

Mastering Parallel Programming With R

In just 24 lessons of one hour or less, Sams Teach Yourself R in 24 Hours helps you learn all the R skills you need to solve a wide spectrum of real-world data analysis problems. You'll master the entire data analysis workflow, learning to build code that's efficient, reproducible, and suitable for sharing with others. This book's straightforward, step-by-step approach teaches you how to import, manipulate, summarize, model, and plot data with R; formalize your analytical code; and build powerful R packages using current best practices. Practical, hands-on examples show you how to apply what you learn. Quizzes and exercises help you test your knowledge and stretch your skills. Learn How To Install, configure, and explore the R environment, including RStudio Use basic R syntax, objects, and packages Create and manage data structures, including vectors, matrices, and arrays Understand lists and data frames Work with dates, times, and factors Use common R functions, and learn to write your own Import and export data and connect to databases and spreadsheets Use the popular tidyr, dplyr and data.table packages Write more efficient R code with profiling, vectorization, and initialization Plot data and extend your graphical capabilities with ggplot2 and Lattice graphics Develop common types of models Construct high-quality packages, both simple and complex Write R classes: S3, S4, and Reference Classes Use R to generate dynamic reports Build web applications with Shiny Register your book at informit.com/register for convenient access to updates and corrections as they become available. This book's source code can be found at <http://www.mango-solutions.com/wp/teach-yourself-r-in-24-hours-book>.

This monograph-like book assembles the thoroughly revised and cross-reviewed lectures given at the School on Data Parallelism, held in Les Menuires, France, in May 1996. The book is a unique survey on the current status and future perspectives of the currently very promising and popular data parallel programming model. Much attention is paid to the style of writing and complementary coverage of the relevant issues throughout the 12 chapters. Thus these lecture notes are ideally suited for advanced courses or self-instruction on data parallel programming. Furthermore, the book is indispensable reading for anybody doing research in data parallel programming and related areas.

An Essential Reference for Intermediate and Advanced R Programmers Advanced R presents useful tools and techniques for attacking many types of R programming problems, helping you avoid mistakes and dead ends. With more than ten years of experience programming in R, the author illustrates the elegance, beauty, and flexibility at the heart of R. The book develops the necessary skills to produce quality code that can be used in a variety of circumstances. You will learn: The fundamentals of R, including standard data types and functions Functional programming as a useful framework for solving wide classes of problems The positives and negatives of metaprogramming How to write fast, memory-efficient code This book not only helps current R users become R programmers but also shows existing programmers what's special about R. Intermediate R programmers can dive deeper into R and learn new strategies for solving diverse problems while programmers from other languages can learn the details of R and understand why R works the way it does.

The Book of R is a comprehensive, beginner-friendly guide to R, the world's most popular programming language for statistical analysis. Even if you have no programming experience and little more than a grounding in the basics of mathematics, you'll find everything you need to begin using R effectively for statistical analysis. You'll start with the basics, like how to handle data and write simple programs, before moving on to more advanced topics, like producing statistical summaries of your data and performing statistical tests and modeling. You'll even learn how to create impressive data visualizations with R's basic graphics tools and contributed packages, like ggplot2 and ggvis, as well as interactive 3D visualizations using the rgl package. Dozens of hands-on exercises (with downloadable solutions) take you from theory to practice, as you learn: –The fundamentals of programming in R, including how to write data frames, create functions, and use variables, statements, and loops –Statistical concepts like exploratory data analysis, probabilities, hypothesis tests, and regression modeling, and how to execute them in R –How to access R's thousands of functions, libraries, and data sets –How to draw valid and useful conclusions from your data –How to create publication-quality graphics of your results Combining detailed explanations with real-world examples and exercises, this book will provide you with a solid understanding of both statistics and the depth of R's functionality. Make The Book of R your doorway into the growing world of data analysis.

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

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

Understand the philosophy of the Clojure language and dive into its inner workings to unlock its advanced features, methodologies, and constructs About This Book Learn to handle data using sequences, reducers, and transducers in Clojure Explore the lesser known and more advanced features, constructs, and methodologies of the Clojure language and its ecosystem, such as asynchronous channels, actors, logic programming, and reactive programming Sharpen your Clojure skills through illustrative and comprehensive examples Who This Book Is For If you're looking to learn more about the core libraries and dive deep into the Clojure language, then this book is ideal for you. Prior knowledge of the Clojure language is required. What You Will Learn Maximize the impact of parallelization, functional composition, and process transformation by composing reducers and

transducers Process and manipulate data using sequences, reducers, and transducers in Clojure Modify and add features to the Clojure language using macros Explore the features of category theory and custom data sources for logic programming in Clojure Orchestrate parallelism and concurrency using built-in primitives as well as community libraries in Clojure Handle data with asynchronous and reactive programming methodologies and leverage it using the core.async library Test your code with unit tests, specs, and type checks to write testable code Troubleshoot and style your Clojure code to make it more maintainable In Detail Clojure is a general-purpose language from the Lisp family with an emphasis on functional programming. It has some interesting concepts and features such as immutability, gradual typing, thread-safe concurrency primitives, and macro-based metaprogramming, which makes it a great choice to create modern, performant, and scalable applications. Mastering Clojure gives you an insight into the nitty-gritty details and more advanced features of the Clojure programming language to create more scalable, maintainable, and elegant applications. You'll start off by learning the details of sequences, concurrency primitives, and macros. Packed with a lot of examples, you'll get a walkthrough on orchestrating concurrency and parallelism, which will help you understand Clojure reducers, and we'll walk through composing transducers so you know about functional composition and process transformation inside out. We also explain how reducers and transducers can be used to handle data in a more performant manner. Later on, we describe how Clojure also supports other programming paradigms such as pure functional programming and logic programming. Furthermore, you'll level up your skills by taking advantage of Clojure's powerful macro system. Parallel, asynchronous, and reactive programming techniques are also described in detail. Lastly, we'll show you how to test and troubleshoot your code to speed up your development cycles and allow you to deploy the code faster. Style and approach This is an easy-to-follow project-based guide that throws you directly into the excitement of Clojure code. Mastering Clojure is for anyone who is interested in expanding their knowledge of language features and advanced functional programming.

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions Master the craft of predictive modeling in R by developing strategy, intuition, and a solid foundation in essential concepts About This Book Grasping the major methods of predictive modeling and moving beyond black box thinking to a deeper level of

understanding Leveraging the flexibility and modularity of R to experiment with a range of different techniques and data types Packed with practical advice and tips explaining important concepts and best practices to help you understand quickly and easily Who This Book Is For Although budding data scientists, predictive modelers, or quantitative analysts with only basic exposure to R and statistics will find this book to be useful, the experienced data scientist professional wishing to attain master level status , will also find this book extremely valuable.. This book assumes familiarity with the fundamentals of R, such as the main data types, simple functions, and how to move data around. Although no prior experience with machine learning or predictive modeling is required, there are some advanced topics provided that will require more than novice exposure. What You Will Learn Master the steps involved in the predictive modeling process Grow your expertise in using R and its diverse range of packages Learn how to classify predictive models and distinguish which models are suitable for a particular problem Understand steps for tidying data and improving the performing metrics Recognize the assumptions, strengths, and weaknesses of a predictive model Understand how and why each predictive model works in R Select appropriate metrics to assess the performance of different types of predictive model Explore word embedding and recurrent neural networks in R Train models in R that can work on very large datasets In Detail R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems. The book begins with a dedicated chapter on the language of models and the predictive modeling process. You will understand the learning curve and the process of tidying data. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model works, how to use R to train it, and how to measure and assess its performance using real-world datasets. How do you train models that can handle really large datasets? This book will also show you just that. Finally, you will tackle the really important topic of deep learning by implementing applications on word embedding and recurrent neural networks. By the end of this book, you will have explored and tested the most popular modeling techniques in use on real- world datasets and mastered a diverse range of techniques in predictive analytics using R. Style and approach This book takes a step-by-step approach in explaining the intermediate to advanced concepts in predictive analytics. Every concept is explained in depth, supplemented with practical examples applicable in a real-world setting.

Gain sharp insights into your data and solve real-world data science problems with R—from data munging to modeling and visualization About This Book Handle your data with precision and care for optimal business intelligence Restructure and transform your data to inform decision-making Packed with practical advice and tips to help you get to grips with data mining Who This Book Is For If you are a data scientist or R developer who wants to explore and optimize your use of R's advanced features and tools, this is the book for you. A basic knowledge of R is required, along with an understanding of database logic. What You Will Learn Connect to and load data from R's range of powerful databases Successfully fetch and parse structured and unstructured data Transform and restructure your data with efficient R packages Define and build complex statistical models with

glm Develop and train machine learning algorithms Visualize social networks and graph data Deploy supervised and unsupervised classification algorithms Discover how to visualize spatial data with R In Detail R is an essential language for sharp and successful data analysis. Its numerous features and ease of use make it a powerful way of mining, managing, and interpreting large sets of data. In a world where understanding big data has become key, by mastering R you will be able to deal with your data effectively and efficiently. This book will give you the guidance you need to build and develop your knowledge and expertise. Bridging the gap between theory and practice, this book will help you to understand and use data for a competitive advantage. Beginning with taking you through essential data mining and management tasks such as munging, fetching, cleaning, and restructuring, the book then explores different model designs and the core components of effective analysis. You will then discover how to optimize your use of machine learning algorithms for classification and recommendation systems beside the traditional and more recent statistical methods. Style and approach Covering the essential tasks and skills within data science, Mastering Data Analysis provides you with solutions to the challenges of data science. Each section gives you a theoretical overview before demonstrating how to put the theory to work with real-world use cases and hands-on examples.

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.

In large projects, programmers tend to get overwhelmed by their complexity. It can be hard to keep track of all the interdependencies in the code-base and how its state changes on runtime. The solution to these problems is Functional Programming, a paradigm specifically designed to deal with the complexity of software development. Mastering ...

Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools. RStudio Master Instructor Garrett Golemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time. Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users Use R programming tools such as if else statements, for loops, and S3 classes Learn how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them

R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions in the real world. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems. This book is designed to be both a guide and a reference for moving beyond the basics of predictive modeling. The book begins with a dedicated chapter on the language of models and the predictive modeling process. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model works, how to use R to train it, and how to measure and assess its performance using real world data sets. By the end of this book, you will have explored and tested the most popular modeling techniques in use on real world data sets and mastered a diverse range of techniques in predictive analytics.

R is the world's most popular language for developing statistical software: Archaeologists use it to track the spread of ancient civilizations, drug companies use it to discover which medications are safe and effective, and actuaries use it to assess financial risks and keep economies running smoothly. The Art of R Programming takes you on a guided tour of software development with R, from basic types and data structures to advanced topics like closures, recursion, and anonymous functions. No statistical knowledge is required, and your programming skills can range from hobbyist to pro. Along the way, you'll learn about functional

and object-oriented programming, running mathematical simulations, and rearranging complex data into simpler, more useful formats. You'll also learn to: –Create artful graphs to visualize complex data sets and functions –Write more efficient code using parallel R and vectorization –Interface R with C/C++ and Python for increased speed or functionality –Find new R packages for text analysis, image manipulation, and more –Squash annoying bugs with advanced debugging techniques Whether you're designing aircraft, forecasting the weather, or you just need to tame your data, The Art of R Programming is your guide to harnessing the power of statistical computing.

Parallel Programming with Co-Arrays describes the basic techniques used to design parallel algorithms for high-performance, scientific computing. It is intended for upper-level undergraduate students and graduate students who need to develop parallel codes with little or no previous introduction to parallel computing. It is also intended as a reference manual for researchers active in the field of scientific computing. All the algorithms in the book are based on partition operators. These operators provide a unifying principle that fits seemingly disparate techniques into an overall framework for algorithm design. The book uses the co-array programming model to illustrate how to write code for concrete examples, but it emphasizes that the important concepts for algorithm design are independent of the programming model. With these concepts in mind, the reader can write algorithms in different programming models based on personal taste and comfort.

Master machine learning techniques with R to deliver insights for complex projects About This Book Get to grips with the application of Machine Learning methods using an extensive set of R packages Understand the benefits and potential pitfalls of using machine learning methods Implement the numerous powerful features offered by R with this comprehensive guide to building an independent R-based ML system Who This Book Is For If you want to learn how to use R's machine learning capabilities to solve complex business problems, then this book is for you. Some experience with R and a working knowledge of basic statistical or machine learning will prove helpful. What You Will Learn Gain deep insights to learn the applications of machine learning tools to the industry Manipulate data in R efficiently to prepare it for analysis Master the skill of recognizing techniques for effective visualization of data Understand why and how to create test and training data sets for analysis Familiarize yourself with fundamental learning methods such as linear and logistic regression Comprehend advanced learning methods such as support vector machines Realize why and how to apply unsupervised learning methods In Detail Machine learning is a field of Artificial Intelligence to build systems that learn from data. Given the growing prominence of R—a cross-platform, zero-cost statistical programming environment—there has never been a better time to start applying machine learning to your data. The book starts with introduction to Cross-Industry Standard Process for Data Mining. It takes you through Multivariate Regression in detail. Moving on, you will also address Classification and Regression trees. You will learn a couple of “Unsupervised techniques”. Finally, the book will walk you through text analysis and time series. The book will deliver practical and real-world solutions to problems and variety of tasks such as complex recommendation systems. By the end of this book, you will gain expertise in performing R machine learning and will be able to build complex ML projects using R and its packages. Style and approach This is a book explains complicated concepts with easy to follow theory and real-world, practical applications. It demonstrates the power of R and machine learning extensively while highlighting the constraints. Mastering Parallel Programming with R Packt Publishing Ltd

Deep dive into C# and .NET architecture to build efficient, powerful applications About This Book Uniquely structured content to help you understand what goes on under the hood of .NET's managed code platform to master .NET programming Deep dive into C# programming and how the code executes via the CLR Packed with hands-on practical examples, you'll understand how to write applications to make full use of the new features of .NET 4.6, .NET Core and C# 6/7 Who This Book Is For This book was written exclusively for .NET developers. If you've been creating C# applications for your clients, at work or at home, this book will help you develop the skills you need to create modern, powerful, and efficient applications in C#. No knowledge of C# 6/7 or .NET 4.6 is needed to follow along—all the latest features are included to help you start writing cross-platform applications immediately. You will need to be familiar with Visual Studio, though all the new features in Visual Studio 2015 will also be covered. What You Will Learn Understand C# core concepts in depth, from sorting algorithms to the Big O notation Get up to speed with the latest changes in C# 6/7 Interface SQL Server and NoSQL databases with .NET Learn SOLID principles and the most relevant GoF Patterns with practical examples in C# 6.0 Defend C# applications against attacks Use Roslyn, a self-hosted framework to compile and advanced edition in both C# and Visual basic .NET languages Discern LINQ and associated Lambda expressions, generics, and delegates Design a .NET application from the ground up Understand the internals of a .NET assembly Grasp some useful advanced features in optimization and parallelism In Detail Mastering C# and .NET Framework will take you in to the depths of C# 6.0/7.0 and .NET 4.6, so you can understand how the platform works when it runs your code, and how you can use this knowledge to write efficient applications. Take full advantage of the new revolution in .NET development, including open source status and cross-platform capability, and get to grips with the architectural changes of CoreCLR. Start with how the CLR executes code, and discover the niche and advanced aspects of C# programming – from delegates and generics, through to asynchronous programming. Run through new forms of type declarations and assignments, source code callers, static using syntax, auto-property initializers, dictionary initializers, null conditional operators, and many others. Then unlock the true potential of the .NET platform. Learn how to write OWASP-compliant applications, how to properly implement design patterns in C#, and how to follow the general SOLID principles and its implementations in C# code. We finish by focusing on tips and tricks that you'll need to get the most from C# and .NET. This book also covers .NET Core 1.1 concepts as per the latest RTM release in the last chapter. Style and approach This book uses hands-on practical code examples that will take you into the depths of C# and .NET. Packed with hands-on practical examples, it is great as a tutorial, or as a reference guide.

Python is one of the most popular programming languages, with numerous libraries and frameworks that facilitate high-performance computing. Concurrency and parallelism in Python are essential when it comes to multiprocessing and multithreading; they behave differently, but their common aim is to reduce the execution time. This book serves as a ...

Get to grips with the advanced concepts of interactive computing to make the most out of IPython About This Book Most updated book on Interactive computing with IPython 4.0; Detailed, example-rich guide that lets you use the most advanced level interactive programming with IPython; Get flexible interactive programming with IPython using this comprehensive guide Who This Book Is For This book is for IPython developers who want to make the most of IPython and perform advanced scientific computing with IPython utilizing the ease of interactive computing. It is ideal for users who wish to learn about the interactive and parallel computing properties of IPython 4.0, along with its integration with third-party tools and concepts such as testing and documenting results. What You Will Learn Develop skills to use IPython for high performance computing (HPC) Understand the IPython interactive shell Use XeroMQ and MPI to pass messages Integrate third-party tools like R, Julia, and JavaScript with IPython Visualize the data Acquire knowledge to test and document the data Get to grips with the

recent developments in the Jupyter notebook system In Detail IPython is an interactive computational environment in which you can combine code execution, rich text, mathematics, plots, and rich media. This book will get IPython developers up to date with the latest advancements in IPython and dive deep into interactive computing with IPython. This an advanced guide on interactive and parallel computing with IPython will explore advanced visualizations and high-performance computing with IPython in detail. You will quickly brush up your knowledge of IPython kernels and wrapper kernels, then we'll move to advanced concepts such as testing, Sphinx, JS events, interactive work, and the ZMQ cluster. The book will cover topics such as IPython Console Lexer, advanced configuration, and third-party tools. By the end of this book, you will be able to use IPython for interactive and parallel computing in a high-performance computing environment. Style and approach This is a comprehensive guide to IPython for interactive, exploratory and parallel computing. It will let the IPython get up to date with the latest advancements in IPython and dive deeper into interactive computing with IPython

Parallel and High Performance Computing offers techniques guaranteed to boost your code's effectiveness. Summary Complex calculations, like training deep learning models or running large-scale simulations, can take an extremely long time. Efficient parallel programming can save hours—or even days—of computing time. Parallel and High Performance Computing shows you how to deliver faster run-times, greater scalability, and increased energy efficiency to your programs by mastering parallel techniques for multicore processor and GPU hardware.

About the technology Write fast, powerful, energy efficient programs that scale to tackle huge volumes of data. Using parallel programming, your code spreads data processing tasks across multiple CPUs for radically better performance. With a little help, you can create software that maximizes both speed and efficiency. About the book Parallel and High Performance Computing offers techniques guaranteed to boost your code's effectiveness. You'll learn to evaluate hardware architectures and work with industry standard tools such as OpenMP and MPI.

You'll master the data structures and algorithms best suited for high performance computing and learn techniques that save energy on handheld devices. You'll even run a massive tsunami simulation across a bank of GPUs. What's inside Planning a new parallel project

Understanding differences in CPU and GPU architecture Addressing underperforming kernels and loops Managing applications with batch scheduling About the reader For experienced programmers proficient with a high-performance computing language like C, C++, or Fortran.

About the author Robert Robey works at Los Alamos National Laboratory and has been active in the field of parallel computing for over 30 years. Yuliana Zamora is currently a PhD student and Siebel Scholar at the University of Chicago, and has lectured on programming modern hardware at numerous national conferences.

Table of Contents PART 1 INTRODUCTION TO PARALLEL COMPUTING 1 Why parallel computing? 2 Planning for parallelization 3 Performance limits and profiling 4 Data design and performance models 5 Parallel algorithms and patterns PART 2 CPU: THE PARALLEL WORKHORSE 6 Vectorization: FLOPs for free 7 OpenMP that performs 8 MPI: The parallel backbone PART 3 GPUS: BUILT TO ACCELERATE 9 GPU architectures and concepts 10 GPU programming model 11 Directive-based GPU programming 12 GPU languages: Getting down to basics 13 GPU profiling and tools PART 4 HIGH PERFORMANCE COMPUTING ECOSYSTEMS 14 Affinity: Truce with the kernel 15 Batch schedulers: Bringing order to chaos 16 File operations for a parallel world 17 Tools and resources for better code

SAS software has been in existence for a long time and has been implemented in large, data-intensive environments, including data warehouses. This SAS book covers practical programming considerations to make when involving SAS in a data warehouse environment. You'll be able to develop the skills you need to apply SAS in your working environment.

The book provides an introduction to programming with Pascal and extends this to show how Borland Delphi is used to development

Microsoft Windows programs. It is packed full of real-life application and splits into three main sections: * Pascal programming* Pascal applications* Delphi programming Practical applications include: software interrupts, hardware interrupts, graphics, date and time, system commands, RS-232 and parallel ports.

Distributed systems intertwine with our everyday lives. The benefits and current shortcomings of the underpinning technologies are experienced by a wide range of people and their smart devices. With the rise of large-scale IoT and similar distributed systems, cloud bursting technologies, and partial outsourcing solutions, private entities are encouraged to increase their efficiency and offer unparalleled availability and reliability to their users. The Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing is a vital reference source that provides valuable insight into current and emergent research occurring within the field of distributed computing. It also presents architectures and service frameworks to achieve highly integrated distributed systems and solutions to integration and efficient management challenges faced by current and future distributed systems. Highlighting a range of topics such as data sharing, wireless sensor networks, and scalability, this multi-volume book is ideally designed for system administrators, integrators, designers, developers, researchers, academicians, and students.

If you want to learn how to quantitatively answer scientific questions for practical purposes using the powerful R language and the open source R tool ecosystem, this book is ideal for you. It is ideally suited for scientists who understand scientific concepts, know a little R, and want to be able to start applying R to be able to answer empirical scientific questions. Some R exposure is helpful, but not compulsory.

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.

Summary Modern data science solutions need to be clean, easy to read, and scalable. In Mastering Large Datasets with Python, author J.T. Wolohan teaches you how to take a small project and scale it up using a functionally influenced approach to Python coding. You'll explore methods and built-in Python tools that lend themselves to clarity and

scalability, like the high-performing parallelism method, as well as distributed technologies that allow for high data throughput. The abundant hands-on exercises in this practical tutorial will lock in these essential skills for any large-scale data science project. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Programming techniques that work well on laptop-sized data can slow to a crawl—or fail altogether—when applied to massive files or distributed datasets. By mastering the powerful map and reduce paradigm, along with the Python-based tools that support it, you can write data-centric applications that scale efficiently without requiring codebase rewrites as your requirements change. About the book Mastering Large Datasets with Python teaches you to write code that can handle datasets of any size. You'll start with laptop-sized datasets that teach you to parallelize data analysis by breaking large tasks into smaller ones that can run simultaneously. You'll then scale those same programs to industrial-sized datasets on a cluster of cloud servers. With the map and reduce paradigm firmly in place, you'll explore tools like Hadoop and PySpark to efficiently process massive distributed datasets, speed up decision-making with machine learning, and simplify your data storage with AWS S3. What's inside An introduction to the map and reduce paradigm Parallelization with the multiprocessing module and pathos framework Hadoop and Spark for distributed computing Running AWS jobs to process large datasets About the reader For Python programmers who need to work faster with more data. About the author J. T. Wolohan is a lead data scientist at Booz Allen Hamilton, and a PhD researcher at Indiana University, Bloomington. Table of Contents: PART 1 1 | Introduction 2 | Accelerating large dataset work: Map and parallel computing 3 | Function pipelines for mapping complex transformations 4 | Processing large datasets with lazy workflows 5 | Accumulation operations with reduce 6 | Speeding up map and reduce with advanced parallelization PART 2 7 | Processing truly big datasets with Hadoop and Spark 8 | Best practices for large data with Apache Streaming and mrjob 9 | PageRank with map and reduce in PySpark 10 | Faster decision-making with machine learning and PySpark PART 3 11 | Large datasets in the cloud with Amazon Web Services and S3 12 | MapReduce in the cloud with Amazon's Elastic MapReduce

Master the principles to make applications robust, scalable and responsive About This Book Implement concurrent applications using the Java 9 Concurrency API and its new components Improve the performance of your applications and process more data at the same time, taking advantage of all of your resources Construct real-world examples related to machine learning, data mining, natural language processing, and more Who This Book Is For This book is for competent Java developers who have basic understanding of concurrency, but knowledge of effective implementation of concurrent programs or usage of streams for making processes more efficient is not required What You Will Learn Master the principles that every concurrent application must follow See how to parallelize a sequential algorithm to obtain

better performance without data inconsistencies and deadlocks Get the most from the Java Concurrency API components Separate the thread management from the rest of the application with the Executor component Execute phased-based tasks in an efficient way with the Phaser components Solve problems using a parallelized version of the divide and conquer paradigm with the Fork / Join framework Find out how to use parallel Streams and Reactive Streams Implement the “map and reduce” and “map and collect” programming models Control the concurrent data structures and synchronization mechanisms provided by the Java Concurrency API Implement efficient solutions for some actual problems such as data mining, machine learning, and more In Detail Concurrency programming allows several large tasks to be divided into smaller sub-tasks, which are further processed as individual tasks that run in parallel. Java 9 includes a comprehensive API with lots of ready-to-use components for easily implementing powerful concurrency applications, but with high flexibility so you can adapt these components to your needs. The book starts with a full description of the design principles of concurrent applications and explains how to parallelize a sequential algorithm. You will then be introduced to Threads and Runnables, which are an integral part of Java 9's concurrency API. You will see how to use all the components of the Java concurrency API, from the basics to the most advanced techniques, and will implement them in powerful real-world concurrency applications. The book ends with a detailed description of the tools and techniques you can use to test a concurrent Java application, along with a brief insight into other concurrency mechanisms in JVM. Style and approach This is a complete guide that implements real-world examples of algorithms related to machine learning, data mining, and natural language processing in client/server environments. All the examples are explained using a step-by-step approach.

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.

This book is for programmers and developers who want to improve the performance of their R programs by making them run faster with large data sets or who are trying to solve a pesky performance problem.

Advanced analytics on your Big Data with latest Apache Spark 2.x About This Book An advanced guide with a combination of instructions and practical examples to extend the most up-to date Spark functionalities. Extend your data processing capabilities to process huge chunk of data in minimum time using advanced concepts in Spark. Master the art of real-time processing with the help of Apache Spark 2.x Who This Book Is For If you are a developer with some experience with Spark and want to strengthen your knowledge of how to get around in the world of Spark, then this book is ideal for you. Basic knowledge of Linux, Hadoop and Spark is assumed. Reasonable knowledge of Scala is expected. What You Will Learn Examine Advanced Machine Learning and DeepLearning with MLib, SparkML, SystemML, H2O

and DeepLearning4J Study highly optimised unified batch and real-time data processing using SparkSQL and Structured Streaming Evaluate large-scale Graph Processing and Analysis using GraphX and GraphFrames Apply Apache Spark in Elastic deployments using Jupyter and Zeppelin Notebooks, Docker, Kubernetes and the IBM Cloud Understand internal details of cost based optimizers used in Catalyst, SystemML and GraphFrames Learn how specific parameter settings affect overall performance of an Apache Spark cluster Leverage Scala, R and python for your data science projects In Detail Apache Spark is an in-memory cluster-based parallel processing system that provides a wide range of functionalities such as graph processing, machine learning, stream processing, and SQL. This book aims to take your knowledge of Spark to the next level by teaching you how to expand Spark's functionality and implement your data flows and machine/deep learning programs on top of the platform. The book commences with an overview of the Spark ecosystem. It will introduce you to Project Tungsten and Catalyst, two of the major advancements of Apache Spark 2.x. You will understand how memory management and binary processing, cache-aware computation, and code generation are used to speed things up dramatically. The book extends to show how to incorporate H2O, SystemML, and DeepLearning4j for machine learning, and Jupyter Notebooks and Kubernetes/Docker for cloud-based Spark. During the course of the book, you will learn about the latest enhancements to Apache Spark 2.x, such as interactive querying of live data and unifying DataFrames and Datasets. You will also learn about the updates on the APIs and how DataFrames and Datasets affect SQL, machine learning, graph processing, and streaming. You will learn to use Spark as a big data operating system, understand how to implement advanced analytics on the new APIs, and explore how easy it is to use Spark in day-to-day tasks. Style and approach This book is an extensive guide to Apache Spark modules and tools and shows how Spark's functionality can be extended for real-time processing and storage with worked examples.

The high-level language of R is recognized as one of the most powerful and flexible statistical software environments, and is rapidly becoming the standard setting for quantitative analysis, statistics and graphics. R provides free access to unrivalled coverage and cutting-edge applications, enabling the user to apply numerous statistical methods ranging from simple regression to time series or multivariate analysis. Building on the success of the author's bestselling *Statistics: An Introduction using R*, *The R Book* is packed with worked examples, providing an all inclusive guide to R, ideal for novice and more accomplished users alike. The book assumes no background in statistics or computing and introduces the advantages of the R environment, detailing its applications in a wide range of disciplines. Provides the first comprehensive reference manual for the R language, including practical guidance and full coverage of the graphics facilities. Introduces all the statistical models covered by R, beginning with simple classical tests such as chi-square and t-test. Proceeds to examine more advanced methods, from regression and analysis of variance, through to generalized linear models, generalized mixed models, time series, spatial statistics, multivariate statistics and much more. *The R Book* is aimed at undergraduates, postgraduates and professionals in science, engineering and medicine. It is also ideal for students and professionals in statistics, economics, geography and the social sciences.

Download Ebook Mastering Parallel Programming With R

Julia is a well-constructed programming language with fast execution speed, eliminating the classic problem of performing analysis in one language and translating it for performance into a second. This book will help you develop and enhance your programming skills in Julia to solve real-world automation challenges. This book starts off with a refresher on installing and running Julia on different platforms. Next, you will compare the different ways of working with Julia and explore Julia's key features in-depth by looking at design and build. You will see how data works using simple statistics and analytics, and discover Julia's speed, its real strength, which makes it particularly useful in highly intensive computing tasks and observe how Julia can cooperate with external processes in order to enhance graphics and data visualization. Finally, you will look into meta-programming and learn how it adds great power to the language and establish networking and distributed computing with Julia.

Extract valuable data from your social media sites and make better business decisions using R About This Book Explore the social media APIs in R to capture data and tame it Employ the machine learning capabilities of R to gain optimal business value A hands-on guide with real-world examples to help you take advantage of the vast opportunities that come with social media data Who This Book Is For If you have basic knowledge of R in terms of its libraries and are aware of different machine learning techniques, this book is for you. Those with experience in data analysis who are interested in mining social media data will find this book useful. What You Will Learn Access APIs of popular social media sites and extract data Perform sentiment analysis and identify trending topics Measure CTR performance for social media campaigns Implement exploratory data analysis and correlation analysis Build a logistic regression model to detect spam messages Construct clusters of pictures using the K-means algorithm and identify popular personalities and destinations Develop recommendation systems using Collaborative Filtering and the Apriori algorithm In Detail With an increase in the number of users on the web, the content generated has increased substantially, bringing in the need to gain insights into the untapped gold mine that is social media data. For computational statistics, R has an advantage over other languages in providing readily-available data extraction and transformation packages, making it easier to carry out your ETL tasks. Along with this, its data visualization packages help users get a better understanding of the underlying data distributions while its range of "standard" statistical packages simplify analysis of the data. This book will teach you how powerful business cases are solved by applying machine learning techniques on social media data. You will learn about important and recent developments in the field of social media, along with a few advanced topics such as Open Authorization (OAuth). Through practical examples, you will access data from R using APIs of various social media sites such as Twitter, Facebook, Instagram, GitHub, Foursquare, LinkedIn, Blogger, and other networks. We will provide you with detailed explanations on the implementation of various use cases using R programming. With this handy guide, you will be ready to embark on your journey as an independent social media analyst. Style and approach This easy-to-follow guide is packed with hands-on, step-by-step examples that will enable you to convert your real-world social media data into useful, practical information.

Mastering Cloud Computing is designed for undergraduate students learning to develop cloud computing applications. Tomorrow's applications won't live on a single computer but will be deployed from and reside on a virtual server, accessible anywhere, any time. Tomorrow's application developers need to understand the requirements of building apps for these virtual systems, including concurrent programming, high-performance computing, and data-intensive systems. The book introduces the principles of distributed and parallel computing underlying cloud architectures and specifically focuses on virtualization, thread programming, task programming, and map-reduce programming. There are examples demonstrating all of these and more, with exercises and labs throughout. Explains how to make design

Download Ebook Mastering Parallel Programming With R

choices and tradeoffs to consider when building applications to run in a virtual cloud environment Real-world case studies include scientific, business, and energy-efficiency considerations

Step-by-step guide written in a lucid language for mastering C# DESCRIPTION This book starts by introducing the concepts of .NET framework. It then discusses OOP and explores how one can work with OOP in C#. There are two chapters on OOP: the first one covers the basics of object-oriented programming (OOP); and the second one delineates advanced concepts related to OOP and how they can be implemented in C#. Next, the book discusses Language Integrated Query (LINQ) and how to work with it in C#, followed by multithreading, asynchronous and parallel programming concepts with relevant code examples to illustrate the concepts covered. Generics, collections, generic collections, delegates, lambda expressions are also covered in this section. In the last section of the book, serialization, file I/O and how to work with them in C# are discussed concisely. A separate chapter on C# 8.0 is added to highlight its new features. .. There is an appendix chapter as well that discusses how one can get started working with Visual Studio 2019. KEY FEATURES Packed with plentiful code examples A comprehensive guide to mastering C# Discusses the OOP principles Targets beginner to advanced readers of C# Provides insight into the latest features of C#, including C# 8.0 Covers Visual Studio 2019 WHAT WILL YOU LEARN By the end of reading this book, the reader will have mastered the concepts of C# programming language, and be able to build high performance, scalable and robust applications using C#. WHO THIS BOOK IS FOR This book is intended for people who are aspiring a career in .NET as well as for professionals who would like to master the concepts of C# programming language. Since this book is for beginner to advanced readers of C#, a basic knowledge of C# will be helpful. Table of Contents 1. Microsoft .NET Internals 2. Getting Started with C# 3. Object-oriented Programming - Part 1 4. Object-oriented Programming - Part 2 5. Properties and Indexers 6. Programming LINQ in C# 7. Generic Collections in C# 8. Delegates and Extension Methods 9. Exception Handling in C# 10. Asynchrony and Parallel programming 11. Multithreading in C# 12. File I/O and Serialization 13. Advanced Concepts in C# 14. New Features of C# 8.0 15 Appendix

Master the Shiny web framework—and take your R skills to a whole new level. By letting you move beyond static reports, Shiny helps you create fully interactive web apps for data analyses. Users will be able to jump between datasets, explore different subsets or facets of the data, run models with parameter values of their choosing, customize visualizations, and much more. Hadley Wickham from RStudio shows data scientists, data analysts, statisticians, and scientific researchers with no knowledge of HTML, CSS, or JavaScript how to create rich web apps from R. This in-depth guide provides a learning path that you can follow with confidence, as you go from a Shiny beginner to an expert developer who can write large, complex apps that are maintainable and performant. Get started: Discover how the major pieces of a Shiny app fit together Put Shiny in action: Explore Shiny functionality with a focus on code samples, example apps, and useful techniques Master reactivity: Go deep into the theory and practice of reactive programming and examine reactive graph components Apply best practices: Examine useful techniques for making your Shiny apps work well in production

CUDA is a computing architecture designed to facilitate the development of parallel programs. In conjunction with a comprehensive software platform, the CUDA Architecture enables programmers to draw on the immense power of graphics processing units (GPUs) when building high-performance applications. GPUs, of course, have long been available for demanding graphics and game applications. CUDA now brings this valuable resource to programmers working on applications in other domains, including science, engineering, and finance. No knowledge of graphics programming is required—just the ability to program in a modestly extended version of C. CUDA by Example, written by two senior members of the CUDA software platform team, shows programmers how to employ this new technology. The authors introduce each area of

CUDA development through working examples. After a concise introduction to the CUDA platform and architecture, as well as a quick-start guide to CUDA C, the book details the techniques and trade-offs associated with each key CUDA feature. You'll discover when to use each CUDA C extension and how to write CUDA software that delivers truly outstanding performance. Major topics covered include Parallel programming Thread cooperation Constant memory and events Texture memory Graphics interoperability Atomics Streams CUDA C on multiple GPUs Advanced atomics Additional CUDA resources All the CUDA software tools you'll need are freely available for download from NVIDIA. <http://developer.nvidia.com/object/cuda-by-example.html>

Unlock the complexities of machine learning algorithms in Spark to generate useful data insights through this data analysis tutorial About This Book Process and analyze big data in a distributed and scalable way Write sophisticated Spark pipelines that incorporate elaborate extraction Build and use regression models to predict flight delays Who This Book Is For Are you a developer with a background in machine learning and statistics who is feeling limited by the current slow and "small data" machine learning tools? Then this is the book for you! In this book, you will create scalable machine learning applications to power a modern data-driven business using Spark. We assume that you already know the machine learning concepts and algorithms and have Spark up and running (whether on a cluster or locally) and have a basic knowledge of the various libraries contained in Spark. What You Will Learn Use Spark streams to cluster tweets online Run the PageRank algorithm to compute user influence Perform complex manipulation of DataFrames using Spark Define Spark pipelines to compose individual data transformations Utilize generated models for off-line/on-line prediction Transfer the learning from an ensemble to a simpler Neural Network Understand basic graph properties and important graph operations Use GraphFrames, an extension of DataFrames to graphs, to study graphs using an elegant query language Use K-means algorithm to cluster movie reviews dataset In Detail The purpose of machine learning is to build systems that learn from data. Being able to understand trends and patterns in complex data is critical to success; it is one of the key strategies to unlock growth in the challenging contemporary marketplace today. With the meteoric rise of machine learning, developers are now keen on finding out how can they make their Spark applications smarter. This book gives you access to transform data into actionable knowledge. The book commences by defining machine learning primitives by the MLlib and H2O libraries. You will learn how to use Binary classification to detect the Higgs Boson particle in the huge amount of data produced by CERN particle collider and classify daily health activities using ensemble Methods for Multi-Class Classification. Next, you will solve a typical regression problem involving flight delay predictions and write sophisticated Spark pipelines. You will analyze Twitter data with help of the doc2vec algorithm and K-means clustering. Finally, you will build different pattern mining models using MLlib, perform complex manipulation of DataFrames using Spark and Spark SQL, and deploy your app in a Spark streaming environment. Style and approach This book takes a practical approach to help you get to grips with using Spark for analytics and to implement machine learning algorithms. We'll teach you about advanced applications of machine learning through illustrative examples. These examples will equip you to harness the potential of machine learning, through Spark, in a variety of enterprise-grade systems.

[Copyright: 1d399c9167a9fa3ebf92ad2e8ff496eb](http://www.it-ebooks.info/book/1d399c9167a9fa3ebf92ad2e8ff496eb)