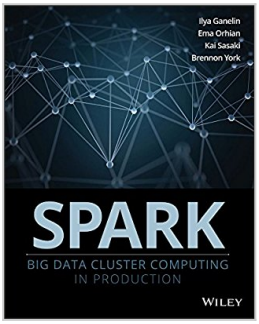

Spark: Big Data Cluster Computing in Production

By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York

Production-targeted Spark guidance with real-world use cases

Spark: Big Data Cluster Computing in Production goes beyond general Spark overviews to provide targeted guidance toward using lightning-fast big-data clustering in production. Written by an expert team well-known in the big data community, this book walks you through the challenges in moving from proof-of-concept or demo Spark applications to live Spark in production. Real use cases provide deep insight into common problems, limitations, challenges, and opportunities, while expert tips and tricks help you get the most out of Spark performance. Coverage includes Spark SQL, Tachyon, Kerberos, ML Lib, YARN, and Mesos, with clear, actionable guidance on resource scheduling, db connectors, streaming, security, and much more.

Spark has become the tool of choice for many Big Data problems, with more active contributors than any other Apache Software project. General introductory books abound, but this book is the first to provide deep insight and real-world advice on using Spark in production. Specific guidance, expert tips, and invaluable foresight make this guide an incredibly useful resource for real production settings.

- Review Spark hardware requirements and estimate cluster size
- Gain insight from real-world production use cases
- Tighten security, schedule resources, and fine-tune performance
- Overcome common problems encountered using Spark in production

Spark works with other big data tools including MapReduce and Hadoop, and uses languages you already know like Java, Scala, Python, and R. Lightning speed makes Spark too good to pass up, but understanding limitations and challenges in advance goes a long way toward easing actual production implementation. *Spark: Big Data Cluster Computing in Production* tells you everything you need to know, with real-world production insight and expert guidance, tips, and tricks.

Spark: Big Data Cluster Computing in Production

By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York

Production-targeted Spark guidance with real-world use cases

Spark: Big Data Cluster Computing in Production goes beyond general Spark overviews to provide targeted guidance toward using lightning-fast big-data clustering in production. Written by an expert team well-known in the big data community, this book walks you through the challenges in moving from proof-of-concept or demo Spark applications to live Spark in production. Real use cases provide deep insight into common problems, limitations, challenges, and opportunities, while expert tips and tricks help you get the most out of Spark performance. Coverage includes Spark SQL, Tachyon, Kerberos, ML Lib, YARN, and Mesos, with clear, actionable guidance on resource scheduling, db connectors, streaming, security, and much more.

Spark has become the tool of choice for many Big Data problems, with more active contributors than any other Apache Software project. General introductory books abound, but this book is the first to provide deep insight and real-world advice on using Spark in production. Specific guidance, expert tips, and invaluable foresight make this guide an incredibly useful resource for real production settings.

- Review Spark hardware requirements and estimate cluster size
- Gain insight from real-world production use cases
- Tighten security, schedule resources, and fine-tune performance
- Overcome common problems encountered using Spark in production

Spark works with other big data tools including MapReduce and Hadoop, and uses languages you already know like Java, Scala, Python, and R. Lightning speed makes Spark too good to pass up, but understanding limitations and challenges in advance goes a long way toward easing actual production implementation. *Spark: Big Data Cluster Computing in Production* tells you everything you need to know, with real-world production insight and expert guidance, tips, and tricks.

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York Bibliography

- Sales Rank: #967305 in eBooks
- Published on: 2016-03-28
- Released on: 2016-03-28
- Format: Kindle eBook

Editorial Review

From the Back Cover

TIPS, TRICKS, AND SOLUTIONS FOR USING SPARK IN PRODUCTION

Spark's popularity means the field is expanding—in terms of both use and capability. Faster than Hadoop and MapReduce, but compatible with Java®, Scala, Python®, and R, this open source clustering framework is becoming a must-have skill. *Spark: Big Data Cluster Computing in Production* goes beyond the basics to show you how to bring Spark to real-world production environments. With expert instruction, real-life use cases, and frank discussion, this guide helps you move past the challenges and bring proof-of-concept Spark applications live.

- Fine-tune your Spark app to run on production data
- Manage resources, organize storage, and master monitoring
- Learn about potential problems from real-world use cases, and see where Spark fits best
- Estimate cluster size and nail down hardware requirements
- Tune up performance with memory management, partitioning, shuffling, and more
- Ensure data security with Kerberos
- Head off Spark streaming problems in production
- Integrate Spark with Yarn, Mesos, Tachyon, and more

About the Author

Ilya Ganelin is a data engineer working at Capital One Data Innovation Lab. Ilya is an active contributor to the core components of Apache Spark and a committer to Apache Apex.

Ema Orhian is a Big Data Engineer interested in scaling algorithms. She is the main committer on jaws-spark-sql-rest, a data warehouse explorer on top of Spark SQL.

Kai Sasaki is a software engineer working in distributed computing and machine learning. He is a Spark contributor who develops mainly MLlib, ML libraries.

Brennon York has been a core contributor to Apache Spark since 2014 including development on GraphX and the core build environment.

Users Review

From reader reviews:

Ruth Cook:

Book is to be different for each grade. Book for children until finally adult are different content. As we know that book is very important normally. The book Spark: Big Data Cluster Computing in Production has been making you to know about other information and of course you can take more information. It doesn't matter what advantages for you. The book Spark: Big Data Cluster Computing in Production is not only giving you more new information but also to become your friend when you truly feel bored. You can spend your current

spend time to read your publication. Try to make relationship together with the book Spark: Big Data Cluster Computing in Production. You never truly feel lose out for everything in case you read some books.

Adrian Kao:

The knowledge that you get from Spark: Big Data Cluster Computing in Production is a more deep you digging the information that hide into the words the more you get thinking about reading it. It does not mean that this book is hard to recognise but Spark: Big Data Cluster Computing in Production giving you joy feeling of reading. The writer conveys their point in particular way that can be understood through anyone who read this because the author of this guide is well-known enough. This book also makes your own personal vocabulary increase well. So it is easy to understand then can go along with you, both in printed or e-book style are available. We propose you for having that Spark: Big Data Cluster Computing in Production instantly.

Christopher Wilkerson:

The book with title Spark: Big Data Cluster Computing in Production has a lot of information that you can discover it. You can get a lot of gain after read this book. This particular book exist new expertise the information that exist in this e-book represented the condition of the world today. That is important to yo7u to know how the improvement of the world. This book will bring you throughout new era of the the positive effect. You can read the e-book with your smart phone, so you can read it anywhere you want.

John Cheung:

A lot of people always spent their free time to vacation or go to the outside with them household or their friend. Do you know? Many a lot of people spent they free time just watching TV, as well as playing video games all day long. If you would like try to find a new activity that is look different you can read the book. It is really fun for yourself. If you enjoy the book that you simply read you can spent the entire day to reading a publication. The book Spark: Big Data Cluster Computing in Production it is rather good to read. There are a lot of people that recommended this book. We were holding enjoying reading this book. Should you did not have enough space to create this book you can buy often the e-book. You can m0ore effortlessly to read this book out of your smart phone. The price is not to fund but this book features high quality.

Read Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York for online ebook

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York 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 Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York books to read online.

Online Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York ebook PDF download

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York Doc

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York Mobipocket

Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York EPub