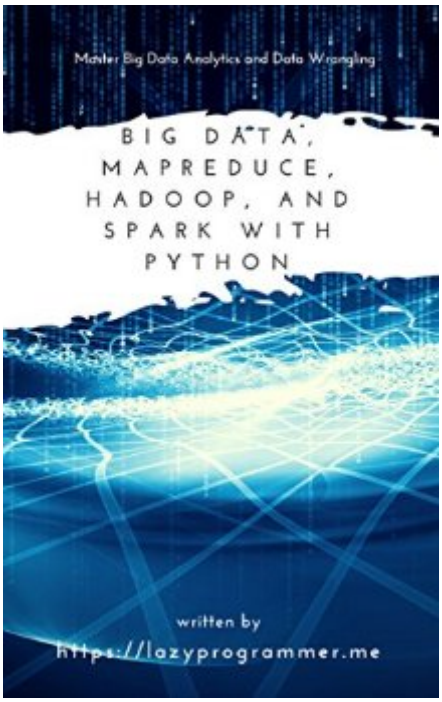
Big Data, MapReduce, Hadoop, And Spark With Python: Master Big Data Analytics And Data Wrangling With MapReduce Fundamentals Using Hadoop, Spark, And Python



Synopsis

Whatâ ™s the big deal with big data?It was recently reported in the Wall Street Journal that the government is collecting so much data on its citizens that they canâ ™t even use it effectively.A few â œunicornsâ • have popped up in the past decade or so, promising to help solve the big data problems that billion dollar corporations and the people running your country canâ ™t.It goes without saying that programming with frameworks that can do big data processing is a highly-coveted skill.Machine learning and artificial intelligence algorithms, which have garnered increased attention (and fear-mongering) in recent years, mainly due to the rise of deep learning, are completely dependent on data to learn.The more data the algorithm learns from, the smarter it can become. The problem is, the amount of data we collect has outpaced gains in CPU performance. Therefore, scalable methods for processing data are needed.In the early 2000s, Google invented MapReduce, a framework to systematically and methodically process big data in a scalable way by distributing the work across multiple machines.Later, the technology was adopted into an open-source framework called Hadoop, and then Spark emerged as a new big data framework which addressed some problems with MapReduce.In this book we will cover all 3 - the fundamental MapReduce paradigm, how to program with Hadoop, and how to program with Spark.Advance your CareerIf Spark is a better version of MapReduce, why are we even talking about it?Good question!Corporations, being slow-moving entities, are often still using Hadoop due to historical reasons. Just search for â œbig dataâ • and â œHadoopâ • on LinkedIn and you will see that there are a large number of high-salary openings for developers who know how to use Hadoop.In addition to giving you deeper insight into how big data processing works, learning about the fundamentals of MapReduce and Hadoop first will help you really appreciate how much easier Spark is to work with.Any startup or technical engineering team will appreciate a solid background with all of these technologies. Many will require you to know all of them, so that you can help maintain and patch their existing systems, and build newer and more efficient systems that improve the performance and robustness of the old systems.Amazingly, all the technologies we discuss in this book can be downloaded and installed for FREE. That means all you need to invest after purchasing this book is your effort and your time. The only prerequisites are that you are comfortable with Python coding and the command line shell. For the machine learning chapter youâ ™ll want to be familiar with using machine learning libraries.BONUS: At the end of this book, Iâ ™ll show you a super simple way to train a deep neural network on Spark with the classic MNIST dataset. It will demonstrate how easy it is to apply deep learning to big data.

Book Information

File Size: 276 KB

Print Length: 59 pages

Simultaneous Device Usage: Unlimited

Publication Date: August 15, 2016

Sold by:Â Digital Services LLC

Language: English

ASIN: B01KH9YWSY

Text-to-Speech: Enabled

X-Ray: Not Enabled

Word Wise: Not Enabled

Lending: Not Enabled

Enhanced Typesetting: Enabled

Best Sellers Rank: #225,686 Paid in Kindle Store (See Top 100 Paid in Kindle Store) #179 inÂ Kindle Store > Kindle Short Reads > 90 minutes (44-64 pages) > Computers & Technology #179 inÂ Books > Computers & Technology > Databases & Big Data > Data Processing #3022 inÂ Kindle Store > Kindle eBooks > Computers & Technology

Big Data, MapReduce, Hadoop, and Spark with Python: Master Big Data Analytics and Data Wrangling with MapReduce Fundamentals using Hadoop, Spark, and Python Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics Series) Python: PYTHON CRASH COURSE - Beginner's Course To Learn The Basics Of Python Programming In 24 Hours!: (Python, Python Programming, Python for Dummies, Python for Beginners, python crash course) Python: Learn Python In A DAY! - The Ultimate Crash Course to Learning the Basics of Python In No Time (Python, Python Course, Python Development, Python Books, Python for Beginners) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics Series) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) PYTHON: Python in 8 Hours, For Beginners, Learn Python Fast! A Smart Way to Learn Python, Plain & Simple, Learn

Python Programming Language in Easy Steps, A Beginner's Guide, Start Coding Today! Python: Learn Web Scraping with Python In A DAY! - The Ultimate Crash Course to Learning the Basics of Web Scraping with Python In No Time (Web Scraping ... Python Books, Python for Beginners) Python: Learn Python FAST - The Ultimate Crash Course to Learning the Basics of the Python Programming Language In No Time (Python, Python Programming, ... (Learn Coding Fast with Hands-On Project 7) Python Data Analytics: Data Analysis and Science using pandas, matplotlib and the Python Programming Language Learning Spark: Analytics With Spark Framework Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) Programming #45: Python Programming Professional Made Easy & Android Programming In a Day! (Python Programming, Python Language, Python for beginners, ... Programming Languages, Android Programming) Big Data Driven Supply Chain Management: A Framework for Implementing Analytics and Turning Information Into Intelligence (FT Press Analytics) Measuring the Digital World: Using Digital Analytics to Drive Better Digital Experiences (FT Press Analytics) Introducing Data Science: Big Data, Machine Learning, and more, using Python tools