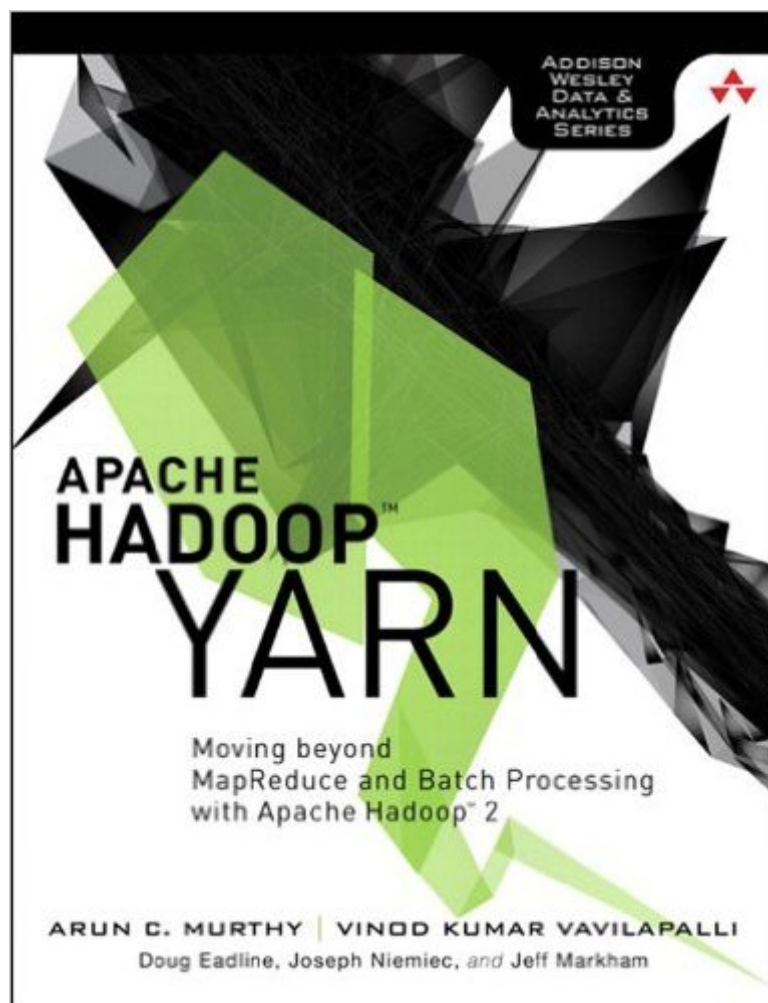# Apache Hadoop YARN: Moving Beyond MapReduce And Batch Processing With Apache Hadoop 2 (Addison-Wesley Data & Analytics Series)

Synopsis

â œThis book is a critically needed resource for the newly released Apache Hadoop 2.0, highlighting YARN as the significant breakthrough that broadens Hadoop beyond the MapReduce paradigm.â • â "From the Foreword by Raymie Stata, CEO of Altiscale   The Insiderâ ™s Guide to Building Distributed, Big Data Applications with Apache Hadoopâ„¢ YARN Â Apache Hadoop is helping drive the Big Data revolution. Now, its data processing has been completely overhauled: Apache Hadoop YARN provides resource management at data center scale and easier ways to create distributed applications that process petabytes of data. And now in   Apache Hadoopâ„¢ YARN,   two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances. Â YARN project founder Arun Murthy and project lead Vinod Kumar Vavilapalli demonstrate how YARN increases scalability and cluster utilization, enables new programming models and services, and opens new options beyond Java and batch processing. They walk you through the entire YARN project lifecycle, from installation through deployment. Â Youâ ™ll find many examples drawn from the authorsâ ™ cutting-edge experienceâ "first as Hadoopâ ™s earliest developers and implementers at Yahoo! and now as Hortonworks developers moving the platform forward and helping customers succeed with it. Â Coverage includes   YARNâ ™s goals, design, architecture, and componentsâ "how it expands the Apache Hadoop ecosystem   Exploring YARN on a single nodeÂ   Administering YARN clusters and Capacity SchedulerÂ   Running existing MapReduce applicationsÂ   Developing a large-scale clustered YARN applicationÂ   Discovering new open source frameworks that run under YARN

Book Information

File Size: 21405 KB

Print Length: 318 pages

Simultaneous Device Usage: Up to 5 simultaneous devices, per publisher limits

Publisher: Addison-Wesley Professional; 1 edition (March 14, 2014)

Publication Date: March 14, 2014

Sold by:Â Digital Services LLC

Language: English

ASIN: B00J0872UE

Text-to-Speech: Enabled

X-Ray: Not Enabled

Word Wise: Not Enabled

Lending: Not Enabled

Enhanced Typesetting: Enabled

Best Sellers Rank: #690,826 Paid in Kindle Store (See Top 100 Paid in Kindle Store)   #56 inÂ Books > Computers & Technology > Web Development & Design > Website Analytics   #168 inÂ Kindle Store > Kindle eBooks > Computers & Technology > Networking > Client-Server Systems   #505 inÂ Books > Computers & Technology > Networking & Cloud Computing > Data in the Enterprise > Client-Server Systems

Customer Reviews

I can't give this book a great rating. In so many places it took five, six, seven or more sentences to say what could be said in one or two well-written, forceful sentences. To some degree I can look past that, were it not for this: even after an unnecessary circumlocution, the point was never firmly made, if at all. Sadly, I had to take my understanding of YARN back to the internet, post my book purchase.

This book mostly provides introduction and theory on Hadoop with YARN features with installation guide.The book severely lacking hands-on examples to get the feeling of working on Hadoop. The books's web page says for last one month "We are in the process of uploading all the code from the book. Please come back again in a while! "[...]

Yes, this seems to have been rushed to print and, unfortunately, yes the proof-readers were not up to snuff.However, once you get over the strange sentences that don't quite read properly (and you will, by re-reading them multiple times), this is a great text for someone looking to get a grip on the understanding of YARN and how MRv2 (+ graph, Storm, Spark etc) fit into things.With the myriad of technologies and stacks now available in the "Big Data" space, Hadoop has finally realized the "grid" dreams of the early 2000's on commodity, on-demand clusters.

Was hoping for a lot more. As others have noted, this was banged together very quickly. You will learn more simply browsing the web and piecing together bits and pieces from there. Avoid this book.

This book is a decent introduction to the idea and principles of Yarn. It does not provide many technical details or code, but enough to understand what Yarn is about and how it works. On the

negative side, the terrible pseudo-scientific style of writing, verbose, vague and repetitive. Each statement is repeated 5-10 times in different paragraphs, which results in 300 pages of dense text. I believe it could be made 3 times shorter and much clearer if authors concentrated on the technical part instead of rhetorical beatifications.

I gave the book 3 stars because it has a good overview of Hadoop Yarn.I only gave it 3 stars because the "fast" installation scripts are so broken I gave up running them and just manually did what they are supposed to do. It took about 8 hours to get my first cluster operational.

I attended Arun Murthyâ ™s lectures before (at Hadoop Meetups, GITPRO etc.) . So I picked this up as soon as it came out.This book is a great logical extension to the â œHadoop:The Definitive Guideâ •. The concepts and architecture of various YARN components are detailed and well-illustrated . Considering that the book comes from the YARN architects , it cannot get any better than this. Some hands-on examples would have been very helpful. Those who want to start on Hadoop , start with the Definitive guide before moving to this.--RamaK

There aren't many books covering Hadoop 2.x (and thus YARN) out yet, but seeing as Arun C. Murthy is one of the original architects of Hadoop, I expected this to be a comprehensive real-world guide to deploying and using Hadoop 2.x and YARN.Unfortunately, it is not. It is more of a "beginner's guide", geared for those who just want to get started with Hadoop and aren't yet using it as a core part of their business. I reviewed another Hadoop 2.x book and perhaps was a bit more "kind" about it on this point but that was because I expected it would be more of a tutorial and it was.This book is well-written, well-organized, and its layout is slick and professional. That's no surprise as it's from a big-name publisher. It's comprehensive in its description of YARN's history, its design and architecture, its features, how it interacts with HDFS and MapReduce, and how it's supposed to be used with other tools and frameworks, and getting a *basic* Hadoop 2.x cluster up-and-running, (by several means, including Ambari).But you can learn almost all of this from the Hadoop project's website, and that info will also be more up-to-date.The big disappointment is that there's pretty much no useful info on, for example, how to configure Hadoop & YARN memory limits and tuning parameters, how to plan for and configure YARN scheduler queues, best-practices for deployment of more than a handful of nodes, or recommended JVM, OS, disk or network configuration. And forget about learning how to do anything at scale. And forget about hearing the hard truth about running and using hadoop in the real world.If you're a beginner to Hadoop and/or

just want a single tome from which you can learn about Hadoop 2.x and YARN from a somewhat high-level, this is a great book for you. But if you want to learn more than you can find online in an afternoon, keep moving. I'm sure something more useful will come out sooner or later.

Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics Series) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics) 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 Series) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics) Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data & Analytics Series) R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data and Analytics) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference (Addison-Wesley Data & Analytics) Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! First Principles of Discrete Systems and Digital Signal Processing (Addison-Wesley Series in Electrical Engineering) Principles of Compiler Design (Addison-Wesley series in computer science and information processing) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault The Design and Implementation of the 4.4 BSD Operating System (Addison-Wesley UNIX and Open Systems Series) Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Adobe Reader) (Addison-Wesley Signature Series (Fowler)) Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions (Addison-Wesley Signature Series (Fowler)) Circuits, Interconnections, and Packaging for Vlsi (Addison-Wesley VLSI systems series) Patterns of Enterprise Application Architecture (Addison-Wesley Signature Series (Fowler)) TCP/IP Illustrated, Vol. 1: The Protocols (Addison-Wesley Professional Computing Series)