

Learning Hadoop 2

By Garry Turkington



Learning Hadoop 2 By Garry Turkington

Design and implement data processing, lifecycle management, and analytic workflows with the cutting-edge toolbox of Hadoop 2

About This Book

- Construct state-of-the-art applications using higher-level interfaces and tools beyond the traditional MapReduce approach
- Use the unique features of Hadoop 2 to model and analyze Twitter’s global stream of user generated data
- Develop a prototype on a local cluster and deploy to the cloud (Amazon Web Services)

Who This Book Is For

If you are a system or application developer interested in learning how to solve practical problems using the Hadoop framework, then this book is ideal for you. You are expected to be familiar with the Unix/Linux command-line interface and have some experience with the Java programming language. Familiarity with Hadoop would be a plus.

What You Will Learn

- Write distributed applications using the MapReduce framework
- Go beyond MapReduce and process data in real time with Samza and iteratively with Spark
- Familiarize yourself with data mining approaches that work with very large datasets
- Prototype applications on a VM and deploy them to a local cluster or to a cloud infrastructure (Amazon Web Services)
- Conduct batch and real time data analysis using SQL-like tools
- Build data processing flows using Apache Pig and see how it enables the easy incorporation of custom functionality
- Define and orchestrate complex workflows and pipelines with Apache Oozie
- Manage your data lifecycle and changes over time


In Detail

This book introduces you to the world of building data-processing applications with the wide variety of tools supported by Hadoop 2. Starting with the core components of the framework—HDFS and YARN—this book will guide you through how to build applications using a variety of approaches.

You will learn how YARN completely changes the relationship between MapReduce and Hadoop and allows the latter to support more varied processing approaches and a broader array of applications. These include real-time processing with Apache Samza and iterative computation with Apache Spark. Next up, we discuss Apache Pig and the dataflow data model it provides. You will discover how to use Pig to analyze a Twitter dataset.

With this book, you will be able to make your life easier by using tools such as Apache Hive, Apache Oozie, Hadoop Streaming, Apache Crunch, and Kite SDK. The last part of this book discusses the likely future direction of major Hadoop components and how to get involved with the Hadoop community.

 [Download Learning Hadoop 2 ...pdf](#)

 [Read Online Learning Hadoop 2 ...pdf](#)

Learning Hadoop 2

By Garry Turkington

Learning Hadoop 2 By Garry Turkington

Design and implement data processing, lifecycle management, and analytic workflows with the cutting-edge toolbox of Hadoop 2

About This Book

- Construct state-of-the-art applications using higher-level interfaces and tools beyond the traditional MapReduce approach
- Use the unique features of Hadoop 2 to model and analyze Twitter’s global stream of user generated data
- Develop a prototype on a local cluster and deploy to the cloud (Amazon Web Services)

Who This Book Is For

If you are a system or application developer interested in learning how to solve practical problems using the Hadoop framework, then this book is ideal for you. You are expected to be familiar with the Unix/Linux command-line interface and have some experience with the Java programming language. Familiarity with Hadoop would be a plus.

What You Will Learn

- Write distributed applications using the MapReduce framework
- Go beyond MapReduce and process data in real time with Samza and iteratively with Spark
- Familiarize yourself with data mining approaches that work with very large datasets
- Prototype applications on a VM and deploy them to a local cluster or to a cloud infrastructure (Amazon Web Services)
- Conduct batch and real time data analysis using SQL-like tools
- Build data processing flows using Apache Pig and see how it enables the easy incorporation of custom functionality
- Define and orchestrate complex workflows and pipelines with Apache Oozie
- Manage your data lifecycle and changes over time

In Detail

This book introduces you to the world of building data-processing applications with the wide variety of tools supported by Hadoop 2. Starting with the core components of the framework—HDFS and YARN—this book will guide you through how to build applications using a variety of approaches.

You will learn how YARN completely changes the relationship between MapReduce and Hadoop and allows the latter to support more varied processing approaches and a broader array of applications. These include real-time processing with Apache Samza and iterative computation with Apache Spark. Next up, we discuss Apache Pig and the dataflow data model it provides. You will discover how to use Pig to analyze a Twitter dataset.

With this book, you will be able to make your life easier by using tools such as Apache Hive, Apache Oozie, Hadoop Streaming, Apache Crunch, and Kite SDK. The last part of this book discusses the likely future direction of major Hadoop components and how to get involved with the Hadoop community.

Learning Hadoop 2 By Garry Turkington Bibliography

- Sales Rank: #1199237 in eBooks
- Published on: 2015-02-13
- Released on: 2015-02-13
- Format: Kindle eBook

 [Download Learning Hadoop 2 ...pdf](#)

 [Read Online Learning Hadoop 2 ...pdf](#)

Download and Read Free Online Learning Hadoop 2 By Garry Turkington

Editorial Review

About the Author

Garry Turkington

Garry Turkington has over 15 years of industry experience, most of which has been focused on the design and implementation of large-scale distributed systems. In his current role as the CTO at Improve Digital, he is primarily responsible for the realization of systems that store, process, and extract value from the company's large data volumes. Before joining Improve Digital, he spent time at Amazon.co.uk, where he led several software development teams, building systems that process the Amazon catalog data for every item worldwide. Prior to this, he spent a decade in various government positions in both the UK and the USA. He has BSc and PhD degrees in Computer Science from Queens University Belfast in Northern Ireland, and a Master's degree in Engineering in Systems Engineering from Stevens Institute of Technology in the USA. He is the author of Hadoop Beginners Guide, published by Packt Publishing in 2013, and is a committer on the Apache Samza project.

Gabriele Modena

Gabriele Modena is a data scientist at Improve Digital. In his current position, he uses Hadoop to manage, process, and analyze behavioral and machine-generated data. Gabriele enjoys using statistical and computational methods to look for patterns in large amounts of data. Prior to his current job in ad tech he held a number of positions in Academia and Industry where he did research in machine learning and artificial intelligence. He holds a BSc degree in Computer Science from the University of Trento, Italy and a Research MSc degree in Artificial Intelligence: Learning Systems, from the University of Amsterdam in the Netherlands.

Users Review

From reader reviews:

Ryan Calhoun:

The book Learning Hadoop 2 gives you the sense of being enjoy for your spare time. You can utilize to make your capable considerably more increase. Book can to get your best friend when you getting tension or having big problem together with your subject. If you can make examining a book Learning Hadoop 2 for being your habit, you can get much more advantages, like add your own personal capable, increase your knowledge about some or all subjects. You may know everything if you like available and read a publication Learning Hadoop 2. Kinds of book are a lot of. It means that, science reserve or encyclopedia or other individuals. So , how do you think about this reserve?

Darlene Johnson:

This Learning Hadoop 2 is great guide for you because the content which can be full of information for you who all always deal with world and get to make decision every minute. This particular book reveal it data accurately using great plan word or we can point out no rambling sentences inside it. So if you are read the idea hurriedly you can have whole data in it. Doesn't mean it only provides you with straight forward sentences but tricky core information with splendid delivering sentences. Having Learning Hadoop 2 in your hand like having the world in your arm, facts in it is not ridiculous 1. We can say that no book that offer you world within ten or fifteen tiny right but this guide already do that. So , this can be good reading book. Hi Mr. and Mrs. hectic do you still doubt in which?

William McDowell:

Reading a book to be new life style in this 12 months; every people loves to go through a book. When you learn a book you can get a lot of benefit. When you read ebooks, you can improve your knowledge, because book has a lot of information upon it. The information that you will get depend on what sorts of book that you have read. If you wish to get information about your study, you can read education books, but if you act like you want to entertain yourself you are able to a fiction books, this sort of us novel, comics, as well as soon. The Learning Hadoop 2 provide you with new experience in studying a book.

Joyce Washington:

As we know that book is significant thing to add our understanding for everything. By a reserve we can know everything we really wish for. A book is a group of written, printed, illustrated or blank sheet. Every year was exactly added. This e-book Learning Hadoop 2 was filled about science. Spend your free time to add your knowledge about your technology competence. Some people has different feel when they reading the book. If you know how big benefit of a book, you can experience enjoy to read a book. In the modern era like currently, many ways to get book which you wanted.

**Download and Read Online Learning Hadoop 2 By Garry
Turkington #J4E7CHNQF9A**

Read Learning Hadoop 2 By Garry Turkington for online ebook

Learning Hadoop 2 By Garry Turkington 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 Learning Hadoop 2 By Garry Turkington books to read online.

Online Learning Hadoop 2 By Garry Turkington ebook PDF download

Learning Hadoop 2 By Garry Turkington Doc

Learning Hadoop 2 By Garry Turkington Mobipocket

Learning Hadoop 2 By Garry Turkington EPub