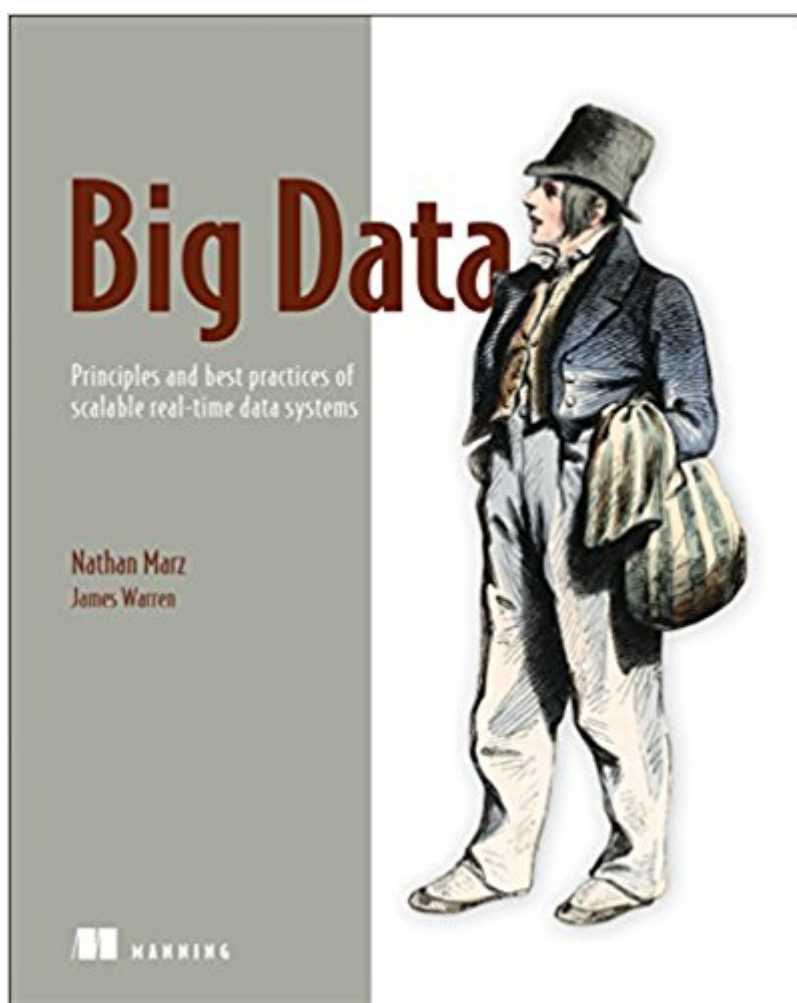The book was found

Big Data: Principles And Best Practices Of Scalable Realtime Data Systems

Synopsis

SummaryBig Data teaches you to build big data systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data. It describes a scalable, easy-to-understand approach to big data systems that can be built and run by a small team. Following a realistic example, this book guides readers through the theory of big data systems, how to implement them in practice, and how to deploy and operate them once they're built.Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.About the BookWeb-scale applications like social networks, real-time analytics, or e-commerce sites deal with a lot of data, whose volume and velocity exceed the limits of traditional database systems. These applications require architectures built around clusters of machines to store and process data of any size, or speed. Fortunately, scale and simplicity are not mutually exclusive.Big Data teaches you to build big data systems using an architecture designed specifically to capture and analyze web-scale data. This book presents the Lambda Architecture, a scalable, easy-to-understand approach that can be built and run by a small team. You'll explore the theory of big data systems and how to implement them in practice. In addition to discovering a general framework for processing big data, you'll learn specific technologies like Hadoop, Storm, and NoSQL databases.This book requires no previous exposure to large-scale data analysis or NoSQL tools. Familiarity with traditional databases is helpful.What's InsideIntroduction to big data systemsReal-time processing of web-scale dataTools like Hadoop, Cassandra, and StormExtensions to traditional database skillsAbout the AuthorsNathan Marz is the creator of Apache Storm and the originator of the Lambda Architecture for big data systems. James Warren is an analytics architect with a background in machine learning and scientific computing.Table of ContentsA new paradigm for Big DataPART 1 BATCH LAYERData model for Big DataData model for Big Data: IllustrationData storage on the batch layerData storage on the batch layer: IllustrationBatch layerBatch layer: IllustrationAn example batch layer: Architecture and algorithmsAn example batch layer: ImplementationPART 2 SERVING LAYERServing layerServing layer: IllustrationPART 3 SPEED LAYERRealtime viewsRealtime views: IllustrationQueuing and stream processingQueuing and stream processing: IllustrationMicro-batch stream processingMicro-batch stream processing: IllustrationLambda Architecture in depth

Book Information

Language: English

ISBN-10: 1617290343

ISBN-13: 978-1617290343

Product Dimensions: 7.3 x 0.6 x 9.1 inches

Shipping Weight: 1.8 pounds (View shipping rates and policies)

Average Customer Review: 4.2 out of 5 stars   34 customer reviews

Best Sellers Rank: #45,328 in Books (See Top 100 in Books)   #9 inÃ Â Books > Computers & Technology > Web Development & Design > User Generated Content   #36 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Mining   #44 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Processing

Customer Reviews

View larger   About This Book   Services like social networks, web analytics, and intelligent e-commerce often need to manage data at a scale too big for a traditional database. Complexity increases with scale and demand, and handling Big Data is not as simple as just doubling down on your RDBMS or rolling out some trendy new technology. Fortunately, scalability and simplicity are not mutually exclusiveÃƒÂ¢Ã â¬Ã â¢you just need to take a different approach. Big Data systems use many machines working in parallel to store and process data, which introduces fundamental challenges unfamiliar to most developers.   Big Data teaches you to build these systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data. It describes a scalable, easy-to-understand approach to Big Data systems that can be built and run by a small team. Following a realistic example, this book guides readers through the theory of Big Data systems and how to implement them in practice.   Big Data requires no previous exposure to large-scale data analysis or NoSQL tools. Familiarity with traditional databases is helpful, though not required. The goal of the book is to teach you how to think about data systems and how to break down difficult problems into simple solutions. We start from first principles and from those deduce the necessary properties for each component of an architecture.

Nathan Marz is currently working on a new startup. Previously, he was the lead engineer at BackType before being acquired by Twitter in 2011. At Twitter, he started the streaming compute team which provides and develops shared infrastructure to support many critical realtime applications throughout the company. Nathan is the creator of Cascalog and Storm, open-source

projects which are relied upon by over 50 companies around the world, including Yahoo!, Twitter, Groupon, The Weather Channel, Taobao, and many more companies.James Warren is an analytics architect at Storm8 with a background in big data processing, machine learning and scientific computing.

If you are looking for a survey of different approaches of handling big data, you want to read "ELEMENTS OF SCALE: COMPOSING AND SCALING DATA PLATFORMS". ([...]) This book is dedicated to Lambda Architecture (one that is surveyed in the above article.)The book is very organized. Introduction in chapter 1 will be the road map of the whole book. Motivating with a simple web application based on RDBMS, the author showed how the approach to scale it becomes undesirable. After enumerating a list of desired properties, he proposed Lambda architecture, an approach in contrast to fully incremental architecture (with RDBMS).The Lambda architecture is partitioned into three layers:1. batch layer that computes different views on big data2. serving layer that answers user queries using views from the batch layer and speed layer.3. speed layer that compensates an approximate answer over a period time when the batch layer is working on the complete answers.In the remaining chapters, the author dive deep into the rationale and requirements of all the different pieces of Lambda Architecture.To under the context of Lambda Architecture, also refer to the wikipedia for crticism.

This book is written by a specialist in big data. I know that because I worked on the big data pipeline. And now I read the book and I see that all my problems are addressed in this book. Virtually every problem discussed appeared in my pipeline too, as if the author worked with me on my project.The other very useful for me feature of this book is that it is the first book where I could find a concise explanation of Storm Trident framework, even though the book is not about Storm.

Good theoretical review of Big Data architecture. Not so great for implementation details using current frameworks.

Gave me a good idea of what Big Data actually is, and how it's implemented. Another tool in the chest.

This book serves as a guide to building the lambda architecture from scratch, processing and serving big data. It is not a book to teach you about big data technologies though. If you are looking

to learn and implement a system with batch and real time layers, this is the best book there is.

Basically a sell of Lambda Architecture. A bit dated to use for proposed architecture since there many new design patterns now that get around some of the limitations. Nice source of 'things to consider/be aware of' though.

I have rarely seen a thorough discussion of the importance of data modeling, data layers, data processing requirements analysis, and data architecture and storage implementation issues (along with other "traditional" database concepts) in the context of big data. This book delivers a refreshing comprehensive solution to that deficiency. Other books in this area tend to focus a lot more on the "gee whiz" coolness of data science and machine learning applications (which are aspects of big data that I happen to love, but they are not the whole story). You cannot hope to achieve good, effective, and efficient results from your analytics processes without good data flow, from discovery to access to integration, which is why architecture design, data modeling, and attention to data pipelining are essential. I highly recommend this book for anyone who isn't ashamed to admit that data engineering is at least as important as data science in the big data era (says this data scientist!).

Great explanation of both the theory and practice of the lambda architecture. While the practice chapters are nice, it's the theory chapters that really shine. The book explains down to the byte level why components are implemented the way they are. For example, there's an immense amount of detail as to why using a db that doesn't support random writes allows for an application to query the batch layer's results without locking.The only downside to the book is that the architecture and exosystem is so new that there's not really a lot of pragmatic solutions. For example, the theory describes a query layer that can merge the results of batch and real time processing for client applications. However, in real life there are no pragmatic solutions for doing this so you'd have to write your own.It'll be interesting to see how the lambda architecture matures and to see future editions of this book. Hopefully, future editions will be as well written and have a better ecosystem for practice chapters.

Download to continue reading...

Big Data: Principles and best practices of scalable realtime data systems Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Big Data For Business: Your Comprehensive Guide to

Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services Florida Real Estate Principles, Practices & Law (Florida Real Estate Principles, Practices and Law) Florida Real Estate Principles, Practices and Law, 33rd Edition (Florida Real Estate Principles, Practices & Law) Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) RealTime Physics Active Learning Laboratories, Module 3: Electricity and Magnetism RealTime Physics Active Learning Laboratories, Module 4: Light and Optics RealTime Physics Active Learning Laboratories, Module 1: Mechanics Tokyo Realtime: Akihabara Inside Sales Predictability: 7 insider secrets to building a predictable and scalable real estate business with inside sales Startup Evolution Curve From Idea to Profitable and Scalable Business: Startup Marketing Manual Practical Machine Learning with H2O: Powerful, Scalable Techniques for Deep Learning and AI The Startup Checklist: 25 Steps to a Scalable, High-Growth Business Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results My Growing-Up Library: Sesame Street Board Books- Too Big for Diapers / Big Enough for a Bed / Too Big for Bottles / Big Enough for a Bike