Access Free Parallel Computing For Data Science With Examples In R C

Yeah, reviewing a books **Parallel Computing For Data Science With Examples In R C** could mount up your near associates listings. This is just one of the solutions for you to be successful. As understood, capability does not recommend that you have astounding points.

Comprehending as well as pact even more than new will give each success. next-door to, the declaration as without difficulty as sharpness of this Parallel Computing For Data Science With Examples In R C can be taken as with ease as picked to act.

KEY=COMPUTING - DONNA CODY

Parallel Computing for Data Science With Examples in R, C++ and CUDA CRC Press Parallel Computing for Data Science: With Examples in R, C++ and CUDA is one of the first parallel computing books to concentrate exclusively on parallel data structures, algorithms, software tools, and applications in data science. It includes examples not only from the classic "n observations, p variables" matrix format but also from time series, Parallel Computing for Data Science With Examples in R, C++ and Cuda CreateSpace Thought-provoking and accessible in approach, this updated and expanded second edition of the Parallel Computing for Data Science: With Examples in R, C++ and CUDA provides a user-friendly introduction to the subject, Taking a clear structural framework, it guides the reader through the subject's core elements. A flowing writing style combines with the use of illustrations and diagrams throughout the text to ensure the reader understands even the most complex of concepts. This succinct and enlightening overview is a required reading for advanced graduate-level students. We hope you find this book useful in shaping your future career. Feel free to send us your enquiries related to our publications to info@risepress.pw Rise Press Python Data Analysis Perform data collection, data processing, wrangling, visualization, and model building using Python Packt Publishing Ltd Understand data analysis pipelines using machine learning algorithms and techniques with this practical guide Key Features Prepare and clean your data to use it for exploratory analysis, data manipulation, and data wrangling Discover supervised, unsupervised, probabilistic, and Bayesian machine learning methods Get to grips with graph processing and sentiment analysis Book Description Data analysis enables you to generate value from small and big data by discovering new patterns and trends, and Python is one of the most popular tools for analyzing a wide variety of data. With this book, you'll get up and running using Python for data analysis by exploring the different phases and methodologies used in data analysis and learning how to use modern

libraries from the Python ecosystem to create efficient data pipelines. Starting with the essential statistical and data analysis fundamentals using Python, you'll perform complex data analysis and modeling, data manipulation, data cleaning, and data visualization using easy-to-follow examples. You'll then understand how to conduct time series analysis and signal processing using ARMA models. As you advance, you'll get to grips with smart processing and data analytics using machine learning algorithms such as regression, classification, Principal Component Analysis (PCA), and clustering. In the concluding chapters, you'll work on real-world examples to analyze textual and image data using natural language processing (NLP) and image analytics techniques, respectively. Finally, the book will demonstrate parallel computing using Dask. By the end of this data analysis book, you'll be equipped with the skills you need to prepare data for analysis and create meaningful data visualizations for forecasting values from data. What you will learn Explore data science and its various process models Perform data manipulation using NumPy and pandas for aggregating, cleaning, and handling missing values Create interactive visualizations using Matplotlib, Seaborn, and Bokeh Retrieve, process, and store data in a wide range of formats Understand data preprocessing and feature engineering using pandas and scikit-learn Perform time series analysis and signal processing using sunspot cycle data Analyze textual data and image data to perform advanced analysis Get up to speed with parallel computing using Dask Who this book is for This book is for data analysts, business analysts, statisticians, and data scientists looking to learn how to use Python for data analysis. Students and academic faculties will also find this book useful for learning and teaching Python data analysis using a hands-on approach. A basic understanding of math and working knowledge of the Python programming language will help you get started with this book. Data Science with Python and Dask Simon and Schuster Summary Dask is a native parallel analytics tool designed to integrate seamlessly with the libraries you're already using, including Pandas, NumPy, and Scikit-Learn. With Dask you can crunch and work with huge datasets, using the tools you already have. And Data Science with Python and Dask is your guide to using Dask for your data projects without changing the way you work! Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. You'll find registration instructions inside the print book. About the Technology An efficient data pipeline means everything for the success of a data science project. Dask is a flexible library for parallel computing in Python that makes it easy to build intuitive workflows for ingesting and analyzing large, distributed datasets. Dask provides dynamic task scheduling and parallel collections that extend the functionality of NumPy, Pandas, and Scikit-learn, enabling users to scale their code from a single laptop to a cluster of hundreds of machines with ease. About the Book Data Science with Python and Dask teaches you to build scalable projects that can handle massive datasets. After meeting the Dask framework, you'll analyze

data in the NYC Parking Ticket database and use DataFrames to streamline your process. Then, you'll create machine learning models using Dask-ML, build interactive visualizations, and build clusters using AWS and Docker. What's inside Working with large, structured and unstructured datasets Visualization with Seaborn and Datashader Implementing your own algorithms Building distributed apps with Dask Distributed Packaging and deploying Dask apps About the Reader For data scientists and developers with experience using Python and the PyData stack. About the Author Jesse Daniel is an experienced Python developer. He taught Python for Data Science at the University of Denver and leads a team of data scientists at a Denver-based media technology company. Table of Contents PART 1 - The Building Blocks of scalable computing Why scalable computing matters Introducing Dask PART 2 - Working with Structured Data using Dask DataFrames Introducing Dask DataFrames Loading data into DataFrames Cleaning and transforming DataFrames Summarizing and analyzing DataFrames Visualizing DataFrames with Seaborn Visualizing location data with Datashader PART 3 - Extending and deploying Dask Working with Bags and Arrays Machine learning with Dask-ML Scaling and deploying Dask Mastering Parallel Programming with R Packt Publishing Ltd Master the robust features of R parallel programming to accelerate your data science computations About This Book Create R programs that exploit the computational capability of your cloud platforms and computers to the fullest Become an expert in writing the most efficient and highest performance parallel algorithms in R Get to grips with the concept of parallelism to accelerate your existing R programs Who This Book Is For This book is for R programmers who want to step beyond its inherent single-threaded and restricted memory limitations and learn how to implement highly accelerated and scalable algorithms that are a necessity for the performant processing of Big Data. No previous knowledge of parallelism is required. This book also provides for the more advanced technical programmer seeking to go beyond high level parallel frameworks. What You Will Learn Create and structure efficient load-balanced parallel computation in R, using R's built-in parallel package Deploy and utilize cloud-based parallel infrastructure from R, including launching a distributed computation on Hadoop running on Amazon Web Services (AWS) Get accustomed to parallel efficiency, and apply simple techniques to benchmark, measure speed and target improvement in your own code Develop complex parallel processing algorithms with the standard Message Passing Interface (MPI) using RMPI, pbdMPI, and SPRINT packages Build and extend a parallel R package (SPRINT) with your own MPI-based routines Implement accelerated numerical functions in R utilizing the vector processing capability of your Graphics Processing Unit (GPU) with OpenCL Understand parallel programming pitfalls, such as deadlock and numerical instability, and the approaches to handle and avoid them Build a task farm master-worker, spatial grid, and hybrid parallel R programs In Detail R is one of the most popular programming languages used in data

science. Applying R to big data and complex analytic tasks requires the harnessing of scalable compute resources. Mastering Parallel Programming with R presents a comprehensive and practical treatise on how to build highly scalable and efficient algorithms in R. It will teach you a variety of parallelization techniques, from simple use of R's built-in parallel package versions of lapply(), to high-level AWS cloud-based Hadoop and Apache Spark frameworks. It will also teach you low level scalable parallel programming using RMPI and pbdMPI for message passing, applicable to clusters and supercomputers, and how to exploit thousand-fold simple processor GPUs through ROpenCL. By the end of the book, you will understand the factors that influence parallel efficiency, including assessing code performance and implementing load balancing; pitfalls to avoid, including deadlock and numerical instability issues; how to structure your code and data for the most appropriate type of parallelism for your problem domain; and how to extract the maximum performance from your R code running on a variety of computer systems. Style and approach This book leads you chapter by chapter from the easy to more complex forms of parallelism. The author's insights are presented through clear practical examples applied to a range of different problems, with comprehensive reference information for each of the R packages employed. The book can be read from start to finish, or by dipping in chapter by chapter, as each chapter describes a specific parallel approach and technology, so can be read as a standalone. Data Science for Transport A Self-Study Guide with Computer Exercises Springer The quantity, diversity and availability of transport data is increasing rapidly, requiring new skills in the management and interrogation of data and databases. Recent years have seen a new wave of 'big data', 'Data Science', and 'smart cities' changing the world, with the Harvard Business Review describing Data Science as the "sexiest job of the 21st century". Transportation professionals and researchers need to be able to use data and databases in order to establish quantitative, empirical facts, and to validate and challenge their mathematical models, whose axioms have traditionally often been assumed rather than rigorously tested against data. This book takes a highly practical approach to learning about Data Science tools and their application to investigating transport issues. The focus is principally on practical, professional work with real data and tools, including business and ethical issues. "Transport modeling practice was developed in a data poor world, and many of our current techniques and skills are building on that sparsity. In a new data rich world, the required tools are different and the ethical questions around data and privacy are definitely different. I am not sure whether current professionals have these skills; and I am certainly not convinced that our current transport modeling tools will survive in a data rich environment. This is an exciting time to be a data scientist in the transport field. We are trying to get to grips with the opportunities that big data sources offer; but at the same time such data skills need to be fused with an understanding of transport, and of transport modeling. Those with

these combined skills can be instrumental at providing better, faster, cheaper data for transport decision- making; and ultimately contribute to innovative, efficient, data driven modeling techniques of the future. It is not surprising that this course, this book, has been authored by the Institute for Transport Studies. To do this well, you need a blend of academic rigor and practical pragmatism. There are few educational or research establishments better equipped to do that than ITS Leeds". - Tom van Vuren, Divisional Director, Mott MacDonald "WSP is proud to be a thought leader in the world of transport modelling, planning and economics, and has a wide range of opportunities for people with skills in these areas. The evidence base and forecasts we deliver to effectively implement strategies and schemes are ever more data and technology focused a trend we have helped shape since the 1970's, but with particular disruption and opportunity in recent years. As a result of these trends, and to suitably skill the next generation of transport modellers, we asked the world-leading Institute for Transport Studies, to boost skills in these areas, and they have responded with a new MSc programme which you too can now study via this book." - Leighton Cardwell, Technical Director, WSP. "From processing and analysing large datasets, to automation of modelling tasks sometimes requiring different software packages to "talk" to each other, to data visualization, SYSTRA employs a range of techniques and tools to provide our clients with deeper insights and effective solutions. This book does an excellent job in giving you the skills to manage, interrogate and analyse databases, and develop powerful presentations. Another important publication from ITS Leeds." - Fitsum Teklu, Associate Director (Modelling & Appraisal) SYSTRA Ltd "Urban planning has relied for decades on statistical and computational practices that have little to do with mainstream data science. Information is still often used as evidence on the impact of new infrastructure even when it hardly contains any valid evidence. This book is an extremely welcome effort to provide young professionals with the skills needed to analyse how cities and transport networks actually work. The book is also highly relevant to anyone who will later want to build digital solutions to optimise urban travel based on emerging data sources". - Yaron Hollander, author of "Transport Modelling for a Complete Beginner" Big Data Analysis with Python Combine Spark and Python to unlock the powers of parallel computing and machine learning Packt Publishing Ltd Get to grips with processing large volumes of data and presenting it as engaging, interactive insights using Spark and Python. Key FeaturesGet a hands-on, fast-paced introduction to the Python data science stackExplore ways to create useful metrics and statistics from large datasetsCreate detailed analysis reports with real-world dataBook Description Processing big data in real time is challenging due to scalability, information inconsistency, and fault tolerance. Big Data Analysis with Python teaches you how to use tools that can control this data avalanche for you. With this book, you'll learn practical techniques to aggregate data into useful dimensions for posterior analysis, extract

statistical measurements, and transform datasets into features for other systems. The book begins with an introduction to data manipulation in Python using pandas. You'll then get familiar with statistical analysis and plotting techniques. With multiple hands-on activities in store, you'll be able to analyze data that is distributed on several computers by using Dask. As you progress, you'll study how to aggregate data for plots when the entire data cannot be accommodated in memory. You'll also explore Hadoop (HDFS and YARN), which will help you tackle larger datasets. The book also covers Spark and explains how it interacts with other tools. By the end of this book, you'll be able to bootstrap your own Python environment, process large files, and manipulate data to generate statistics, metrics, and graphs. What you will learnUse Python to read and transform data into different formatsGenerate basic statistics and metrics using data on diskWork with computing tasks distributed over a clusterConvert data from various sources into storage or querying formatsPrepare data for statistical analysis, visualization, and machine learningPresent data in the form of effective visualsWho this book is for Big Data Analysis with Python is designed for Python developers, data analysts, and data scientists who want to get hands-on with methods to control data and transform it into impactful insights. Basic knowledge of statistical measurements and relational databases will help you to understand various concepts explained in this book. Introduction to HPC with MPI for Data Science Springer This gentle introduction to High Performance Computing (HPC) for Data Science using the Message Passing Interface (MPI) standard has been designed as a first course for undergraduates on parallel programming on distributed memory models, and requires only basic programming notions. Divided into two parts the first part covers high performance computing using C++ with the Message Passing Interface (MPI) standard followed by a second part providing high-performance data analytics on computer clusters. In the first part, the fundamental notions of blocking versus non-blocking point-to-point communications, global communications (like broadcast or scatter) and collaborative computations (reduce), with Amdalh and Gustafson speed-up laws are described before addressing parallel sorting and parallel linear algebra on computer clusters. The common ring, torus and hypercube topologies of clusters are then explained and global communication procedures on these topologies are studied. This first part closes with the MapReduce (MR) model of computation well-suited to processing big data using the MPI framework. In the second part, the book focuses on high-performance data analytics. Flat and hierarchical clustering algorithms are introduced for data exploration along with how to program these algorithms on computer clusters, followed by machine learning classification, and an introduction to graph analytics. This part closes with a concise introduction to data core-sets that let big data problems be amenable to tiny data problems. Exercises are included at the end of each chapter in order for students to practice the concepts learned, and a final

section contains an overall exam which allows them to evaluate how well they have assimilated the material covered in the book. Scientific Parallel Computing Princeton University Press What does Google's management of billions of Web pages have in common with analysis of a genome with billions of nucleotides? Both apply methods that coordinate many processors to accomplish a single task. From mining genomes to the World Wide Web, from modeling financial markets to global weather patterns, parallel computing enables computations that would otherwise be impractical if not impossible with sequential approaches alone. Its fundamental role as an enabler of simulations and data analysis continues an advance in a wide range of application areas. Scientific Parallel Computing is the first textbook to integrate all the fundamentals of parallel computing in a single volume while also providing a basis for a deeper understanding of the subject. Designed for graduate and advanced undergraduate courses in the sciences and in engineering, computer science, and mathematics, it focuses on the three key areas of algorithms, architecture, languages, and their crucial synthesis in performance. The book's computational examples, whose math prerequisites are not beyond the level of advanced calculus, derive from a breadth of topics in scientific and engineering simulation and data analysis. The programming exercises presented early in the book are designed to bring students up to speed quickly, while the book later develops projects challenging enough to guide students toward research questions in the field. The new paradigm of cluster computing is fully addressed. A supporting web site provides access to all the codes and software mentioned in the book, and offers topical information on popular parallel computing systems. Integrates all the fundamentals of parallel computing essential for today's high-performance requirements Ideal for graduate and advanced undergraduate students in the sciences and in engineering, computer science, and mathematics Extensive programming and theoretical exercises enable students to write parallel codes quickly More challenging projects later in the book introduce research questions New paradigm of cluster computing fully addressed Supporting web site provides access to all the codes and software mentioned in the book Data Science for Public Policy Springer Nature This textbook presents the essential tools and core concepts of data science to public officials, policy analysts, and economists among others in order to further their application in the public sector. An expansion of the quantitative economics frameworks presented in policy and business schools, this book emphasizes the process of asking relevant questions to inform public policy. Its techniques and approaches emphasize data-driven practices, beginning with the basic programming paradigms that occupy the majority of an analysts time and advancing to the practical applications of statistical learning and machine learning. The text considers two divergent, competing perspectives to support its applications, incorporating techniques from both causal inference and prediction. Additionally, the book includes open-sourced data as well as

live code, written in R and presented in notebook form, which readers can use and modify to practice working with data. R Programming for Data Science Lulu.com Data science has taken the world by storm. Every field of study and area of business has been affected as people increasingly realize the value of the incredible quantities of data being generated. But to extract value from those data, one needs to be tra Nature Inspired Computing for Data Science Springer Nature This book discusses the current research and concepts in data science and how these can be addressed using different nature-inspired optimization techniques. Focusing on various data science problems, including classification, clustering, forecasting, and deep learning, it explores how researchers are using nature-inspired optimization techniques to find solutions to these problems in domains such as disease analysis and health care, object recognition, vehicular ad-hoc networking, high-dimensional data analysis, gene expression analysis, microgrids, and deep learning. As such it provides insights and inspiration for researchers to wanting to employ nature-inspired optimization techniques in their own endeavors. Applied Data Science Using PySpark Learn the End-to-End Predictive Model-Building Cycle Apress Discover the capabilities of PySpark and its application in the realm of data science. This comprehensive guide with hand-picked examples of daily use cases will walk you through the end-to-end predictive model-building cycle with the latest techniques and tricks of the trade. Applied Data Science Using PySpark is divided unto six sections which walk you through the book. In section 1, you start with the basics of PySpark focusing on data manipulation. We make you comfortable with the language and then build upon it to introduce you to the mathematical functions available off the shelf. In section 2, you will dive into the art of variable selection where we demonstrate various selection techniques available in PySpark. In section 3, we take you on a journey through machine learning algorithms, implementations, and fine-tuning techniques. We will also talk about different validation metrics and how to use them for picking the best models. Sections 4 and 5 go through machine learning pipelines and various methods available to operationalize the model and serve it through Docker/an API. In the final section, you will cover reusable objects for easy experimentation and learn some tricks that can help you optimize your programs and machine learning pipelines. By the end of this book, you will have seen the flexibility and advantages of PySpark in data science applications. This book is recommended to those who want to unleash the power of parallel computing by simultaneously working with big datasets. What You Will Learn Build an end-to-end predictive model Implement multiple variable selection techniques Operationalize models Master multiple algorithms and implementations Who This Book is For Data scientists and machine learning and deep learning engineers who want to learn and use PySpark for real-time analysis of streaming data. Introduction to Data Science A Python Approach to Concepts, Techniques and Applications Springer This accessible and classroom-tested

textbook/reference presents an introduction to the fundamentals of the emerging and interdisciplinary field of data science. The coverage spans key concepts adopted from statistics and machine learning, useful techniques for graph analysis and parallel programming, and the practical application of data science for such tasks as building recommender systems or performing sentiment analysis. Topics and features: provides numerous practical case studies using real-world data throughout the book; supports understanding through hands-on experience of solving data science problems using Python; describes techniques and tools for statistical analysis, machine learning, graph analysis, and parallel programming; reviews a range of applications of data science, including recommender systems and sentiment analysis of text data; provides supplementary code resources and data at an associated website. IPython Interactive Computing and Visualization Cookbook Over 100 hands-on recipes to sharpen your skills in high-performance numerical computing and data science in the Jupyter Notebook, 2nd Edition Packt Publishing Ltd Learn to use IPython and Jupyter Notebook for your data analysis and visualization work. Key Features Leverage the Jupyter Notebook for interactive data science and visualization Become an expert in high-performance computing and visualization for data analysis and scientific modeling A comprehensive coverage of scientific computing through many hands-on, example-driven recipes with detailed, step-by-step explanations Book Description Python is one of the leading open source platforms for data science and numerical computing. IPython and the associated Jupyter Notebook offer efficient interfaces to Python for data analysis and interactive visualization, and they constitute an ideal gateway to the platform. IPython Interactive Computing and Visualization Cookbook, Second Edition contains many ready-to-use, focused recipes for high-performance scientific computing and data analysis, from the latest IPython/Jupyter features to the most advanced tricks, to help you write better and faster code. You will apply these state-of-the-art methods to various real-world examples, illustrating topics in applied mathematics, scientific modeling, and machine learning. The first part of the book covers programming techniques: code quality and reproducibility, code optimization, high-performance computing through just-in-time compilation, parallel computing, and graphics card programming. The second part tackles data science, statistics, machine learning, signal and image processing, dynamical systems, and pure and applied mathematics. What you will learn Master all features of the Jupyter Notebook Code better: write high-quality, readable, and well-tested programs; profile and optimize your code; and conduct reproducible interactive computing experiments Visualize data and create interactive plots in the Jupyter Notebook Write blazingly fast Python programs with NumPy, ctypes, Numba, Cython, OpenMP, GPU programming (CUDA), parallel IPython, Dask, and more Analyze data with Bayesian or frequentist statistics (Pandas, PyMC, and R), and learn from actual data through machine

learning (scikit-learn) Gain valuable insights into signals, images, and sounds with SciPy, scikit-image, and OpenCV Simulate deterministic and stochastic dynamical systems in Python Familiarize yourself with math in Python using SymPy and Sage: algebra, analysis, logic, graphs, geometry, and probability theory Who this book is for This book is intended for anyone interested in numerical computing and data science: students, researchers, teachers, engineers, analysts, and hobbyists. A basic knowledge of Python/NumPy is recommended. Some skills in mathematics will help you understand the theory behind the computational methods. Encyclopedia of Parallel Computing Springer Science & Business Media Containing over 300 entries in an A-Z format, the Encyclopedia of Parallel Computing provides easy, intuitive access to relevant information for professionals and researchers seeking access to any aspect within the broad field of parallel computing. Topics for this comprehensive reference were selected, written, and peer-reviewed by an international pool of distinguished researchers in the field. The Encyclopedia is broad in scope, covering machine organization, programming languages, algorithms, and applications. Within each area, concepts, designs, and specific implementations are presented. The highly-structured essays in this work comprise synonyms, a definition and discussion of the topic, bibliographies, and links to related literature. Extensive cross-references to other entries within the Encyclopedia support efficient, user-friendly searchers for immediate access to useful information. Key concepts presented in the Encyclopedia of Parallel Computing include; laws and metrics; specific numerical and non-numerical algorithms; asynchronous algorithms; libraries of subroutines; benchmark suites; applications; sequential consistency and cache coherency; machine classes such as clusters, shared-memory multiprocessors, special-purpose machines and dataflow machines; specific machines such as Cray supercomputers, IBM's cell processor and Intel's multicore machines; race detection and auto parallelization; parallel programming languages, synchronization primitives, collective operations, message passing libraries, checkpointing, and operating systems. Topics covered: Speedup, Efficiency, Isoefficiency, Redundancy, Amdahls law, Computer Architecture Concepts, Parallel Machine Designs, Benmarks, Parallel Programming concepts & design, Algorithms, Parallel applications. This authoritative reference will be published in two formats: print and online. The online edition features hyperlinks to cross-references and to additional significant research. Related Subjects: supercomputing, high-performance computing, distributed computing Data Science and Big Data Computing Frameworks and Methodologies Springer This illuminating text/reference surveys the state of the art in data science, and provides practical guidance on big data analytics. Expert perspectives are provided by authoritative researchers and practitioners from around the world, discussing research developments and emerging trends, presenting case studies on helpful frameworks and innovative methodologies, and suggesting best practices

for efficient and effective data analytics. Features: reviews a framework for fast data applications, a technique for complex event processing, and agglomerative approaches for the partitioning of networks; introduces a unified approach to data modeling and management, and a distributed computing perspective on interfacing physical and cyber worlds; presents techniques for machine learning for big data, and identifying duplicate records in data repositories; examines enabling technologies and tools for data mining; proposes frameworks for data extraction, and adaptive decision making and social media analysis. Data Science 6th International Conference, ICDS 2019, Ningbo, China, May 15–20, 2019, Revised Selected Papers Springer Nature This book constitutes the refereed proceedings of the 6th International Conference on Data Science, ICDS 2019, held in Ningbo, China, during May 2019. The 64 revised full papers presented were carefully reviewed and selected from 210 submissions. The research papers cover the areas of Advancement of Data Science and Smart City Applications, Theory of Data Science, Data Science of People and Health, Web of Data, Data Science of Trust and Internet of Things. Parallel Computing is Everywhere IOS Press The most powerful computers work by harnessing the combined computational power of millions of processors, and exploiting the full potential of such large-scale systems is something which becomes more difficult with each succeeding generation of parallel computers. Alternative architectures and computer paradigms are increasingly being investigated in an attempt to address these difficulties. Added to this, the pervasive presence of heterogeneous and parallel devices in consumer products such as mobile phones, tablets, personal computers and servers also demands efficient programming environments and applications aimed at small-scale parallel systems as opposed to large-scale supercomputers. This book presents a selection of papers presented at the conference: Parallel Computing (ParCo2017), held in Bologna, Italy, on 12 to 15 September 2017. The conference included contributions about alternative approaches to achieving High Performance Computing (HPC) to potentially surpass exa- and zetascale performances, as well as papers on the application of quantum computers and FPGA processors. These developments are aimed at making available systems better capable of solving intensive computational scientific/engineering problems such as climate models, security applications and classic NP-problems, some of which cannot currently be managed by even the most powerful supercomputers available. New areas of application, such as robotics, AI and learning systems, data science, the Internet of Things (IoT), and in-car systems and autonomous vehicles were also covered. As always, ParCo2017 attracted a large number of notable contributions covering present and future developments in parallel computing, and the book will be of interest to all those working in the field. Advances in Artificial Intelligence, Computation, and Data Science For Medicine and Life Science Springer Nature Artificial intelligence (AI) has become pervasive in most areas of research and applications. While computation can significantly

reduce mental efforts for complex problem solving, effective computer algorithms allow continuous improvement of AI tools to handle complexity—in both time and memory requirements—for machine learning in large datasets. Meanwhile, data science is an evolving scientific discipline that strives to overcome the hindrance of traditional skills that are too limited to enable scientific discovery when leveraging research outcomes. Solutions to many problems in medicine and life science, which cannot be answered by these conventional approaches, are urgently needed for society. This edited book attempts to report recent advances in the complementary domains of AI, computation, and data science with applications in medicine and life science. The benefits to the reader are manifold as researchers from similar or different fields can be aware of advanced developments and novel applications that can be useful for either immediate implementations or future scientific pursuit. Features: Considers recent advances in AI, computation, and data science for solving complex problems in medicine, physiology, biology, chemistry, and biochemistry Provides recent developments in three evolving key areas and their complementary combinations: AI, computation, and data science Reports on applications in medicine and physiology, including cancer, neuroscience, and digital pathology Examines applications in life science, including systems biology, biochemistry, and even food technology This unique book, representing research from a team of international contributors, has not only real utility in academia for those in the medical and life sciences communities, but also a much wider readership from industry, science, and other areas of technology and education. Parallel R Data Analysis in the Distributed World "O'Reilly Media, Inc." It’s tough to argue with R as a high-quality, cross-platform, open source statistical software product—unless you’re in the business of crunching Big Data. This concise book introduces you to several strategies for using R to analyze large datasets, including three chapters on using R and Hadoop together. You’ll learn the basics of Snow, Multicore, Parallel, Segue, RHIPE, and Hadoop Streaming, including how to find them, how to use them, when they work well, and when they don’t. With these packages, you can overcome R’s single-threaded nature by spreading work across multiple CPUs, or offloading work to multiple machines to address R’s memory barrier. Snow: works well in a traditional cluster environment Multicore: popular for multiprocessor and multicore computers Parallel: part of the upcoming R 2.14.0 release R+Hadoop: provides low-level access to a popular form of cluster computing RHIPE: uses Hadoop’s power with R’s language and interactive shell Segue: lets you use Elastic MapReduce as a backend for lapply-style operations Data Analysis in the Cloud Models, Techniques and Applications Elsevier Data Analysis in the Cloud introduces and discusses models, methods, techniques, and systems to analyze the large number of digital data sources available on the Internet using the computing and storage facilities of the cloud. Coverage includes scalable data mining and knowledge discovery techniques together with cloud

computing concepts, models, and systems. Specific sections focus on map-reduce and NoSQL models. The book also includes techniques for conducting high-performance distributed analysis of large data on clouds. Finally, the book examines research trends such as Big Data pervasive computing, data-intensive exascale computing, and massive social network analysis. Introduces data analysis techniques and cloud computing concepts Describes cloud-based models and systems for Big Data analytics Provides examples of the state-of-the-art in cloud data analysis Explains how to develop large-scale data mining applications on clouds Outlines the main research trends in the area of scalable Big Data analysis Big Data and High Performance Computing IOS Press Big Data has been much in the news in recent years, and the advantages conferred by the collection and analysis of large datasets in fields such as marketing, medicine and finance have led to claims that almost any real world problem could be solved if sufficient data were available. This is of course a very simplistic view, and the usefulness of collecting, processing and storing large datasets must always be seen in terms of the communication, processing and storage capabilities of the computing platforms available. This book presents papers from the International Research Workshop, Advanced High Performance Computing Systems, held in Cetraro, Italy, in July 2014. The papers selected for publication here discuss fundamental aspects of the definition of Big Data, as well as considerations from practice where complex datasets are collected, processed and stored. The concepts, problems, methodologies and solutions presented are of much more general applicability than may be suggested by the particular application areas considered. As a result the book will be of interest to all those whose work involves the processing of very large data sets, exascale computing and the emerging fields of data science Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython "O'Reilly Media, Inc." Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples BIG DATA

ANALYTICS: CLUSTER ANALYSIS AND PATTERN RECOGNITION. EXAMPLES WITH MATLAB Lulu Press, Inc Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. MATLAB has the tool Neural Network Toolbox (Deep Learning Toolbox from version 18) that provides algorithms, functions, and apps to create, train, visualize, and simulate neural networks. You can perform classification, regression, clustering, dimensionality reduction, time-series forecasting, and dynamic system modeling and control.The toolbox includes convolutional neural network and autoencoder deep learning algorithms for image classification and feature learning tasks. To speed up training of large data sets, you can distribute computations and data across multicore processors, GPUs, and computer clusters using Big Data tools (Parallel Computing Toolbox). Unsupervised learning algorithms, including self-organizing maps and competitive layers-Apps for data-fitting, pattern recognition, and clustering-Preprocessing, postprocessing, and network visualization for improving training efficiency and assessing network performance. This book develops cluster analysis and pattern recognition Languages and Compilers for Parallel Computing 30th International Workshop, LCPC 2017, College Station, TX, USA, October 11–13, 2017, Revised Selected Papers Springer Nature This book constitutes the proceedings of the 30th International Workshop on Languages and Compilers for Parallel Computing, LCPC 2017, held in College Station, TX, USA, in October 2017. The 17 full papers presented together with abstracts of 5 keynote talks, 11 invited speakers and 4 poster papers in this volume were carefully reviewed and selected from 26 submissions. LCPC encourages submissions that go outside its original scope of scientific computing to diverse areas that are enable or enhanced by the power of parallel systems such as mobile computing, big data, relevant aspects of machine learning, data centers, cognitive computing, etc. LCPC strongly encourages personal interaction and technical discussions along the initial material. Applied Data Science Lessons Learned for the Data-Driven Business Springer This book has two main goals: to define data science through the work of data scientists and their results, namely data products, while simultaneously providing the reader with relevant lessons learned from applied data science projects at the intersection of academia and industry. As such, it is not a replacement for a classical textbook (i.e., it does not elaborate on fundamentals of methods and principles described elsewhere), but systematically highlights the connection between theory, on the one hand, and its application in specific use cases, on the other. With these goals in mind, the book is divided into three parts: Part I pays tribute to the interdisciplinary nature of data science and provides a common understanding of data science terminology for readers with different backgrounds. These six chapters are geared towards drawing a consistent picture of data science and were predominantly written by the editors themselves. Part II then broadens the spectrum by presenting views and insights from diverse authors – some from academia and some

from industry, ranging from financial to health and from manufacturing to e-commerce. Each of these chapters describes a fundamental principle, method or tool in data science by analyzing specific use cases and drawing concrete conclusions from them. The case studies presented, and the methods and tools applied, represent the nuts and bolts of data science. Finally, Part III was again written from the perspective of the editors and summarizes the lessons learned that have been distilled from the case studies in Part II. The section can be viewed as a meta-study on data science across a broad range of domains, viewpoints and fields. Moreover, it provides answers to the question of what the mission-critical factors for success in different data science undertakings are. The book targets professionals as well as students of data science: first, practicing data scientists in industry and academia who want to broaden their scope and expand their knowledge by drawing on the authors' combined experience. Second, decision makers in businesses who face the challenge of creating or implementing a data-driven strategy and who want to learn from success stories spanning a range of industries. Third, students of data science who want to understand both the theoretical and practical aspects of data science, vetted by real-world case studies at the intersection of academia and industry. Effective Data Science Infrastructure How to make data scientists productive Simon and Schuster Simplify data science infrastructure to give data scientists an efficient path from prototype to production. In Effective Data Science Infrastructure you will learn how to: Design data science infrastructure that boosts productivity Handle compute and orchestration in the cloud Deploy machine learning to production Monitor and manage performance and results Combine cloud-based tools into a cohesive data science environment Develop reproducible data science projects using Metaflow, Conda, and Docker Architect complex applications for multiple teams and large datasets Customize and grow data science infrastructure Effective Data Science Infrastructure: How to make data scientists more productive is a hands-on guide to assembling infrastructure for data science and machine learning applications. It reveals the processes used at Netflix and other data-driven companies to manage their cutting edge data infrastructure. In it, you'll master scalable techniques for data storage, computation, experiment tracking, and orchestration that are relevant to companies of all shapes and sizes. You'll learn how you can make data scientists more productive with your existing cloud infrastructure, a stack of open source software, and idiomatic Python. The author is donating proceeds from this book to charities that support women and underrepresented groups in data science. About the technology Growing data science projects from prototype to production requires reliable infrastructure. Using the powerful new techniques and tooling in this book, you can stand up an infrastructure stack that will scale with any organization, from startups to the largest enterprises. About the book Effective Data Science Infrastructure teaches you to build data pipelines and project workflows that will supercharge data scientists and

their projects. Based on state-of-the-art tools and concepts that power data operations of Netflix, this book introduces a customizable cloud-based approach to model development and MLOps that you can easily adapt to your company’s specific needs. As you roll out these practical processes, your teams will produce better and faster results when applying data science and machine learning to a wide array of business problems. What's inside Handle compute and orchestration in the cloud Combine cloud-based tools into a cohesive data science environment Develop reproducible data science projects using Metaflow, AWS, and the Python data ecosystem Architect complex applications that require large datasets and models, and a team of data scientists About the reader For infrastructure engineers and engineering-minded data scientists who are familiar with Python. About the author At Netflix, Ville Tuulos designed and built Metaflow, a full-stack framework for data science. Currently, he is the CEO of a startup focusing on data science infrastructure. Table of Contents 1 Introducing data science infrastructure 2 The toolchain of data science 3 Introducing Metaflow 4 Scaling with the compute layer 5 Practicing scalability and performance 6 Going to production 7 Processing data 8 Using and operating models 9 Machine learning with the full stack Learning IPython for Interactive Computing and Data Visualization Packt Publishing Ltd Get started with Python for data analysis and numerical computing in the Jupyter notebook About This Book Learn the basics of Python in the Jupyter Notebook Analyze and visualize data with pandas, NumPy, matplotlib, and seaborn Perform highly-efficient numerical computations with Numba, Cython, and ipyparallel Who This Book Is For This book targets students, teachers, researchers, engineers, analysts, journalists, hobbyists, and all data enthusiasts who are interested in analyzing and visualizing real-world datasets. If you are new to programming and data analysis, this book is exactly for you. If you're already familiar with another language or analysis software, you will also appreciate this introduction to the Python data analysis platform. Finally, there are more technical topics for advanced readers. No prior experience is required; this book contains everything you need to know. What You Will Learn Install Anaconda and code in Python in the Jupyter Notebook Load and explore datasets interactively Perform complex data manipulations effectively with pandas Create engaging data visualizations with matplotlib and seaborn Simulate mathematical models with NumPy Visualize and process images interactively in the Jupyter Notebook with scikit-image Accelerate your code with Numba, Cython, and IPython.parallel Extend the Notebook interface with HTML, JavaScript, and D3 In Detail Python is a user-friendly and powerful programming language. IPython offers a convenient interface to the language and its analysis libraries, while the Jupyter Notebook is a rich environment well-adapted to data science and visualization. Together, these open source tools are widely used by beginners and experts around the world, and in a huge variety of fields and endeavors. This book is a beginner-friendly guide to the Python data analysis platform. After an introduction to the Python

language, IPython, and the Jupyter Notebook, you will learn how to analyze and visualize data on real-world examples, how to create graphical user interfaces for image processing in the Notebook, and how to perform fast numerical computations for scientific simulations with NumPy, Numba, Cython, and ipyparallel. By the end of this book, you will be able to perform in-depth analyses of all sorts of data. Style and approach This is a hands-on beginner-friendly guide to analyze and visualize data on real-world examples with Python and the Jupyter Notebook. Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics IGI Global Analyzing data sets has continued to be an invaluable application for numerous industries. By combining different algorithms, technologies, and systems used to extract information from data and solve complex problems, various sectors have reached new heights and have changed our world for the better. The Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics is a collection of innovative research on the methods and applications of data analytics. While highlighting topics including artificial intelligence, data security, and information systems, this book is ideally designed for researchers, data analysts, data scientists, healthcare administrators, executives, managers, engineers, IT consultants, academicians, and students interested in the potential of data application technologies. Algorithms and Architectures for Parallel Processing 15th International Conference, ICA3PP 2015, Zhangjiajie, China, November 18-20, 2015, Proceedings, Part III Springer This four volume set LNCS 9528, 9529, 9530 and 9531 constitutes the refereed proceedings of the 15th International Conference on Algorithms and Architectures for Parallel Processing, ICA3PP 2015, held in Zhangjiajie, China, in November 2015. The 219 revised full papers presented together with 77 workshop papers in these four volumes were carefully reviewed and selected from 807 submissions (602 full papers and 205 workshop papers). The first volume comprises the following topics: parallel and distributed architectures; distributed and network-based computing and internet of things and cyber-physical-social computing. The second volume comprises topics such as big data and its applications and parallel and distributed algorithms. The topics of the third volume are: applications of parallel and distributed computing and service dependability and security in distributed and parallel systems. The covered topics of the fourth volume are: software systems and programming models and performance modeling and evaluation. Deep Learning and Parallel Computing Environment for Bioengineering Systems Academic Press Deep Learning and Parallel Computing Environment for Bioengineering Systems delivers a significant forum for the technical advancement of deep learning in parallel computing environment across bio-engineering diversified domains and its applications. Pursuing an interdisciplinary approach, it focuses on methods used to identify and acquire valid, potentially useful knowledge sources. Managing the gathered knowledge and applying it to multiple domains

including health care, social networks, mining, recommendation systems, image processing, pattern recognition and predictions using deep learning paradigms is the major strength of this book. This book integrates the core ideas of deep learning and its applications in bio engineering application domains, to be accessible to all scholars and academicians. The proposed techniques and concepts in this book can be extended in future to accommodate changing business organizations' needs as well as practitioners' innovative ideas. Presents novel, in-depth research contributions from a methodological/application perspective in understanding the fusion of deep machine learning paradigms and their capabilities in solving a diverse range of problems Illustrates the state-of-the-art and recent developments in the new theories and applications of deep learning approaches applied to parallel computing environment in bioengineering systems Provides concepts and technologies that are successfully used in the implementation of today's intelligent data-centric critical systems and multi-media Cloud-Big data Digital Radiography Physical Principles and Quality Control Springer This is the second edition of a well-received book that enriches the understanding of radiographers and radiologic technologists across the globe, and is designed to meet the needs of courses (units) on radiographic imaging equipment, procedures, production, and exposure. The book also serves as a supplement for courses that address digital imaging techniques, such as radiologic physics, radiographic equipment and quality control. In a broader sense, the purpose of the book is to meet readers' needs in connection with the change from film-based imaging to film-less or digital imaging; today, all radiographic imaging worldwide is based on digital imaging technologies. The book covers a wide range of topics to address the needs of members of various professional radiologic technology associations, such as the American Society of Radiologic Technologists, the Canadian Association of Medical Radiation Technologists, the College of Radiographers in the UK, and the Australian and New Zealand Societies for Radiographers. Big Data Analytics: Systems, Algorithms, Applications Springer Nature This book provides a comprehensive survey of techniques, technologies and applications of Big Data and its analysis. The Big Data phenomenon is increasingly impacting all sectors of business and industry, producing an emerging new information ecosystem. On the applications front, the book offers detailed descriptions of various application areas for Big Data Analytics in the important domains of Social Semantic Web Mining, Banking and Financial Services, Capital Markets, Insurance, Advertisement, Recommendation Systems, Bio-Informatics, the IoT and Fog Computing, before delving into issues of security and privacy. With regard to machine learning techniques, the book presents all the standard algorithms for learning – including supervised, semi-supervised and unsupervised techniques such as clustering and reinforcement learning techniques to perform collective Deep Learning. Multi-layered and nonlinear learning for Big Data are also covered. In turn, the book highlights real-life case studies

on successful implementations of Big Data Analytics at large IT companies such as Google, Facebook, LinkedIn and Microsoft. Multi-sectorial case studies on domain-based companies such as Deutsche Bank, the power provider Opower, Delta Airlines and a Chinese City Transportation application represent a valuable addition. Given its comprehensive coverage of Big Data Analytics, the book offers a unique resource for undergraduate and graduate students, researchers, educators and IT professionals alike. Parallel Programming for Multicore and Cluster Systems Springer Science & Business Media Innovations in hardware architecture, like hyper-threading or multicore processors, mean that parallel computing resources are available for inexpensive desktop computers. In only a few years, many standard software products will be based on concepts of parallel programming implemented on such hardware, and the range of applications will be much broader than that of scientific computing, up to now the main application area for parallel computing. Rauber and Rünger take up these recent developments in processor architecture by giving detailed descriptions of parallel programming techniques that are necessary for developing efficient programs for multicore processors as well as for parallel cluster systems and supercomputers. Their book is structured in three main parts, covering all areas of parallel computing: the architecture of parallel systems, parallel programming models and environments, and the implementation of efficient application algorithms. The emphasis lies on parallel programming techniques needed for different architectures. For this second edition, all chapters have been carefully revised. The chapter on architecture of parallel systems has been updated considerably, with a greater emphasis on the architecture of multicore systems and adding new material on the latest developments in computer architecture. Lastly, a completely new chapter on general-purpose GPUs and the corresponding programming techniques has been added. The main goal of the book is to present parallel programming techniques that can be used in many situations for a broad range of application areas and which enable the reader to develop correct and efficient parallel programs. Many examples and exercises are provided to show how to apply the techniques. The book can be used as both a textbook for students and a reference book for professionals. The material presented has been used for courses in parallel programming at different universities for many years. Azure Data Scientist Associate Certification Guide A hands-on guide to machine learning in Azure and passing the Microsoft Certified DP-100 exam Packt Publishing Ltd Develop the skills you need to run machine learning workloads in Azure and pass the DP-100 exam with ease Key FeaturesCreate end-to-end machine learning training pipelines, with or without codeTrack experiment progress using the cloud-based MLflow-compatible process of Azure ML servicesOperationalize your machine learning models by creating batch and real-time endpointsBook Description The Azure Data Scientist Associate Certification Guide helps you acquire practical knowledge for

machine learning experimentation on Azure. It covers everything you need to pass the DP-100 exam and become a certified Azure Data Scientist Associate. Starting with an introduction to data science, you'll learn the terminology that will be used throughout the book and then move on to the Azure Machine Learning (Azure ML) workspace. You'll discover the studio interface and manage various components, such as data stores and compute clusters. Next, the book focuses on no-code and low-code experimentation, and shows you how to use the Automated ML wizard to locate and deploy optimal models for your dataset. You'll also learn how to run end-to-end data science experiments using the designer provided in Azure ML Studio. You'll then explore the Azure ML Software Development Kit (SDK) for Python and advance to creating experiments and publishing models using code. The book also guides you in optimizing your model's hyperparameters using Hyperdrive before demonstrating how to use responsible AI tools to interpret and debug your models. Once you have a trained model, you'll learn to operationalize it for batch or real-time inferences and monitor it in production. By the end of this Azure certification study guide, you'll have gained the knowledge and the practical skills required to pass the DP-100 exam. What you will learnCreate a working environment for data science workloads on AzureRun data experiments using Azure Machine Learning servicesCreate training and inference pipelines using the designer or codeDiscover the best model for your dataset using Automated MLUse hyperparameter tuning to optimize trained modelsDeploy, use, and monitor models in productionInterpret the predictions of a trained modelWho this book is for This book is for developers who want to infuse their applications with AI capabilities and data scientists looking to scale their machine learning experiments in the Azure cloud. Basic knowledge of Python is needed to follow the code samples used in the book. Some experience in training machine learning models in Python using common frameworks like scikit-learn will help you understand the content more easily. IPython Interactive Computing and Visualization Cookbook, Second Edition Over 100 Hands-On Recipes to Sharpen Your Skills in High-performance Numerical Computing and Data Science in the Jupyter Notebook Learn to use IPython and Jupyter Notebook for your data analysis and visualization work. Key Features Leverage the Jupyter Notebook for interactive data science and visualization Become an expert in high-performance computing and visualization for data analysis and scientific modeling A comprehensive coverage of scientific computing through many hands-on, example-driven recipes with detailed, step-by-step explanations Book Description Python is one of the leading open source platforms for data science and numerical computing. IPython and the associated Jupyter Notebook offer efficient interfaces to Python for data analysis and interactive visualization, and they constitute an ideal gateway to the platform. IPython Interactive Computing and Visualization Cookbook, Second Edition contains many ready-to-use, focused recipes for high-performance scientific computing

and data analysis, from the latest IPython/Jupyter features to the most advanced tricks, to help you write better and faster code. You will apply these state-of-the-art methods to various real-world examples, illustrating topics in applied mathematics, scientific modeling, and machine learning. The first part of the book covers programming techniques: code quality and reproducibility, code optimization, high-performance computing through just-in-time compilation, parallel computing, and graphics card programming. The second part tackles data science, statistics, machine learning, signal and image processing, dynamical systems, and pure and applied mathematics. What you will learn Master all features of the Jupyter Notebook Code better: write high-quality, readable, and well-tested programs; profile and optimize your code; and conduct reproducible interactive computing experiments Visualize data and create interactive plots in the Jupyter Notebook Write blazingly fast Python programs with NumPy, ctypes, Numba, Cython, OpenMP, GPU programming (CUDA), parallel IPython, Dask, and more Analyze data with Bayesian or frequentist statistics (Pandas, PyMC, and R), and learn from actual data through machine learning (scikit-learn) Gain valuable insights into signals, images, and sounds with SciPy, scikit-image, and OpenCV Simulate deterministic and stochastic dynamical systems in Python Familiarize yourself with math in Python using SymPy and Sage: algebra, analysis, logic, graphs, geometry, and probability theory Who this book is for This book is intended for anyone interested in numerical computing and data science: students, researchers, teachers, engineers, analysts, and hobbyists. A basic knowledge of Python/NumPy is recommended. Some skills in mathematics will help you understand the theory behind the computational methods. Java for Data Science Packt Publishing Ltd Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal. Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often

statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This book follows a tutorial approach, providing examples of each of the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly. Java: Data Science Made Easy Packt Publishing Ltd Data collection, processing, analysis, and more About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples A highly practical course covering a broad set of topics - from the basics of Machine Learning to Deep Learning and Big Data frameworks. Who This Book Is For This course is meant for Java developers who are comfortable developing applications in Java, and now want to enter the world of data science or wish to build intelligent applications. Aspiring data scientists with some understanding of the Java programming language will also find this book to be very helpful. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing your existing Java stack, this book is for you! What You Will Learn Understand the key concepts of data science Explore the data science ecosystem available in Java Work with the Java APIs and techniques used to perform efficient data analysis Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images, and create your own search Learn how to build deep neural networks with DeepLearning4j Build data science applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this course, we cover the basic as well as advanced data science concepts and how they are implemented using the popular Java tools and libraries.The course starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a

discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. You will examine the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. Throughout this course, the chapters will illustrate a challenging data science problem, and then go on to present a comprehensive, Java-based solution to tackle that problem. You will cover a wide range of topics – from classification and regression, to dimensionality reduction and clustering, deep learning and working with Big Data. Finally, you will see the different ways to deploy the model and evaluate it in production settings. By the end of this course, you will be up and running with various facets of data science using Java, in no time at all. This course contains premium content from two of our recently published popular titles: Java for Data Science Mastering Java for Data Science Style and approach This course follows a tutorial approach, providing examples of each of the concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly. Data Intensive Computing Applications for Big Data IOS Press The book 'Data Intensive Computing Applications for Big Data' discusses the technical concepts of big data, data intensive computing through machine learning, soft computing and parallel computing paradigms. It brings together researchers to report their latest results or progress in the development of the above mentioned areas. Since there are few books on this specific subject, the editors aim to provide a common platform for researchers working in this area to exhibit their novel findings. The book is intended as a reference work for advanced undergraduates and graduate students, as well as multidisciplinary, interdisciplinary and transdisciplinary research workers and scientists on the subjects of big data and cloud/parallel and distributed computing, and explains didactically many of the core concepts of these approaches for practical applications. It is organized into 24 chapters providing a comprehensive overview of big data analysis using parallel computing and addresses the complete data science workflow in the cloud, as well as dealing with privacy issues and the challenges faced in a data-intensive cloud computing environment. The book explores both fundamental and high-level concepts, and will serve as a manual for those in the industry, while also helping beginners to understand the basic and advanced aspects of big data and cloud computing.