Apache Oozie: The Workflow Scheduler for Hadoop

By Mohammad Kamrul Islam, Aravind Srinivasan

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan

Get a solid grounding in Apache Oozie, the workflow scheduler system for managing Hadoop jobs. With this hands-on guide, two experienced Hadoop practitioners walk you through the intricacies of this powerful and flexible platform, with numerous examples and real-world use cases.

Once you set up your Oozie server, you'll dive into techniques for writing and coordinating workflows, and learn how to write complex data pipelines. Advanced topics show you how to handle shared libraries in Oozie, as well as how to implement and manage Oozie's security capabilities.

- Install and configure an Oozie server, and get an overview of basic concepts
- Journey through the world of writing and configuring workflows
- Learn how the Oozie coordinator schedules and executes workflows based on triggers
- Understand how Oozie manages data dependencies
- Use Oozie bundles to package several coordinator apps into a data pipeline
- Learn about security features and shared library management
- Implement custom extensions and write your own EL functions and actions
- Debug workflows and manage Oozie's operational details

⬇ **Download** Apache Oozie: The Workflow Scheduler for Hadoop ...pdf

📄 **Read Online** Apache Oozie: The Workflow Scheduler for Hadoop ...pdf

Apache Oozie: The Workflow Scheduler for Hadoop

By Mohammad Kamrul Islam, Aravind Srinivasan

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan

Get a solid grounding in Apache Oozie, the workflow scheduler system for managing Hadoop jobs. With this hands-on guide, two experienced Hadoop practitioners walk you through the intricacies of this powerful and flexible platform, with numerous examples and real-world use cases.

Once you set up your Oozie server, you'll dive into techniques for writing and coordinating workflows, and learn how to write complex data pipelines. Advanced topics show you how to handle shared libraries in Oozie, as well as how to implement and manage Oozie's security capabilities.

- Install and configure an Oozie server, and get an overview of basic concepts
- Journey through the world of writing and configuring workflows
- Learn how the Oozie coordinator schedules and executes workflows based on triggers
- Understand how Oozie manages data dependencies
- Use Oozie bundles to package several coordinator apps into a data pipeline
- Learn about security features and shared library management
- Implement custom extensions and write your own EL functions and actions
- Debug workflows and manage Oozie's operational details

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan Bibliography

- Rank: #1027977 in eBooks
- Published on: 2015-05-12
- Released on: 2015-05-12
- Format: Kindle eBook

Download and Read Free Online Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan

Editorial Review

About the Author

Mohammad Kamrul Islam is currently working at Uber in data engineering team as a Staff Software Engineer. Previously, he worked at Linkedin for more than two years as Staff Software Engineer in the Hadoop development team. Before that, he worked at Yahoo for nearly five years as an Oozie architect/technical lead. His fingerprints can befound all over Oozie and is a respected voice in the Oozie community. He has been intimately involved with the Apache Hadoop ecosystem since 2009. Mohammad has a Ph.D. in Computer Science with a specialization in parallel job scheduling from Ohio State University. He received his MSCS degree from Wright State University, Ohio andBSCS from Bangladesh University of Engineering and Technology (BUET). He is a Project Management Committee (PMC) member of both Apache Oozie and Apache TEZ and frequently contributes to Apache YARN/MapReduce and Apache Hive. He was elected as the PMC chair and Vice-President of Oozie as part of the Apache Software Foundation from 2013 through 2015.

Aravind Srinivasan has been involved with Hadoop in general and Oozie in particular since 2008. He is currently a Lead Application Architect at Altiscale, a Hadoop-as-a-service company, where he helps customers with Hadoop application design and architecture. His association with Big Data and Hadoop started during his time at Yahoo, where he spent almost six years working on various data pipelines for advertising systems. He has extensive experience building complicated, low latency data pipelines and also in porting legacy pipelines to Oozie. He drove a lot of Oozie's requirements as a customer in its early days of adoption inside Yahoo and later spent some time as a Product Manager in Yahoo's Hadoop team where he contributed further to Oozie's roadmap. He also spent a year after Yahoo at Think Big Analytics, a Hadoop consulting firm, where he got to consult on some interesting and challenging Big Data integration projects at Facebook. He has a Masters in Computer Science from Arizona State and lives in Silicon Valley.

Users Review

From reader reviews:

Kay Roberts:

As people who live in the actual modest era should be revise about what going on or information even knowledge to make these individuals keep up with the era which is always change and move ahead. Some of you maybe will certainly update themselves by studying books. It is a good choice for you personally but the problems coming to a person is you don't know what type you should start with. This Apache Oozie: The Workflow Scheduler for Hadoop is our recommendation to cause you to keep up with the world. Why, because book serves what you want and wish in this era.

Diane Merryman:

Do you among people who can't read satisfying if the sentence chained within the straightway, hold on guys this aren't like that. This Apache Oozie: The Workflow Scheduler for Hadoop book is readable through you who hate those straight word style. You will find the information here are arrange for enjoyable reading through experience without leaving possibly decrease the knowledge that want to give to you. The writer involving Apache Oozie: The Workflow Scheduler for Hadoop content conveys the thought easily to understand by many people. The printed and e-book are not different in the content but it just different such as it. So , do you continue to thinking Apache Oozie: The Workflow Scheduler for Hadoop is not loveable to be your top list reading book?

Suzanne Ferris:

Is it anyone who having spare time and then spend it whole day by watching television programs or just resting on the bed? Do you need something totally new? This Apache Oozie: The Workflow Scheduler for Hadoop can be the answer, oh how comes? A fresh book you know. You are thus out of date, spending your extra time by reading in this fresh era is common not a geek activity. So what these guides have than the others?

Shawn Mathison:

That reserve can make you to feel relax. This particular book Apache Oozie: The Workflow Scheduler for Hadoop was multi-colored and of course has pictures around. As we know that book Apache Oozie: The Workflow Scheduler for Hadoop has many kinds or genre. Start from kids until teens. For example Naruto or Private investigator Conan you can read and think that you are the character on there. Therefore not at all of book usually are make you bored, any it offers up you feel happy, fun and loosen up. Try to choose the best book in your case and try to like reading in which.

Download and Read Online Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan #TD7M0HZLW4Y

Read Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan for online ebook

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan 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 Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan books to read online.

Online Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan ebook PDF download

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan Doc

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan Mobipocket

Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan EPub

TD7M0HZLW4Y: Apache Oozie: The Workflow Scheduler for Hadoop By Mohammad Kamrul Islam, Aravind Srinivasan