scientists can give you a lot of pals because by you investigating this one book you have thing that they don't and make anyone more like an interesting person. This specific book can be one of a step for you to get success. This guide offer you information that possibly your friend doesn't realize, by knowing more than some other make you to be great people. So , why hesitate? Let me have Data Analytics with Hadoop: An Introduction for Data Scientists.

Caitlin Cruz:

As a college student exactly feel bored for you to reading. If their teacher requested them to go to the library in order to make summary for some guide, they are complained. Just small students that has reading's heart and soul or real their leisure activity. They just do what the professor want, like asked to go to the library. They go to generally there but nothing reading very seriously. Any students feel that reading through is not important, boring and can't see colorful pics on there. Yeah, it is for being complicated. Book is very important for you personally. As we know that on this era, many ways to get whatever we would like. Likewise word says, ways to reach Chinese's country. Therefore this Data Analytics with Hadoop: An Introduction for Data Scientists can make you really feel more interested to read.

Download and Read Online Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim #KXTL2YZO4EW

Read Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim for online ebook

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim Free PDF d0wnl0ad, audio books, books to read, good books to read, cheap books, good books, online books, books online, book reviews epub, read books online, books to read online, online library, greatbooks to read, PDF best books to read, top books to read Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim books to read online.

Online Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim ebook PDF download

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim Doc

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim Mobipocket

Data Analytics with Hadoop: An Introduction for Data Scientists By Benjamin Bengfort, Jenny Kim EPub