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: #352822 in Books
- Brand: Wiley
- Published on: 2016-03-21
- Original language: English
- Number of items: 1
- Dimensions: 9.30" h x .50" w x 7.40" l, .0 pounds
- Binding: Paperback
- 216 pages

Download and Read Free Online Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York

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:

John Pasko:

Have you spare time to get a day? What do you do when you have more or little spare time? Yep, you can choose the suitable activity for spend your time. Any person spent all their spare time to take a move, shopping, or went to often the Mall. How about open or even read a book titled Spark: Big Data Cluster Computing in Production? Maybe it is being best activity for you. You recognize beside you can spend your time together with your favorite's book, you can more intelligent than before. Do you agree with their

opinion or you have other opinion?

Kim Marshall:

As people who live in the modest era should be upgrade about what going on or info even knowledge to make these individuals keep up with the era which is always change and progress. Some of you maybe may update themselves by reading books. It is a good choice for you personally but the problems coming to you is you don't know what type you should start with. This Spark: Big Data Cluster Computing in Production is our recommendation to help you keep up with the world. Why, since this book serves what you want and need in this era.

Holly Sheehan:

Playing with family in the park, coming to see the marine world or hanging out with close friends is thing that usually you have done when you have spare time, then why you don't try point that really opposite from that. One particular activity that make you not feeling tired but still relaxing, trilling like on roller coaster you have been ride on and with addition info. Even you love Spark: Big Data Cluster Computing in Production, you may enjoy both. It is excellent combination right, you still need to miss it? What kind of hang type is it? Oh seriously its mind hangout folks. What? Still don't buy it, oh come on its referred to as reading friends.

Tanya Caggiano:

Your reading sixth sense will not betray an individual, why because this Spark: Big Data Cluster Computing in Production publication written by well-known writer who knows well how to make book which might be understand by anyone who else read the book. Written within good manner for you, dripping every ideas and writing skill only for eliminate your own personal hunger then you still uncertainty Spark: Big Data Cluster Computing in Production as good book not just by the cover but also from the content. This is one guide that can break don't assess book by its deal with, so do you still needing an additional sixth sense to pick this kind of!? Oh come on your studying sixth sense already alerted you so why you have to listening to yet another sixth sense.

Download and Read Online Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York #F5J7I2EKSO3

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

F5J7I2EKSO3: Spark: Big Data Cluster Computing in Production By Ilya Ganelin, Ema Orhian, Kai Sasaki, Brennon York