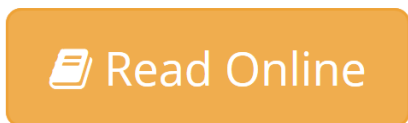


# Spark for Data Science

By Bikramaditya Singhal, Srinivas Duvvuri



**Spark for Data Science** By Bikramaditya Singhal, Srinivas Duvvuri

## Key Features

- Perform data analysis and build predictive models on huge datasets that leverage Apache Spark
- Learn to integrate data science algorithms and techniques with the fast and scalable computing features of Spark to address big data challenges
- Work through practical examples on real-world problems with sample code snippets

## Book Description

This is the era of Big Data and Internet of Things! Big Data implies big innovation and enables a competitive advantage for businesses. Apache Spark was designed to perform Big Data analytics at scale, and so Spark is equipped with the necessary algorithms and supports multiple programming languages.

Whether you are a technologist, a data scientist, or a beginner to Big Data analytics, this book will provide you with all the skills necessary to perform statistical data analysis, data visualization, predictive modeling, and build scalable data products or solutions using Python, Scala, and R.

With ample case studies and real-world examples, Spark for Data Science will help you ensure the successful execution of your data science projects.

## What you will learn

- Consolidate, clean, and transform your data acquired from various data sources
- Perform statistical analysis of data to find hidden insights
- Explore graphical techniques to see what your data looks like
- Use machine learning techniques to build predictive models
- Build scalable data products and solutions
- Start programming using the RADD API
- Become an expert by improving your data analytical skills

## About the Author

**Bikramaditya Singhal** works as a Senior Data Science Analyst with Broadridge Financial Solutions (India) Pvt. Ltd. He has over 6 years of experience in

statistical analysis, machine learning, and also in developing, designing, and architecting data-driven solutions.

His passion for technology and applied mathematics propelled him to pursue a career in data science. He is a strong believer in continuous innovation. He worked with Microsoft India and cofounded a company that provides data-driven insights to clients globally.

He has been a speaker at various conferences and meetups on data science, machine learning, and Apache Spark. His current skillset includes statistical data analysis, machine learning, R, Python, Scala, and ETL tools. With a unique blend of science as well as the technology aspect of Big Data, he has been instrumental in providing solutions to Big Data analytics problems.

**Srinivas Duvvuri** is currently heading the Fixed Income Suite of products at Broadridge India, and is also a principal member of the Broadridge Technology Council. In addition, he is involved in setting up the Big Data COE at Broadridge. He has over 22 years of experience in software product development and engineering complex, high-performance, scalable, multi-platform software solutions based on cutting edge technologies.

His experience predominantly spans product development in multiple domains including financial services, infrastructure management, OLAP, telecom billing, and customer care. Prior to Broadridge, he held leadership positions at a start-up and at leading IT majors such as CA, Hyperion (Oracle), and Globalstar, and also has a patent in Relational OLAP. Srinivas has a B.Tech in Aeronautics Engineering and an M.Tech in Computer Science, from IIT, Madras.

 [Download Spark for Data Science ...pdf](#)

 [Read Online Spark for Data Science ...pdf](#)

# Spark for Data Science

*By Bikramaditya Singhal, Srinivas Duvvuri*

**Spark for Data Science** By Bikramaditya Singhal, Srinivas Duvvuri

## Key Features

- Perform data analysis and build predictive models on huge datasets that leverage Apache Spark
- Learn to integrate data science algorithms and techniques with the fast and scalable computing features of Spark to address big data challenges
- Work through practical examples on real-world problems with sample code snippets

## Book Description

This is the era of Big Data and Internet of Things! Big Data implies big innovation and enables a competitive advantage for businesses. Apache Spark was designed to perform Big Data analytics at scale, and so Spark is equipped with the necessary algorithms and supports multiple programming languages.

Whether you are a technologist, a data scientist, or a beginner to Big Data analytics, this book will provide you with all the skills necessary to perform statistical data analysis, data visualization, predictive modeling, and build scalable data products or solutions using Python, Scala, and R.

With ample case studies and real-world examples, Spark for Data Science will help you ensure the successful execution of your data science projects.

## What you will learn

- Consolidate, clean, and transform your data acquired from various data sources
- Perform statistical analysis of data to find hidden insights
- Explore graphical techniques to see what your data looks like
- Use machine learning techniques to build predictive models
- Build scalable data products and solutions
- Start programming using the RADD API
- Become an expert by improving your data analytical skills

## About the Author

**Bikramaditya Singhal** works as a Senior Data Science Analyst with Broadridge Financial Solutions (India) Pvt. Ltd. He has over 6 years of experience in statistical analysis, machine learning, and also in developing, designing, and architecting data-driven solutions.

His passion for technology and applied mathematics propelled him to pursue a career in data science. He is a strong believer in continuous innovation. He worked with Microsoft India and cofounded a company that provides data-driven insights to clients globally.

He has been a speaker at various conferences and meetups on data science, machine learning, and Apache Spark. His current skillset includes statistical data analysis, machine learning, R, Python, Scala, and ETL tools. With a unique blend of science as well as the technology aspect of Big Data, he has been instrumental

in providing solutions to Big Data analytics problems.

**Srinivas Duvvuri** is currently heading the Fixed Income Suite of products at Broadridge India, and is also a principal member of the Broadridge Technology Council. In addition, he is involved in setting up the Big Data COE at Broadridge. He has over 22 years of experience in software product development and engineering complex, high-performance, scalable, multi-platform software solutions based on cutting edge technologies.

His experience predominantly spans product development in multiple domains including financial services, infrastructure management, OLAP, telecom billing, and customer care. Prior to Broadridge, he held leadership positions at a start-up and at leading IT majors such as CA, Hyperion (Oracle), and Globalstar, and also has a patent in Relational OLAP. Srinivas has a B.Tech in Aeronautics Engineering and an M.Tech in Computer Science, from IIT, Madras.

### Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri Bibliography

- Rank: #1563027 in eBooks
- Published on: 2016-09-30
- Released on: 2016-09-30
- Format: Kindle eBook

 [Download Spark for Data Science ...pdf](#)

 [Read Online Spark for Data Science ...pdf](#)

## Download and Read Free Online Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri

---

### Editorial Review

#### About the Author

Srinivas Duvvuri Srinivas Duvvuri is currently Senior Vice President Development, heading the development teams for Fixed Income Suite of products at Broadridge Financial Solutions (India) Pvt Ltd. In addition, he also leads the Big Data and Data Science COE and is the principal member of the Broadridge India Technology Council. He is self learnt Data Scientist. The Big Data /Data Science COE in the past 3 years, has successfully completed multiple POC's and some of the use cases are moving towards production deployment. He has over 25+ years of experience in software product development. His experience spans predominantly in product development in, multiple domains Financial Services, Infrastructure Management, OLAP, Telecom Billing and Customer Care, CAD/CAM. Prior to Broadridge, he s held leadership positions at a Startup and leading IT majors such as CA, Hyperion (Oracle), Globalstar. He has a patent in Relational OLAP. Srinivas loves to teach and mentor budding Engineers. He has established strong Academic connect and interacts with a host of educational institutions, He is an active speaker in various conferences, summits and meetups on topics such as Big data, Data Science Srinivas is a B.Tech in Aeronautical Engineering and M.Tech in Computer Science, from IIT, Madras.

Bikramaditya Singhal Bikramaditya Singhal is a data scientist with about 7 years of industry experience. He is an expert in statistical analysis, predictive analytics, machine learning, Bitcoin, Blockchain, and programming in C, R, and Python. He has extensive experience in building scalable data analytics solutions in many industry sectors. He also has an active interest on industrial IoT, machine to machine communication, decentralized computation through Blockchain and Artificial Intelligence. Bikram currently leads the data science team of Digital Enterprise Solutions' group at Tech Mahindra Ltd. He also worked in companies such as Microsoft India, Broadridge, Chelsio Communications and also cofounded a company named Mund Consulting' which focused on Big Data analytics. Bikram is an active speaker in various conferences, summits and meetups on topics such as big data, data science, IIoT and Blockchain.

### Users Review

#### From reader reviews:

##### Alyssa Cox:

A lot of people always spent their very own free time to vacation as well as go to the outside with them friends and family or their friend. Did you know? Many a lot of people spent these people free time just watching TV, or maybe playing video games all day long. In order to try to find a new activity that's look different you can read a new book. It is really fun for yourself. If you enjoy the book you read you can spent the entire day to reading a book. The book Spark for Data Science it is very good to read. There are a lot of people who recommended this book. These folks were enjoying reading this book. When you did not have enough space bringing this book you can buy the actual e-book. You can m0ore quickly to read this book from the smart phone. The price is not to cover but this book features high quality.

##### Todd McCrea:

Do you have something that you like such as book? The e-book lovers usually prefer to opt for book like comic, quick story and the biggest the first is novel. Now, why not attempting Spark for Data Science that

give your fun preference will be satisfied by reading this book. Reading behavior all over the world can be said as the opportunity for people to know world considerably better than how they react toward the world. It can't be stated constantly that reading habit only for the geeky person but for all of you who wants to end up being success person. So , for all of you who want to start reading through as your good habit, you may pick Spark for Data Science become your own starter.

**Glenna Monaghan:**

Your reading 6th sense will not betray you actually, why because this Spark for Data Science guide written by well-known writer who knows well how to make book which can be understand by anyone who have read the book. Written inside good manner for you, dripping every ideas and composing skill only for eliminate your hunger then you still question Spark for Data Science as good book not simply by the cover but also with the content. This is one reserve that can break don't evaluate book by its deal with, so do you still needing a different sixth sense to pick this kind of!? Oh come on your reading through sixth sense already alerted you so why you have to listening to another sixth sense.

**Nicholas Buchanan:**

This Spark for Data Science is new way for you who has interest to look for some information given it relief your hunger of information. Getting deeper you onto it getting knowledge more you know or you who still having bit of digest in reading this Spark for Data Science can be the light food for you personally because the information inside this particular book is easy to get by anyone. These books create itself in the form that is reachable by anyone, sure I mean in the e-book type. People who think that in reserve form make them feel drowsy even dizzy this guide is the answer. So there is no in reading a guide especially this one. You can find what you are looking for. It should be here for a person. So , don't miss that! Just read this e-book variety for your better life in addition to knowledge.

**Download and Read Online Spark for Data Science By  
Bikramaditya Singhal, Srinivas Duvvuri #42PSHT6FUIN**

## **Read Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri for online ebook**

Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri 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 for Data Science By Bikramaditya Singhal, Srinivas Duvvuri books to read online.

### **Online Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri ebook PDF download**

**Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri Doc**

**Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri Mobipocket**

**Spark for Data Science By Bikramaditya Singhal, Srinivas Duvvuri EPub**