BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER
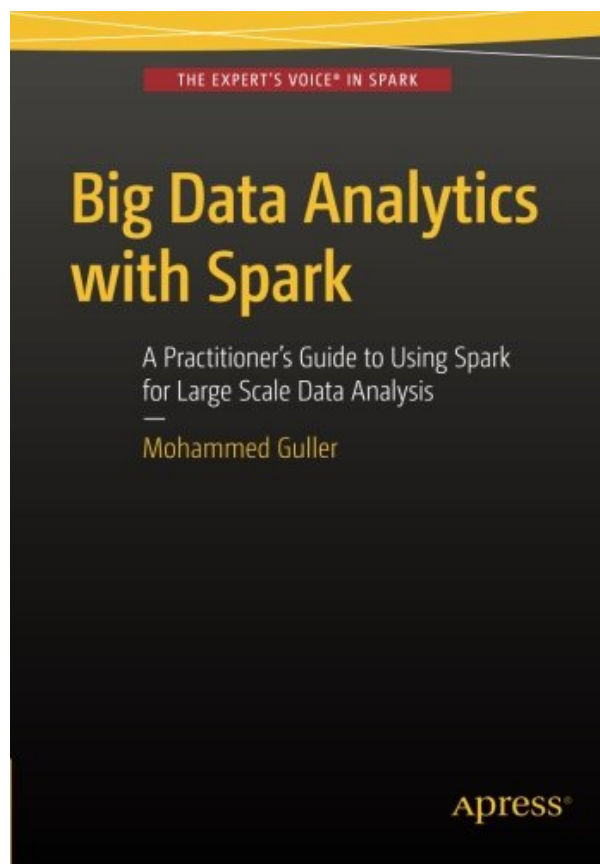
THE EXPERT'S VOICE® IN SPARK

Big Data Analytics with Spark

A Practitioner's Guide to Using Spark for Large Scale Data Analysis

—

Mohammed Guller
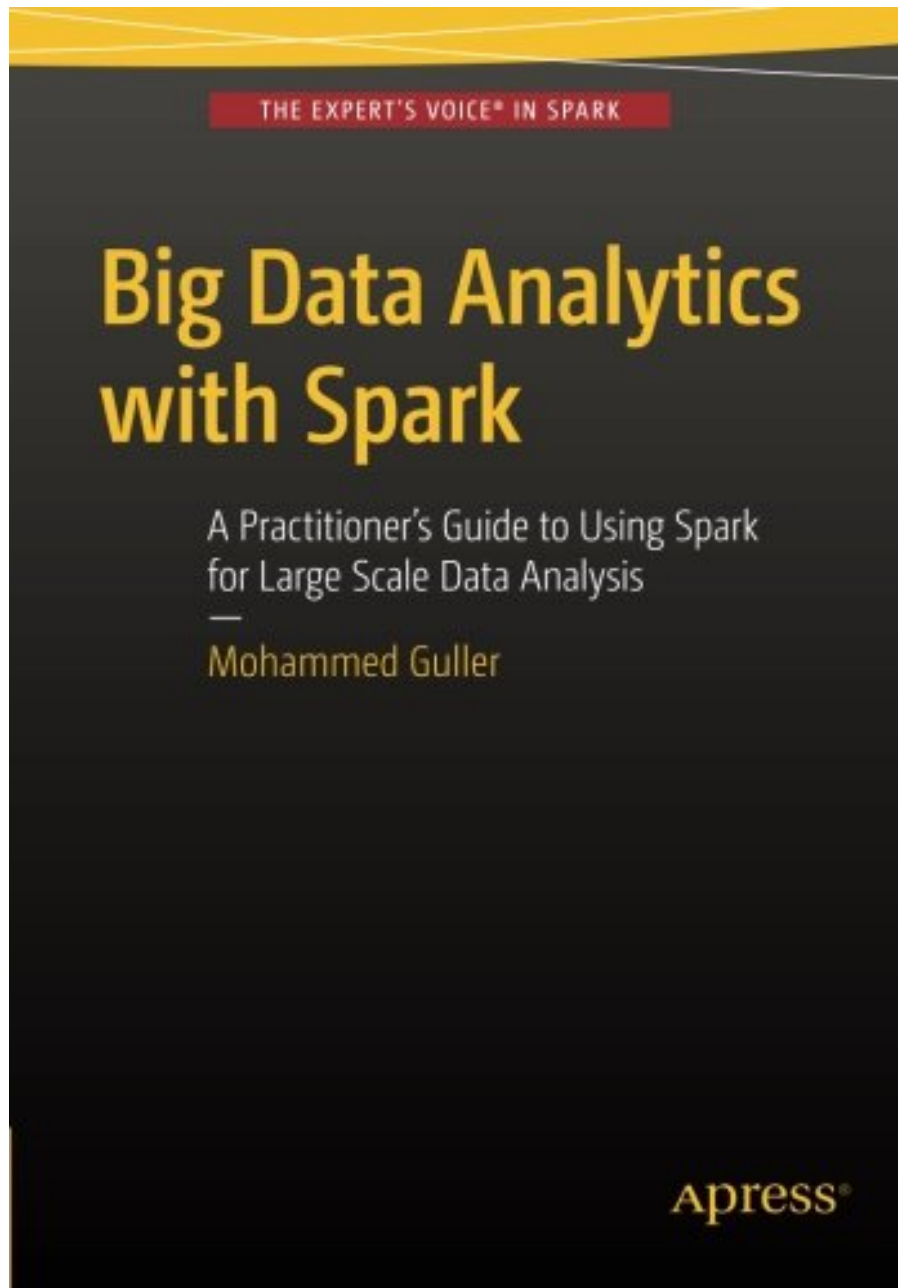
Apress®

Click link bellow and free register to download ebook:
 BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER

[DOWNLOAD FROM OUR ONLINE LIBRARY](#)

BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER PDF

It will not take even more time to obtain this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller It will not take more cash to publish this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller Nowadays, individuals have been so smart to utilize the modern technology. Why don't you utilize your device or other tool to conserve this downloaded soft documents book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller In this manner will let you to constantly be gone along with by this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller Obviously, it will certainly be the best buddy if you read this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller until completed.

About the Author

Mohammed Guller is the principal architect at Glassbeam, where he leads the development of advanced and predictive analytics products. He is a big data and Spark expert. He is frequently invited to speak at big data–related conferences. He is passionate about building new products, big data analytics, and machine learning.

Over the last 20 years, Mohammed has successfully led the development of several innovative technology products from concept to release. Prior to joining Glassbeam, he was the founder of TrustRecs.com, which he started after working at IBM for five years. Before IBM, he worked in a number of hi-tech start-ups, leading new product development.

Mohammed has a master's of business administration from the University of California, Berkeley, and a master's of computer applications from RCC, Gujarat University, India.

BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER PDF

When you are hurried of task due date as well as have no suggestion to obtain motivation, **Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller** book is one of your options to take. Schedule Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller will certainly offer you the best resource as well as thing to get motivations. It is not only about the tasks for politic company, administration, economics, and other. Some bought jobs to make some fiction jobs additionally need motivations to get rid of the job. As just what you need, this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller will probably be your option.

It is not secret when hooking up the creating abilities to reading. Checking out *Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller* will make you obtain even more sources as well as resources. It is a manner in which can boost exactly how you overlook as well as understand the life. By reading this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller, you could greater than exactly what you get from various other book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller This is a popular publication that is published from popular publisher. Seen type the writer, it can be trusted that this book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller will certainly give several inspirations, about the life and also experience and every little thing within.

You might not should be uncertainty about this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller It is simple means to obtain this book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller You can just see the established with the link that we supply. Here, you can acquire guide Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller by on the internet. By downloading Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller, you can locate the soft documents of this publication. This is the exact time for you to start reading. Even this is not printed book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller; it will exactly provide even more benefits. Why? You may not bring the published book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller or stack guide in your residence or the workplace.

BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER PDF

Big Data Analytics with Spark is a step-by-step guide for learning Spark, which is an open-source fast and general-purpose cluster computing framework for large-scale data analysis. You will learn how to use Spark for different types of big data analytics projects, including batch, interactive, graph, and stream data analysis as well as machine learning. In addition, this book will help you become a much sought-after Spark expert.

Spark is one of the hottest Big Data technologies. The amount of data generated today by devices, applications and users is exploding. Therefore, there is a critical need for tools that can analyze large-scale data and unlock value from it. Spark is a powerful technology that meets that need. You can, for example, use Spark to perform low latency computations through the use of efficient caching and iterative algorithms; leverage the features of its shell for easy and interactive Data analysis; employ its fast batch processing and low latency features to process your real time data streams and so on. As a result, adoption of Spark is rapidly growing and is replacing Hadoop MapReduce as the technology of choice for big data analytics.

This book provides an introduction to Spark and related big-data technologies. It covers Spark core and its add-on libraries, including Spark SQL, Spark Streaming, GraphX, and MLlib. Big Data Analytics with Spark is therefore written for busy professionals who prefer learning a new technology from a consolidated source instead of spending countless hours on the Internet trying to pick bits and pieces from different sources.

The book also provides a chapter on Scala, the hottest functional programming language, and the program that underlies Spark. You'll learn the basics of functional programming in Scala, so that you can write Spark applications in it.

What's more, Big Data Analytics with Spark provides an introduction to other big data technologies that are commonly used along with Spark, like Hive, Avro, Kafka and so on. So the book is self-sufficient; all the technologies that you need to know to use Spark are covered. The only thing that you are expected to know is programming in any language.

There is a critical shortage of people with big data expertise, so companies are willing to pay top dollar for people with skills in areas like Spark and Scala. So reading this book and absorbing its principles will provide a boost?possibly a big boost?to your career.

- Sales Rank: #201081 in Books
- Published on: 2015-11-25
- Released on: 2015-12-30

- Original language: English
- Number of items: 1
- Dimensions: 10.00" h x .69" w x 7.00" l, .0 pounds
- Binding: Paperback
- 277 pages

About the Author

Mohammed Guller is the principal architect at Glassbeam, where he leads the development of advanced and predictive analytics products. He is a big data and Spark expert. He is frequently invited to speak at big data–related conferences. He is passionate about building new products, big data analytics, and machine learning.

Over the last 20 years, Mohammed has successfully led the development of several innovative technology products from concept to release. Prior to joining Glassbeam, he was the founder of TrustRecs.com, which he started after working at IBM for five years. Before IBM, he worked in a number of hi-tech start-ups, leading new product development.

Mohammed has a master's of business administration from the University of California, Berkeley, and a master's of computer applications from RCC, Gujarat University, India.

Most helpful customer reviews

4 of 4 people found the following review helpful.
Good (but very basic) book. More breath than depth
By Abhishek Srivastava
I liked the book. it nice and simple and good for beginners.

One issue I found is that this book covers the the entire spark eco-system (spark core, spark sql, spark streaming, mllib) in a very brief way. it doesn't go deep into any of the topics. This makes it very similar to the spark programming guide on the web which also adopts a similar pattern.

What would have been nice if the author would have gone deep into spark-core. Upon reading this book, it didn't teach me anything more than the spark programming guide

5 of 6 people found the following review helpful.
If you want to learn Spark, buy this book. Highly recommended
By Ian Stirk
Hi,

I have written a detailed chapter-by-chapter review of this book on www DOT i-programmer DOT info, the first and last parts of this review are given here. For my review of all chapters, search i-programmer DOT info for STIRK together with the book's title.

This book aims to provide a "...concise and easy-to-understand tutorial for big data and Spark". How does it

fare?

Spark is increasing the tool of choice for big data processing, being much faster than Hadoop's MapReduce. After putting Spark into a big data context, the book aims to cover Spark's core library, together with its more specialized libraries for Streaming, Machine Learning, SQL, and Graphing.

The book is aimed at developers that are new to Spark, some general background programming knowledge required, but little else.

Chapter 1 Big Data Technology Landscape

This chapter opens with a discussion about the current big data age, with data as the lifeblood of organizations, and growing exponentially. The standard 3Vs definition of big data is explored (velocity, variety, volume). Traditional relational database management systems (RDBMS) are unable to process these large volumes in a timely manner – this is where the scalability of big data systems comes into its own.

Next, the chapter discusses some technologies that are either used with Spark, or Spark competes with. The first technology is Hadoop, this is fault tolerant and scalable, and runs on commodity hardware. The three major components of Hadoop are discussed: YARN (Yet Another Resource Negotiator), MapReduce (distributed processing model), and HDFS (Hadoop Distributed File System). Spark is increasingly being used in place of MapReduce owning to its faster speed. The section briefly discusses Hive, a data warehouse with a SQL like interface, Spark SQL is expected to supersede Hive on many systems.

The chapter continues with a look at some common binary formats for serializing (storing on disk) big data, and their pros and cons. Specifically Avro, Thrift, Protocol Buffers, and SequenceFile are examined. Next, some column storage formats, which have performance advantages when the client requires a subset of columns, were briefly discussed, namely: RCFile, ORC, and Parquet.

Then a brief overview of messaging systems is provided, together with the advantages of having a layer of abstraction between producers and consumers. Specifically, Kafka and ZeroMQ are discussed with the aid of useful supporting diagrams.

NoSQL is then examined. The various types of NoSQL databases have different aims to the traditional RDBMS, typically trading Atomicity, Consistency, Isolation, Durability (ACID) for scalability and flexibility. The specific NoSQL databases briefly discussed are Cassandra and HBase. I sometimes wonder if it is meaningful to group NoSQL databases together. Is it meaningful to divide sports into Football and NoFootball? Are all the NoFootball sports meaningful as a group?

The chapter ends with a look at some distributed SQL query engines, these do not use MapReduce batch jobs, and are thus more oriented to interactive querying. The engines briefly examined are: Impala, Presto, and Apache Drill.

This chapter provides an excellent overview of big data technology. It should be noted there are many more technologies than described, but the examples given are sufficient to explain the topic areas. This is possibly the best backgrounder to big data I've read.

The discussions are very well written, concise and clear, with helpful diagrams, and no wasted words. There's a good flow between the topics, and useful links between chapters. There are website links for further information. These traits apply to all the chapters in the book.

.
.
.
Conclusion

This book aims to provide a "...concise and easy-to-understand tutorial for big data and Spark", and clearly succeeds. The book is exceptionally well written. Helpful explanations, diagrams, practical step-by-step walkthroughs, annotated code, inter-chapter links, and website links abound throughout.

The book is aimed at developers that are new to Spark, and explains concepts from the beginning. If you work through the book you should become competent in the use of Spark, there is much more to learn of course, but this book gives a solid foundation in both core Spark and its major specialized libraries: Streaming, Machine Learning, SQL, and Graphing.

The book is based on workshops given by the author, and clearly the feedback from these has been useful in creating this book, since it seems to have answered all the questions I had.

This book provides everything you need to know to get started with Spark, explained in an easy-to-follow manner. If you want to learn Spark, buy this book. Highly recommended

2 of 2 people found the following review helpful.
Well Organized and informative book
By Technocrat
This book is a very well written definitive overview of Spark. This is a great book for those who want to learn about spark but dont know where to start from.
Fundamentals are very well explained in the book for developers who are new to spark. It starts with great overview of big data technology and helps in building basics and then moves on to explore more advanced topics. The book covers Spark core and its specialized add-on libraries too. This book also contains a plenty of sample examples which are really useful. Even if you are new to the subject this book has enough information to get a developer started on spark projects.
In short this book is really well organized, very informative and easy to follow. Highly recommended .

See all 8 customer reviews...

BIG DATA ANALYTICS WITH SPARK: A PRACTITIONER'S GUIDE TO USING SPARK FOR LARGE SCALE DATA ANALYSIS BY MOHAMMED GULLER PDF

You could carefully include the soft file **Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller** to the gizmo or every computer hardware in your workplace or residence. It will certainly help you to constantly continue checking out Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller each time you have extra time. This is why, reading this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller doesn't offer you issues. It will provide you important sources for you which intend to begin creating, covering the similar book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller are various book area.

About the Author

Mohammed Guller is the principal architect at Glassbeam, where he leads the development of advanced and predictive analytics products. He is a big data and Spark expert. He is frequently invited to speak at big data–related conferences. He is passionate about building new products, big data analytics, and machine learning.

Over the last 20 years, Mohammed has successfully led the development of several innovative technology products from concept to release. Prior to joining Glassbeam, he was the founder of TrustRecs.com, which he started after working at IBM for five years. Before IBM, he worked in a number of hi-tech start-ups, leading new product development.

Mohammed has a master's of business administration from the University of California, Berkeley, and a master's of computer applications from RCC, Gujarat University, India.

It will not take even more time to obtain this Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller It will not take more cash to publish this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller Nowadays, individuals have been so smart to utilize the modern technology. Why don't you utilize your device or other tool to conserve this downloaded soft documents book Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller In this manner will let you to constantly be gone along with by this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller Obviously, it will certainly be the best buddy if you read this publication Big Data Analytics With Spark: A Practitioner's Guide To Using Spark For Large Scale Data Analysis By Mohammed Guller until completed.