Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies

By Kevin Sitto, Marshall Presser

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser

If your organization is about to enter the world of big data, you not only need to decide whether Apache Hadoop is the right platform to use, but also which of its many components are best suited to your task. This field guide makes the exercise manageable by breaking down the Hadoop ecosystem into short, digestible sections. You'll quickly understand how Hadoop's projects, subprojects, and related technologies work together.

Each chapter introduces a different topic—such as core technologies or data transfer—and explains why certain components may or may not be useful for particular needs. When it comes to data, Hadoop is a whole new ballgame, but with this handy reference, you'll have a good grasp of the playing field.

Topics include:

- **Core technologies**—Hadoop Distributed File System (HDFS), MapReduce, YARN, and Spark
- **Database and data management**—Cassandra, HBase, MongoDB, and Hive
- **Serialization**—Avro, JSON, and Parquet
- **Management and monitoring**—Puppet, Chef, Zookeeper, and Oozie
- **Analytic helpers**—Pig, Mahout, and MLLib
- **Data transfer**—Scoop, Flume, distcp, and Storm
- **Security, access control, auditing**—Sentry, Kerberos, and Knox
- **Cloud computing and virtualization**—Serengeti, Docker, and Whirr

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies

By Kevin Sitto, Marshall Presser

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser

If your organization is about to enter the world of big data, you not only need to decide whether Apache Hadoop is the right platform to use, but also which of its many components are best suited to your task. This field guide makes the exercise manageable by breaking down the Hadoop ecosystem into short, digestible sections. You'll quickly understand how Hadoop's projects, subprojects, and related technologies work together.

Each chapter introduces a different topic—such as core technologies or data transfer—and explains why certain components may or may not be useful for particular needs. When it comes to data, Hadoop is a whole new ballgame, but with this handy reference, you'll have a good grasp of the playing field.

Topics include:

- **Core technologies**—Hadoop Distributed File System (HDFS), MapReduce, YARN, and Spark
- **Database and data management**—Cassandra, HBase, MongoDB, and Hive
- **Serialization**—Avro, JSON, and Parquet
- **Management and monitoring**—Puppet, Chef, Zookeeper, and Oozie
- **Analytic helpers**—Pig, Mahout, and MLLib
- **Data transfer**—Scoop, Flume, distcp, and Storm
- **Security, access control, auditing**—Sentry, Kerberos, and Knox
- **Cloud computing and virtualization**—Serengeti, Docker, and Whirr

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser Bibliography

- Sales Rank: #478384 in eBooks
- Published on: 2015-03-02
- Released on: 2015-03-02
- Format: Kindle eBook

Download and Read Free Online Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser

Editorial Review

About the Author

Kevin Sitto is a Field Solutions Engineer with Pivotal Software, providing consulting services to help folks understand and address their big data needs.

He lives in Maryland with his wife and two kids and enjoys making homebrew beer when he's not writing books about big data.

Marshall Presser is a Field Chief Technology Officer for Pivotal and is based in McLean VA. In addition to helping customers solve complex analytic problems with the Greenplum Database, he leads the Hadoop Virtual Field Team, working on issues of integrating Hadoop with relational databases.

Prior to coming to Pivotal (formerly Greenplum), he spent 12 years at Oracle, specializing in High Availability, Business Continuity, Clustering, Parallel Database Technology, Disaster Recovery and Large Scale Database Systems. Marshall has also worked for a number of hardware vendors implementing clusters and other parallel architectures. His background includes parallel computation, operating system and compiler development as well as private consulting for organizations in heath care, financial services, and federal and state governments.

Marshall holds a B.A in Mathematics and an M.A. in Economics and Statistics from the University of Pennsylvania and a M.Sc. in Computing from Imperial College, London.

Users Review

From reader reviews:

Ruth Irizarry:

Do you have favorite book? If you have, what is your favorite's book? E-book is very important thing for us to know everything in the world. Each publication has different aim or even goal; it means that publication has different type. Some people sense enjoy to spend their time to read a book. They are really reading whatever they acquire because their hobby will be reading a book. Think about the person who don't like reading a book? Sometime, person feel need book if they found difficult problem or even exercise. Well, probably you'll have this Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies.

Holly Hughes:

As people who live in the modest era should be up-date about what going on or info even knowledge to make these individuals keep up with the era which is always change and move ahead. Some of you maybe

will update themselves by examining books. It is a good choice for you personally but the problems coming to you actually is you don't know which you should start with. This Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies is our recommendation so you keep up with the world. Why, because this book serves what you want and need in this era.

Heather Robertson:

Now a day people that Living in the era wherever everything reachable by connect to the internet and the resources within it can be true or not require people to be aware of each info they get. How people have to be smart in obtaining any information nowadays? Of course the answer then is reading a book. Examining a book can help persons out of this uncertainty Information specifically this Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies book because this book offers you rich data and knowledge. Of course the data in this book hundred per-cent guarantees there is no doubt in it you know.

Gregory Kile:

In this period of time globalization it is important to someone to get information. The information will make professionals understand the condition of the world. The healthiness of the world makes the information quicker to share. You can find a lot of sources to get information example: internet, magazine, book, and soon. You can observe that now, a lot of publisher this print many kinds of book. Typically the book that recommended for your requirements is Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies this book consist a lot of the information from the condition of this world now. This kind of book was represented how can the world has grown up. The vocabulary styles that writer use for explain it is easy to understand. The particular writer made some research when he makes this book. Honestly, that is why this book suitable all of you.

Download and Read Online Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser #TF5DVHRK41O

Read Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser for online ebook

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser 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 Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser books to read online.

Online Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser ebook PDF download

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser Doc

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser Mobipocket

Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser EPub

TF5DVHRK41O: Field Guide to Hadoop: An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies By Kevin Sitto, Marshall Presser