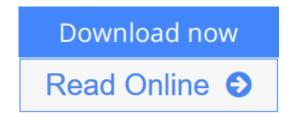# Big Data: Principles and best practices of scalable realtime data systems

By Nathan Marz, James Warren

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren

Summary

Big 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 Book

Web-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 Inside

- Introduction to big data systems
- Real-time processing of web-scale data
- Tools like Hadoop, Cassandra, and Storm
- Extensions to traditional database skills

About the Authors

Nathan 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 Contents

1. A new paradigm for Big DataPART 1 BATCH LAYER
2. Data model for Big Data
3. Data model for Big Data: Illustration
4. Data storage on the batch layer
5. Data storage on the batch layer: Illustration
6. Batch layer
7. Batch layer: Illustration
8. An example batch layer: Architecture and algorithms
9. An example batch layer: ImplementationPART 2 SERVING LAYER
10. Serving layer
11. Serving layer: IllustrationPART 3 SPEED LAYER
12. Realtime views
13. Realtime views: Illustration
14. Queuing and stream processing
15. Queuing and stream processing: Illustration
16. Micro-batch stream processing
17. Micro-batch stream processing: Illustration
18. Lambda Architecture in depth

Big Data: Principles and best practices of scalable realtime data systems

By Nathan Marz, James Warren

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren

Summary

Big 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 Book

Web-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 Inside

- Introduction to big data systems
- Real-time processing of web-scale data
- Tools like Hadoop, Cassandra, and Storm
- Extensions to traditional database skills

About the Authors

Nathan 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 Contents

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren Bibliography

- Sales Rank: #28321 in Books
- Published on: 2015-05-10
- Original language: English
- Number of items: 1
- Dimensions: 9.10" h x .60" w x 7.30" l, .0 pounds
- Binding: Paperback
- 328 pages

Download and Read Free Online Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren

Editorial Review

About the Author

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.

Users Review

From reader reviews:

Ismael Roop:

This book untitled Big Data: Principles and best practices of scalable realtime data systems to be one of several books that will best seller in this year, this is because when you read this reserve you can get a lot of benefit onto it. You will easily to buy this kind of book in the book retailer or you can order it through online. The publisher on this book sells the e-book too. It makes you more easily to read this book, because you can read this book in your Smartphone. So there is no reason for your requirements to past this e-book from your list.

Saul Robinson:

The reason why? Because this Big Data: Principles and best practices of scalable realtime data systems is an unordinary book that the inside of the publication waiting for you to snap this but latter it will jolt you with the secret the item inside. Reading this book beside it was fantastic author who have write the book in such incredible way makes the content inside of easier to understand, entertaining approach but still convey the meaning totally. So , it is good for you for not hesitating having this anymore or you going to regret it. This phenomenal book will give you a lot of gains than the other book have got such as help improving your skill and your critical thinking method. So , still want to delay having that book? If I were you I will go to the e-book store hurriedly.

Clarence Ross:

Many people spending their moment by playing outside along with friends, fun activity together with family

or just watching TV the entire day. You can have new activity to pay your whole day by examining a book. Ugh, do you consider reading a book can really hard because you have to use the book everywhere? It alright you can have the e-book, having everywhere you want in your Smart phone. Like Big Data: Principles and best practices of scalable realtime data systems which is having the e-book version. So , try out this book? Let's find.

Betty Dansby:

Don't be worry if you are afraid that this book may filled the space in your house, you may have it in e-book method, more simple and reachable. This Big Data: Principles and best practices of scalable realtime data systems can give you a lot of buddies because by you taking a look at this one book you have issue that they don't and make you actually more like an interesting person. This particular book can be one of a step for you to get success. This book offer you information that possibly your friend doesn't understand, by knowing more than various other make you to be great folks. So , why hesitate? Let us have Big Data: Principles and best practices of scalable realtime data systems.

Download and Read Online Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren #GVAOIEQ3U6S

Read Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren for online ebook

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren 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 Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren books to read online.

Online Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren ebook PDF download

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren Doc

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren Mobipocket

Big Data: Principles and best practices of scalable realtime data systems By Nathan Marz, James Warren EPub