



# 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:

##### Marcus Galvan:

As people who live in the particular modest era should be upgrade about what going on or information even knowledge to make these individuals keep up with the era that is certainly always change and progress. Some of you maybe may update themselves by reading books. It is a good choice for you but the problems coming to you is you don't know which one you should start with. This Spark for Data Science is our recommendation so you keep up with the world. Why, because this book serves what you want and need in this era.

##### Helen Leduc:

This Spark for Data Science are reliable for you who want to be a successful person, why. The explanation of this Spark for Data Science can be one of several great books you must have will be giving you more than just simple studying food but feed a person with information that probably will shock your preceding knowledge. This book is handy, you can bring it just about everywhere and whenever your conditions both in

e-book and printed versions. Beside that this Spark for Data Science forcing you to have an enormous of experience such as rich vocabulary, giving you demo of critical thinking that we all know it useful in your day task. So , let's have it and luxuriate in reading.

### **Alma Driver:**

As we know that book is essential thing to add our information for everything. By a guide we can know everything we want. A book is a pair of written, printed, illustrated as well as blank sheet. Every year seemed to be exactly added. This guide Spark for Data Science was filled regarding science. Spend your free time to add your knowledge about your scientific research competence. Some people has various feel when they reading any book. If you know how big benefit of a book, you can experience enjoy to read a e-book. In the modern era like today, many ways to get book you wanted.

### **Cody Chenault:**

Some individuals said that they feel bored stiff when they reading a book. They are directly felt the idea when they get a half elements of the book. You can choose typically the book Spark for Data Science to make your reading is interesting. Your own skill of reading ability is developing when you including reading. Try to choose simple book to make you enjoy to study it and mingle the impression about book and examining especially. It is to be first opinion for you to like to wide open a book and examine it. Beside that the reserve Spark for Data Science can to be your friend when you're feel alone and confuse with what must you're doing of their time.

**Download and Read Online Spark for Data Science By  
Bikramaditya Singhal, Srinivas Duvvuri #LHS7AJ9NP14**

## **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**