The book was found

Hadoop Operations

A Guide for Developers and Administrators

Hadoop Operations

O'REILLY®

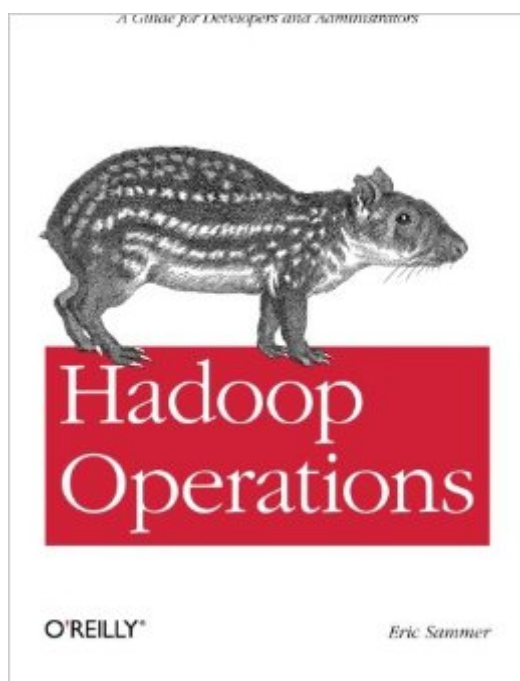Eric Sammer

Synopsis

If you've been asked to maintain large and complex Hadoop clusters, this book is a must. Demand for operations-specific material has skyrocketed now that Hadoop is becoming the de facto standard for truly large-scale data processing in the data center. Eric Sammer, Principal Solution Architect at Cloudera, shows you the particulars of running Hadoop in production, from planning, installing, and configuring the system to providing ongoing maintenance.Rather than run through all possible scenarios, this pragmatic operations guide calls out what works, as demonstrated in critical deployments.Get a high-level overview of HDFS and MapReduce: why they exist and how they work Plan a Hadoop deployment, from hardware and OS selection to network requirements Learn setup and configuration details with a list of critical properties Manage resources by sharing a cluster across multiple groups Get a runbook of the most common cluster maintenance tasks Monitor Hadoop clusters--and learn troubleshooting with the help of real-world war stories Use basic tools and techniques to handle backup and catastrophic failure

Book Information

Paperback: 298 pages

Publisher: O'Reilly Media; 1 edition (October 19, 2012)

Language: English

ISBN-10: 1449327052

ISBN-13: 978-1449327057

Product Dimensions: 7 x 0.8 x 9.2 inches

Shipping Weight: 1.2 pounds (View shipping rates and policies)

Average Customer Review: 4.6 out of 5 stars See all reviews (19 customer reviews)

Best Sellers Rank: #265,168 in Books (See Top 100 in Books)   #23 in Books > Computers & Technology > Programming > Parallel Programming   #68 in Books > Computers & Technology > Networking & Cloud Computing > Cloud Computing   #142 in Books > Computers & Technology > Databases & Big Data > Data Mining

Customer Reviews

I recently received this book and, having experience with Hadoop, ripped through it cover-to-cover -- that is, you cantake this review as indicative of the entire book.Whether the topic is HDFS and how data is ingested and replicated, or how Map/Reduce "finds" the most suitablenode to run it's tasks on, or what the cost and performance advantages are of adopting the shared-nothing, commoditymodel recommended for Hadoop clusters, etc., etc., etc., this book provides the how,

what, when, where and why ofHadoop (the missing manual, of sorts).Cluster Administrators as well as Map/Reduce programmers benefit from it's through, no-shortcuts-taken, breakdown ofthe Hadoop platform. I highly recommend it.

As a developer looking to support devops it was a great read. Obviously with any book on hadoop time is not kind and while this book cover hadoop 2.0 AND mentions future work it predates hadoop 2.2 so some information or needs will not be met.That said, still tons of good information here on how hadoop works and on topics like security and monitoring.

My "big data" (getting tired of a buzzword a little bit ...) project is now moving into a production phase where our customer will be deploying a Hadoop and other related technologies into their production data center.This book could not have come into better time as production team is look to both contract the support team and to have a manual for Hadoop operations.There are two things that I like about this book:1. It covers all of the topics that matter. It covers most important aspects of the Hadoop platform and its architecture but from the operational perspective - HDFS architecture and cluster configuration, MapReduce and YARN execution models, cluster setup and most importantly a very detailes review of options and recommendations related to operating system, network and storage setup.2. There are dedicated chapters to cluster maintenance, backups, monitoring and, very importantly, troubleshooting that go into very solid level of details on many of the problems or intricacies that one should better know about Hadoop in an operational setting.These chapters are obviously written by someone who ran Hadoop many times before and in a large, production setting.War stories and "mystery bottleneck" sections are great.In summary, right book in a right time, although I feel we should have had similar book maybe a year ago. I am guessing that Cloudera wanted to get their solid cut first at consulting and support fees before making such material available ;-) (author Eric Sammer is Cloudera's solutions architect)

Very good book, easy read.Mainly its for Admins, not much stuff for developers.Great high level overview if you are tasked to deploy production cluster (requirements, deployment,tuning,monitoring,backup..)Even if you don't need to deploy Hadoop cluster its great read from a system engineer perspective, talks about hight availability concepts, network bandwidth, memory limitations... Very good explained, first tells what are the limitations then how to outcome them. Also explains what were limitations in earlier hadoop versions and how they were resolved in Hadoop 2 (high availability, feferation)I would not say this book is suitable for developers, unleast

they want to learn some of the admin tasks. Just high overview of MapReduce is explained, but nothing from programing perspective. No other projects explained that are often used with Hadoop like, Hive, Hbase, Pig...

Before you can do anything useful with Hadoop you have to set it up and tune it. Not a simple task. This book will help get you to step one and beyond. You can read it cover to cover and it also makes for a decent reference on many important topics. I've been immersed in Hadoop for nearly three years and still found lots of new information as well as solid reinforcements of prior knowledge. Chapter 4, "Planning a Hadoop Cluster" is full of good information for those from the old school who are not used to the idea of Hadoop being designed from the ground up to run on commodity hardware. Don't argue with them, just make them read that chapter. The author says it best when he states "Exotic deployments of Hadoop usually end in exotic results, and not in a good way." A must have for any Hadoop administrator.

This book is targeted at Administrators but is definitely useful for any developer. The author kept his shoes firm on the topic without any digressions. The author explains topics in an easy to understand tone. He dealt with real issues faced by real people when deploying/maintaning hadoop clusters. Would definitely recommend.

Hadoop Operations by Eric Sammer is marvelous book which explains almost each bit of information is very lucid manner. AS name suggests, book is for operations guys - How data is ingested and replicated, or how MapReduce "finds" the most suitablenode to run parts of job, or what the cost and performance advantages are of adopting the shared-nothing, commodity hardware model recommended for Hadoop cluster, etc.This book is for Operations guys/Administrators and as good supporting material to developers.

Download to continue reading...

Big Data, MapReduce, Hadoop, and Spark with Python: Master Big Data Analytics and Data Wrangling with MapReduce Fundamentals using Hadoop, Spark, and Python 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) Hadoop Operations Water Treatment WSO: Principles and Practices of Water Supply Operations Volume 1 (Water Supply Operations Series) Maingot's Abdominal Operations, 12th Edition (Zinner, Maingot's Abdominal Operations) Maingot's Abdominal Operations (Zinner, Maingot's Abdominal Operations) Hadoop: The Definitive Guide Understanding AS/400Â® System Operations Building a Digital Analytics Organization: Create Value by Integrating Analytical Processes, Technology, and People into Business Operations (FT Press Analytics) GI Collector's Guide, Vol. 2: U.S. Army European Theater of Operations (Hardcover) Urban Operations, 2006, FM 3-06, Field Manual No. 3-06, Military Manual GI Collector's Guide, Vol. 2: U.S. Army European Theater of Operations Modern Communications Receiver Design and Technology (Artech House Intelligence and Information Operations) Linux Operations and Administration Wastewater Treatment Plant Operations Made Easy: A Practical Guide for Licensure Hazardous Waste Operations and Emergency Response Manual Essentials of Fire Fighting and Fire Department Operations (6th Edition)