The book was found

# Hadoop: The Definitive Guide: Storage And Analysis At Internet Scale

O'REILLY®

4th Edition
Revised & Updated

# Hadoop
## The Definitive Guide

STORAGE AND ANALYSIS AT INTERNET SCALE

Tom White

**DOWNLOAD EBOOK**
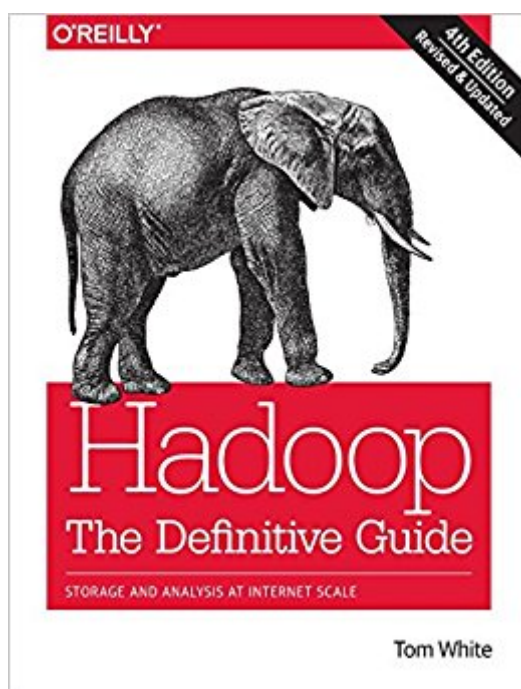
# Synopsis

Get ready to unlock the power of your data. With the fourth edition of this comprehensive guide, youâ ™ll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters.Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. Youâ ™ll learn about recent changes to Hadoop, and explore new case studies on Hadoopâ ™s role in healthcare systems and genomics data processing.Learn fundamental components such as MapReduce, HDFS, and YARNExplore MapReduce in depth, including steps for developing applications with itSet up and maintain a Hadoop cluster running HDFS and MapReduce on YARNLearn two data formats: Avro for data serialization and Parquet for nested dataUse data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer)Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with HadoopLearn the HBase distributed database and the ZooKeeper distributed configuration service

# Book Information

Paperback: 756 pages

Publisher: O'Reilly Media; 4 edition (April 11, 2015)

Language: English

ISBN-10: 1491901632

ISBN-13: 978-1491901632

Product Dimensions:  6.9 x 1.6 x 9.2 inches

Shipping Weight: 2.4 pounds (View shipping rates and policies)

Average Customer Review:    4.3 out of 5 stars      159 customer reviews

Best Sellers Rank: #14,957 in Books (See Top 100 in Books)   #1 inÂ Books > Computers & Technology > Programming > Parallel Programming   #12 inÂ Books > Computers & Technology > Databases & Big Data > Data Mining   #12 inÂ Books > Computers & Technology > Programming > Languages & Tools > Java

# Customer Reviews

Tom White has been an Apache Hadoop committer since February 2007, and is a member of the Apache Software Foundation. He works for Cloudera, a company set up to offer Hadoop support and training. Previously he was as an independent Hadoop consultant, working with companies to

set up, use, and extend Hadoop. He has written numerous articles for O'Reilly, java.net and IBM's developerWorks, and has spoken at several conferences, including at ApacheCon 2008 on Hadoop. Tom has a Bachelor's degree in Mathematics from the University of Cambridge and a Master's in Philosophy of Science from the University of Leeds, UK.

Nice update for an already great book. As other reviews have already mentioned, if you're new to Hadoop, you'll probably need to read a few chapters in order to really get a good picture of what Hadoop is, and how it works. This is not necessarily a problem with the book, but just the nature of the beast when learning as complex as Hadoop. If you're new to Hadoop, take the time to work through the first few chapters, and maybe expect to come back and re-read parts later. If you're an experienced engineer who is looking to get up-to-speed with what has changed in recent versions, you'll definitely find what you're looking for here.I appreciate that this book covers high-level concepts as well as dives deep into the technical details that you will need to know for the design, implementation and day-to-day running of Hadoop and its various associated technologies.

Good overview of the Hadoop ecosystem. Of course, if you need to dive deeper, you will have to pick the component of interest for you and get additional material. This updated version covers the latest and greatest tools in the ecosystem, so, at least for me it was a good investment.

I bought this book for a project at work, to prototype a log analysis system using Hadoop. I haven't bought very many technical books in the last few years, but the quality of most online documentation for Hadoop is poor and books seemed like a better option. This book is considered the "bible" for Hadoop. It was useful, and I kept it open on my desk for quite a while as I worked to get the infrastructure set up. Consider it a high-level intro to lots of different Hadoop topics, and you'll be happy with it. Just don't expect it to answer all of your questions. You'll probably still end up doing a lot of digging through other online sources, because the Hadoop ecosystem is large and complicated, and no book can really cover all of it. Besides this book, I also bought Hadoop In Action (not quite as big as this book, but a useful counter-point) and Data Intensive Text Processing With MapReduce (which gave me a good intro to the Map Reduce algorithm, but wasn't that useful once I had a general idea what was going on).

This book is so good that it forced it me to write my first book review on  :). Everything is so clearly described and thoroughly explained both for Hadoop and related Apache projects. Read (and

re-read) at least the whole section I thoroughly, it is very important for the understanding of the rest of the book.

It's lacking on a lot of details. Maybe it tries to cover too much; Hadoop is a large, and growing daily, ecosystem. It's not simple and extremely fluid. Calling any printed material a "definitive guide" is asking for trouble.One example of failed detail is around configuration for the "Fair scheduler." It basically says (in a few paragraphs) turn it on and let the defaults fly. The Hadoop Operations book has 3 pages on the subject and mentions at least 1, "You must configure this for it to work" variable.Buy it if you want a good overview of everything at the time of publishing. Don't if you expect it to be your handbook for building a Hadoop Cluster.

I feel like this book is very hard to follow and not organized well. This book is better as a reference book then a book to teach you about Hadoop. I had taken a Hadoop training course before reading this and I still had a difficult time following along. With that said, the combination of the training course and this book, I was able to pass the Cloudera Certfied Developer for Apache Hadoop (CCDH) certification exam. I gave this book 2 stars because I believe there was some key pieces of knowledge in this book that was asked on the exam.

Howdy Hoop, look at all of that data!

Wasn't as complete as I would have like with multiple syntax examples. I had to google some of the error messages I would recieve using syntax in the book.

Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale Internet Business Insights: Lessons Learned and Strategies Used by 101 Successful Internet-Based Entrepreneurs (Internet Business Books) ESP8266: Programming NodeMCU Using Arduino IDE - Get Started With ESP8266 (Internet Of Things, IOT, Projects In Internet Of Things, Internet Of Things for Beginners, NodeMCU Programming, ESP8266) Guns Danger & Safety 2nd Edition: An Essential Guide In Firearm Ammunition, Loading, Shooting, Storage and Safety (Guns, Guns & Ammo, Ammunition, Hunting, ... Loading, Targets, Handguns, Gun Storage) Storage Unit Auctions: A Practical Guide to Profiting with Storage Unit Auctions Build Your Own Cedar Storage Chest DIY PLANS HOPE BLANKET TOY BOX STORAGE PATTERNS; So Easy, Beginners Look Like Experts; PDF Download Version so you can get it NOW! Programming Hive: Data Warehouse and Query

Language for Hadoop L590 - Progressive Scale Studies - Scale Study and Practical Theory in Major and Minor Keys for the Young Violinist SEO - The Sassy Way of Ranking #1 in Google - when you have NO CLUE!: Beginner's Guide to Search Engine Optimization and Internet Marketing (Beginner Internet Marketing Series) (Volume 3) The Scale and Arpeggio Bible for Clarinet: (probably the only scale book you'll ever need) Beyond Powerful Radio: A Communicator's Guide to the Internet Ageâ •News, Talk, Information & Personality for Broadcasting, Podcasting, Internet, Radio Pocket Neighborhoods: Creating Small-Scale Community in a Large-Scale World Walt Disney's Railroad Story: The Small-Scale Fascination That Led to a Full-Scale Kingdom Scale Studies for Viola: Based on the Hrimaly Scale Studies for the Violin Scale Model Life: Building Scale Model Kits Magazine (Volume 2) Detailing Scale Model Aircraft (Scale Modeling Handbook) 2018 Rand McNally Large Scale Road Atlas (Rand McNally Large Scale Road Atlas U. S. A.) Hagstrom Westchester County & Metro New York Large Scale: Covering a 75-Mile Radius from Midtown Manhattan (Hagstrom Westchester County Atlas Large Scale Edition) Hagstrom Morris/Sussex/Warren Counties Atlas: Large Scale Edition (Hagstrom Warren, Morris, Sussex Counties Atlas Large Scale) The Usborne Internet-Linked Children's Encyclopedia. [Written and Researched by Felicity Brooks ... [Et Al.] (Internet-Linked Reference Books)