Apache Hadoop YARN: Moving Beyond MapReduce And Batch Processing With Apache Hadoop 2 (Addison-Wesley Data & Analytics Series)
This 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. — From the Foreword by Raymie Stata, CEO of Altiscale

The Insider’s Guide to Building Distributed, Big Data Applications with Apache Hadoop’s YARN

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 Hadoop’sYARN, two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances. 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. You’ll find many examples drawn from the authors’ cutting-edge experience as Hadoop’s earliest developers and implementers at Yahoo! and now as Hortonworks developers moving the platform forward and helping customers succeed with it.

Coverage includes YARN’s goals, design, architecture, and components; how it expands the Apache Hadoop ecosystem; exploring YARN on a single node; administering YARN clusters and Capacity Scheduler; running existing MapReduce applications; developing a large-scale clustered YARN application; 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: Digital Services LLC
Language: English
ASIN: B00J0872UE
Text-to-Speech: Enabled
X-Ray: Not Enabled
Word Wise: Not Enabled
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 book’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 Murthy’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 Hadoop: The Definitive Guide. 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.

Download to continue reading...


Dmca