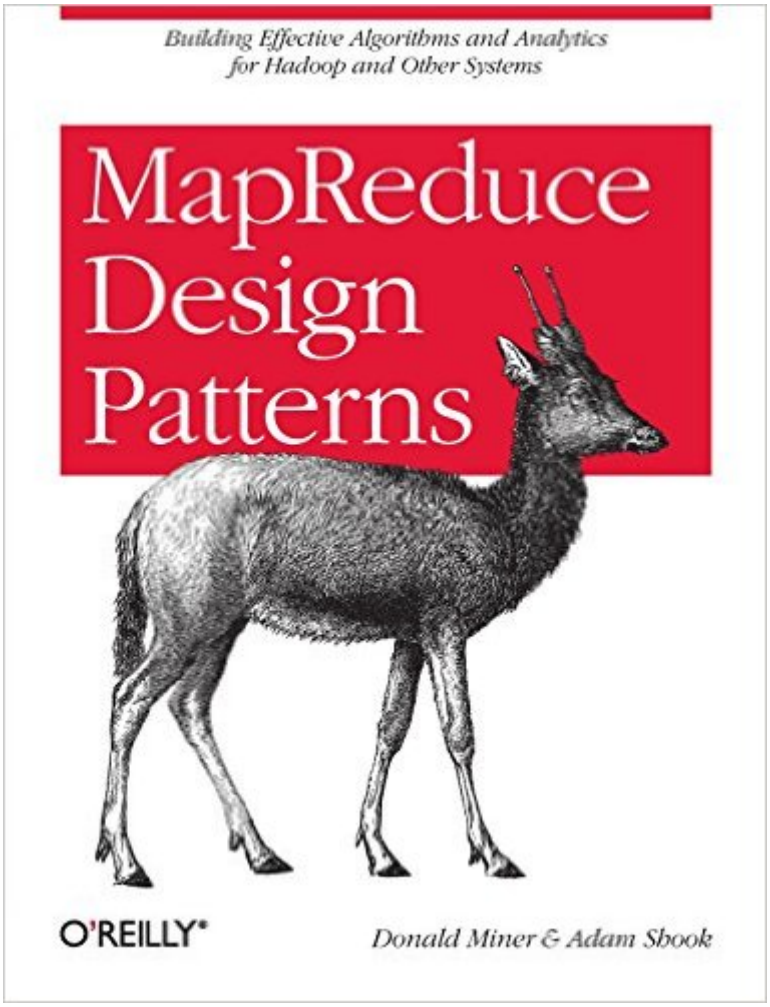
The book was found

MapReduce Design Patterns: Building Effective Algorithms And Analytics For Hadoop And Other Systems



PDF
Adobe
DOWNLOAD EBOOK

Synopsis

Until now, design patterns for the MapReduce framework have been scattered among various research papers, blogs, and books. This handy guide brings together a unique collection of valuable MapReduce patterns that will save you time and effort regardless of the domain, language, or development framework you’re using. Each pattern is explained in context, with pitfalls and caveats clearly identified to help you avoid common design mistakes when modeling your big data architecture. This book also provides a complete overview of MapReduce that explains its origins and implementations, and why design patterns are so important. All code examples are written for Hadoop.Summarization patterns: get a top-level view by summarizing and grouping data Filtering patterns: view data subsets such as records generated from one user Data organization patterns: reorganize data to work with other systems, or to make MapReduce analysis easier Join patterns: analyze different datasets together to discover interesting relationships Metapatterns: piece together several patterns to solve multi-stage problems, or to perform several analytics in the same job Input and output patterns: customize the way you use Hadoop to load or store data ""A clear exposition of MapReduce programs for common data processing patterns—this book is indespensible for anyone using Hadoop."" --Tom White, author of Hadoop: The Definitive Guide

Book Information

Paperback: 230 pages
Publisher: O’Reilly Media; 1 edition (December 22, 2012)
Language: English
ISBN-10: 1449327176
ISBN-13: 978-1449327170
Product Dimensions: 7 x 0.6 x 9 inches
Shipping Weight: 12.6 ounces (View shipping rates and policies)
Average Customer Review: 4.0 out of 5 starsÂ Â See all reviewsÂ (20 customer reviews)
Best Sellers Rank: #285,020 in Books (See Top 100 in Books) #159 inÂ Books > Computers & Technology > Databases & Big Data > Data Modeling & Design #408 inÂ Books > Textbooks > Computer Science > Software Design & Engineering #848 inÂ Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Software Development

Customer Reviews

In the 1990s O’Reilly books had a well-earned reputation for quality. O’Reilly authors such as Simson Garfinkel explained technical topics with precision, clarity, and wit. I proudly kept a whole

shelf of O'Reilly books at work, and I imbibed copious java from their tenth anniversary mug. I'm sorry to see that O'Reilly's traditional quality has gone the way of the Internet bubble. MapReduce Design Patterns represents the absolute nadir of technical writing, and it never should have been published in its current form.One of the most poorly written parts of the book is Appendix A on Bloom filters. As I was writing my original review of the book, I thought it might be helpful to point readers to a better explanation of the topic. Turning to Wikipedia as a potential reference, I was struck by the number of similarities between it and Appendix A. It now appears that this appendix plagiarizes the Wikipedia article "Bloom filter." To see this, compare the opening paragraph of the Wikipedia article (January 19, 2013) to the first two paragraphs of the book's appendix (which you can see in the sample pages here): Wiki: A Bloom filter, conceived by Burton Howard Bloom in 1970, is a space-efficient probabilistic data structure that is used to test whether an element is a member of a set. (Paragraph 1, sentence 1) MRDP: Conceived by Burton Howard Bloom in 1970, a Bloom filter is a probabilistic data structure used to test whether a member is an element of a set. (Page 221, paragraph 1, sentence 1) Wiki: False positive retrieval results are possible, but false negatives are not; i.e. a query returns either "inside set (may be wrong)" or "definitely not in set".

It's a good book that show you a lot of Mapreduce patterns using Hadoop.But the main trouble it's that you can not trust the examples source code, at all. I Clone the Code Github in my Mac and I've found several bugs..https://github.com/adamjshook/mapreducepatternsI'm running the Book's code using a macbook with:- hadoop-1.0.4- Mac OS ver 10.6.8- Java ver "1.6.0_43"- Eclipse- Data for running the examples from (Stack Exchange Data Dump - Dec 2011 _Update_)For now, these are the bugs I've found:Page: 31The error is in the MedianStdDevCombiner code.I'm looking for a bug in this full example because when you execute it ,you obtain different result from the previous normal Median and Standard deviation using the same input data. The result obtained is nearly double values from the previous example, when need to be the same results.Page: 35-36The error i found is in the Inverted Index Example.In the Mapper Function if "getWikipediaURL" return a null value then you get a nullpointerException because you need to check if the result of this function is null prior to set the "link" variable value.Page 117-118In ReduceSideJoinWithBloomDriver Code from github doesn't exist any reference to load the bloom filter from any argument... [something like DistributedCache.addCacheFile(.....], this file is nearly a Copy/paste from the previous ReduceSideJoin.java.Page 122:In ReplicatedJoinMapper you always get a java.io.FileNotFoundException because this code want to read and decompress a folder , not a concrete "file.gz", inside this folder. You only need to add a index to your files inside the

DistributedCache.

Download to continue reading...

MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Agile Data Science: Building Data Analytics Applications with Hadoop Big Data Analytics with R and Hadoop Data Analytics with Hadoop: An Introduction for Data Scientists The Design of Innovation: Lessons from and for Competent Genetic Algorithms (Genetic Algorithms and Evolutionary Computation) Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications (Real-Time Systems Series) 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) Even You Can Learn Statistics and Analytics: An Easy to Understand Guide to Statistics and Analytics (3rd Edition) Analytics: Data Science, Data Analysis and Predictive Analytics for Business Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data) People Analytics: How Social Sensing Technology Will Transform Business and What It Tells Us about the Future of Work (FT Press Analytics) Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (MIT Press) Algorithms in C++ Part 5: Graph Algorithms (3rd Edition) (Pt.5) Hadoop Application Architectures Pro Apache Hadoop Practical Hive: A Guide to Hadoop's Data Warehouse System Hadoop in Practice Safari Animal Patterns: 30 Exotic Safari Animal Patterns to Feel the Wildlife World (Safari Animal Patterns, animal designs, zendoodle) Genetic Algorithms and Engineering Design (Engineering Design and Automation)

Dmca