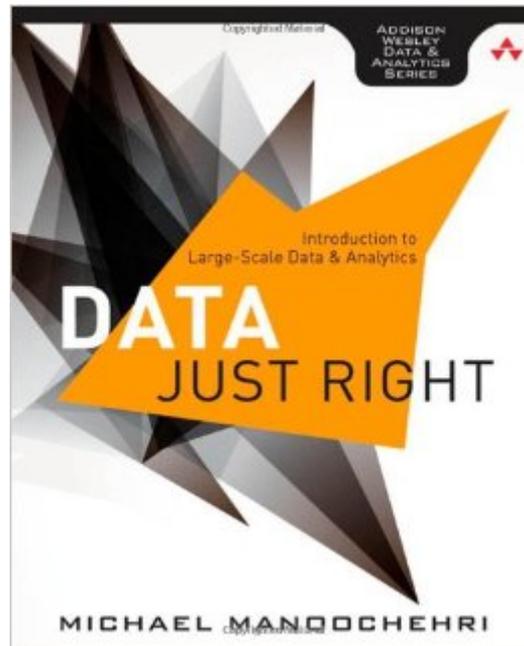# Data Just Right: Introduction To Large-Scale Data & Analytics (Addison-Wesley Data And Analytics)

Synopsis

Making Big Data Work: Real-World Use Cases and Examples, Practical Code, Detailed Solutions Â Large-scale data analysis is now vitally important to virtually every business. Mobile and social technologies are generatingÂ massive datasets; distributed cloud computing offers the resources to store and analyze them; and professionals have radically new technologies at their command, including NoSQL databases. Until now, however, most books on â œBig Dataâ • have been little more than business polemics or product catalogs. Data Just Right is different: Itâ ™s a completely practical and indispensable guide for every Big Data decision-maker, implementer, and strategist. Â Michael Manoochehri, a former Google engineer and data hacker, writes for professionals who need practical solutions that can be implemented with limited resources and time. Drawing on his extensive experience, he helps you focus on building applications, rather than infrastructure, because thatâ ™s where you can derive the most value. Â Manoochehri shows how to address each of todayâ ™s key Big Data use cases in a cost-effective way by combining technologies in hybrid solutions. Youâ ™ll find expert approaches to managing massive datasets, visualizing data, building data pipelines and dashboards, choosing tools for statistical analysis, and more. Throughout, the author demonstrates techniques using many of todayâ ™s leading data analysis tools, including Hadoop, Hive, Shark, R, Apache Pig, Mahout, and Google BigQuery. Â Coverage includes Mastering the four guiding principles of Big Data successâ ”and avoiding common pitfalls Emphasizing collaboration and avoiding problems with siloed data Hosting and sharing multi-terabyte datasets efficiently and economically â œBuilding for infinityâ • to support rapid growth Developing a NoSQL Web app with Redis to collect crowd-sourced data Running distributed queries over massive datasets with Hadoop, Hive, and Shark Building a data dashboard with Google BigQuery Exploring large datasets with advanced visualization Implementing efficient pipelines for transforming immense amounts of data Automating complex processing with Apache Pig and the Cascading Java library Applying machine learning to classify, recommend, and predict incoming information Using R to perform statistical analysis on massive datasets Building highly efficient analytics workflows with Python and Pandas Establishing sensible purchasing strategies: when to build, buy, or outsource Previewing emerging trends and convergences in scalable data technologies and the evolving role of the Data ScientistÂ

Book Information

Series: Addison-Wesley Data and Analytics
Paperback: 256 pages

Publisher: Addison-Wesley Professional; 1 edition (December 29, 2013)

Language: English

ISBN-10: 0321898656

ISBN-13: 978-0321898654

Product Dimensions: 7 x 0.5 x 9.2 inches

Shipping Weight: 13.4 ounces (View shipping rates and policies)

Average Customer Review: 4.2 out of 5 stars Â Â See all reviews Â (6 customer reviews)

Best Sellers Rank: #793,029 in Books (See Top 100 in Books)   #327 in Â Books > Computers & Technology > Networking & Cloud Computing > Network Administration > Storage & Retrieval   #443 in Â Books > Computers & Technology > Databases & Big Data > Data Mining   #836 in Â Books > Textbooks > Computer Science > Database Storage & Design

Customer Reviews

Hive, Hadoop, Shark, Dremel, BigQuery, SciPy, NumPy, Pandas, R, Pig... whether you are new or a seasoned big data expert, there is a big and growing universe of keywords to understand. In this book Manoochehri manages to give a through review on the whys and hows, giving the reader just the right depth in each topic to understand the motivation for each of these different technologies, how they are different to each other, and why you would want to use them. I love that he's not afraid to jump and write code, as - when you do it just right - a few lines of code are much more illustrative than a picture or block of texts would do.Totally recommended. If you want to learn Hadoop, buy a Hadoop book - or an R book if you want to go deeper in that topic. But if you want to understand the current big data universe, how the tools interrelate between each other, and go from data generation to storage to analysis to visualization - this is the book.

If you work with expensive enterprise strength data management/analysis products like SAS and Oracle and you want a book that will give you a map to cover the open source tools for dealing with "big data" (i.e., Hadoop, Hive, and Pig) get this. It does an amazingly good job of explaining the utility of the various tools that are used to manage *HUGE* data. Everything from the practical concerns in designing web facing applications to analytic data-sets are covered at the perfect depth for someone who knows a bit about data and databases. Even if you are not a programmer, the author does an exceptional job of explaining things from the ground up without babying the reader (e.g., what are the advantages of using CSV files vs XML vs JSON vs Thrift vs Avro). There are code snippets scattered throughout that are useful for comparing and contrasting if you know some

programming languages (e.g., SQL queries vs HiveQL) but the book does not attempt to explain the code in great detail. So, you end up with the outline of what a tool does without getting bogged down in the gory details. If you want to go deeper into the solutions the book is full of references to seminal white papers and other external references so you can expand on what is covered.So, if you keep hearing about things like Hadoop, noSQL, Python, SciPy, Pandas, R and you just want to learn "what is the big deal" or "why bother" learning yet another tool, this is the perfect book.

This book provides an interesting overview of main technologies in data science, but strikes a slightly odd balance between technical and descriptive -- there are some brief code examples that can get you on the way or that give you an impression of the functionality of the particular tool, but it remains very superficial. In the end I have neither the impression I have a good overview of the tools available (at least, not beyond what I already had), nor do I know much in detail about each of them. Most items are explained in too simple language, using analogies where technical detail would have been more interesting. It's also slightly repetitive at times. I think the author has tried to please both more technically inclined and others at the same time, which hasn't really worked.So, if you want a very quick overview of what data science is, this is an easy read and provides you just that, but if you want anything deeper out of it, I think this book is somewhat disappointing.

Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data and Analytics) R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data & Analytics Series) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics Series) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics Series) Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference (Addison-Wesley Data & Analytics) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) Sing You Home Large Print (Large Print, companion soundtrack, Large Print) Word Search Puzzles Large Print: Large print word search, Word search books, Word search books for adults, Adult word search books, Word search puzzle books, Extra large print word search The Design and

Implementation of the 4.4 BSD Operating System (Addison-Wesley UNIX and Open Systems Series) First Principles of Discrete Systems and Digital Signal Processing (Addison-Wesley Series in Electrical Engineering) Essential SharePoint 2010: Overview, Governance, and Planning (Addison-Wesley Microsoft Technology) Principles of Compiler Design (Addison-Wesley series in computer science and information processing) Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Adobe Reader) (Addison-Wesley Signature Series (Fowler)) Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions (Addison-Wesley Signature Series (Fowler)) Circuits, Interconnections, and Packaging for Vlsi (Addison-Wesley VLSI systems series) Patterns of Enterprise Application Architecture (Addison-Wesley Signature Series (Fowler)) Essential SharePointÃ Â® 2013: Practical Guidance for Meaningful Business Results (3rd Edition) (Addison-Wesley Microsoft Technology)