

Parallel Computing For Data Science With Examples In R C

The rapid advance of Internet of Things (IoT) technologies has resulted in the number of IoT-connected devices growing exponentially, with billions of connected devices worldwide. While this development brings with it great opportunities for many fields of science, engineering, business and everyday life, it also presents challenges such as an architectural bottleneck – with a very large number of IoT devices connected to a rather small number of servers in Cloud data centers – and the problem of data deluge. Edge computing aims to alleviate the computational burden of the IoT for the Cloud by pushing some of the computations and logics of processing from the Cloud to the Edge of the Internet. It is becoming commonplace to allocate tasks and applications such as data filtering, classification, semantic enrichment and data aggregation to this layer, but to prevent this new layer from itself becoming another bottleneck for the whole computing stack from IoT to the Cloud, the Edge computing layer needs to be capable of implementing massively parallel and distributed algorithms efficiently. This book, Advances in Edge Computing: Massive Parallel Processing and Applications, addresses these challenges in 11 chapters. Subjects covered include: Fog storage software architecture; IoT-based crowdsourcing; the industrial Internet of Things; privacy issues; smart home management in the Cloud and the Fog; and a cloud robotic solution to assist medical applications. Providing an overview of developments in the field, the book will be of interest to all those working with the Internet of Things and Edge computing.

Parallel Programming: Concepts and Practice provides an upper level introduction to parallel programming. In addition to covering general parallelism concepts, this text teaches practical programming skills for both shared memory and distributed memory architectures. The authors ’ open-source system for automated code evaluation provides easy access to parallel computing resources, making the book particularly suitable for classroom settings. Covers parallel programming approaches for single computer nodes and HPC clusters. OpenMP, multithreading, SIMD vectorization, MPI, UPC++ Contains numerous practical parallel programming exercises Includes access to an automated code evaluation tool that enables students the opportunity to program in a web browser and receive immediate feedback on the result validity of their program Features an example-based teaching of concept to enhance learning outcomes The book provides a practical guide to computational scientists and engineers to help advance their research by exploiting the superpower of supercomputers with many processors and complex networks. This book focuses on the design and analysis of basic parallel algorithms, the key components for composing larger packages for a wide range of applications.

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

Patterns for Parallel Programming

A Tour of Data Science

Handbook of Parallel Computing and Statistics

A Python Approach to Concepts, Techniques and Applications

Concepts and Practice

Introduction to Data Science

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

The Parallel Programming Guide for Every Software Developer From grids and clusters to next-generation game consoles, parallel computing is going mainstream. Innovations such as Hyper-Threading Technology, HyperTransport Technology, and multicore microprocessors from IBM, Intel, and Sun are accelerating the movement's growth. Only one thing is missing: programmers with the skills to meet the soaring demand for parallel software. That's where Patterns for Parallel Programming comes in. It's the first parallel programming guide written specifically to serve working software developers, not just computer scientists. The authors introduce a complete, highly accessible pattern language that will help any experienced developer "think parallel" and start writing effective parallel code almost immediately. Instead of formal theory, they deliver proven solutions to the challenges faced by parallel programmers, and pragmatic guidance for using today's parallel APIs in the real world. Coverage includes: Understanding the parallel computing landscape and the challenges faced by parallel developers Finding the concurrency in a software design problem and decomposing it into concurrent tasks Managing the use of data across tasks Creating an algorithm structure that effectively exploits the concurrency you've identified Connecting your algorithmic structures to the APIs needed to implement them Specific software constructs for implementing parallel programs Working with today's leading parallel programming environments: OpenMP, MPI, and Java Patterns have helped thousands of programmers master object-oriented development and other complex programming technologies. With this book, you will learn that they're the best way to master parallel programming too.

This book constitutes the proceedings of the 16th International Conference on Parallel Computing Technologies, PaCT 2021, which was held during September 13-18, 2021. The conference was planned to take place in Kaliningrad, Russia, but changed to an online event due to the COVID-19 pandemic. The 24 full and 12 short papers included in this book were carefully reviewed and selected from 62 submissions. They were organized in topical sections as follows: parallel programming methods and tools; applications; memory-efficient data structures; experimental studies; job management; essential algorithms; computing services; and cellular automata.

Scientific computing has become an indispensable tool in numerous fields, such as physics, mechanics, biology, finance and industry. For example, it enables us, thanks to efficient algorithms adapted to current computers, to simulate, without the help of models or experimentations, the deflection of beams in bending, the sound level in a theater room or a fluid flowing around an aircraft wing. This book presents the scientific computing techniques applied to parallel computing for the numerical simulation of large-scale problems; these problems result from systems modeled by partial differential equations. Computing concepts will be tackled via examples. Implementation and programming techniques resulting from the finite element method will be presented for direct solvers, iterative solvers and domain decomposition methods, along with an introduction to MPI and OpenMP.

Parallel Computing for Data Science

Encyclopedia of Parallel Computing

5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I

Introduction to Parallel Programming

Advances in Computing and Data Sciences

Big Data and High Performance Computing

There is a software gap between the hardware potential and the performance that can be attained using today's software parallel program development tools. The tools need manual intervention by the programmer to parallelize the code. Programming a parallel computer requires closely studying the target algorithm or application, more so than in the traditional sequential programming we have all learned. The programmer must be aware of the communication and data dependencies of the algorithm or application. This book provides the techniques to explore the possible ways to program a parallel computer for a given application.

Parallel Scientific Computing and Optimization introduces new developments in the construction, analysis, and implementation of parallel computing algorithms. This book presents 23 self-contained chapters, including survey chapters and surveys, written by distinguished researchers in the field of parallel computing. Each chapter is devoted to some aspects of the subject: parallel algorithms for matrix computations, parallel optimization, management of parallel programming models and data, with the largest focus on parallel scientific computing in industrial applications. This volume is intended for scientists and graduate students specializing in computer science and applied mathematics who are engaged in parallel scientific computing.

Parallel computing has been the enabling technology of high-end machines for many years. Now, it has finally become the ubiquitous key to the efficient use of any kind of multi-processor computer architecture, from smart phones, tablets, embedded systems and cloud computing up to exascale computers. _x000D_ This book presents the proceedings of ParCo2013 - the latest edition of the biennial International Conference on Parallel Computing - held from 10 to 13 September 2013, in Garching, Germany. The conference focused on several key parallel computing areas. Themes included parallel programming models for multi- and manycore CPUs, GPUs, FPGAs and heterogeneous platforms, the performance engineering processes that must be adapted to efficiently use these new and innovative platforms, novel numerical algorithms and approaches to large-scale simulations of problems in science and engineering. _x000D_ The conference programme also included twelve mini-symposia (including an industry session and a special PhD Symposium), which comprehensively represented and intensified the discussion of current hot topics in high performance and parallel computing. These special sessions covered large-scale supercomputing, novel challenges arising from parallel architectures (multi-/manycore, heterogeneous platforms, FPGAs), multi-level algorithms as well as multi-scale, multi-physics and multi-dimensional problems. _x000D_ It is clear that parallel computing - including the processing of large data sets ("Big Data") - will remain a persistent driver of research in all fields of innovative computing, which makes this book relevant to all those with an interest in this field.

Mathematics of Computing -- Parallelism.

R Programming for Data Science

Applied Parallel Computing

Vector Models for Data-parallel Computing

Data Analysis in the Distributed World

Introduction to Parallel Computing

Models, Algorithms and Applications

Parallel processing has been an enabling technology in scientific computing for more than 20 years. This book is the first in-depth discussion of parallel computing in 10 years; it reflects the mix of topics that mathematicians, computer scientists, and computational scientists focus on to make parallel processing effective for scientific problems. Presently, the impact of parallel processing on scientific computing varies greatly across disciplines, but it plays a vital role in most problem domains and is absolutely essential in many of them. Parallel Processing for Scientific Computing is divided into four parts: The first concerns performance modeling, analysis, and optimization; the second focuses on parallel algorithms and software for an array of problems common to many modeling and simulation applications; the third emphasizes tools and environments that can ease and enhance the process of application development; and the fourth provides a sampling of applications that require parallel computing for scaling to solve larger and realistic models that can advance science and engineering.

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.

New insight in many scientific and engineering fields is unthinkable without the use of numerical simulations running efficiently on modern computers. The faster we get new results, the bigger and accurate are the problems that we can solve. It is the combination of mathematical ideas plus efficient programming that drives the progress in many disciplines. Future champions in the area thus will have to be qualified in their application domain, they will need a profound understanding of some mathematical ideas, and they need the skills to deliver fast code. The present textbook targets students which have programming skills already and do not shy away from mathematics, though they might be educated in computer science or an application domain. It introduces the basic concepts and ideas behind applied mathematics and parallel programming that we need to write numerical simulations for today ’ s multicore workstations. Our intention is not to dive into one particular application domain or to introduce a new programming language – we lay the generic foundations for future courses and projects in the area. The text is written in an accessible style which is easy to digest for students without years and years of mathematics education. It values clarity and intuition over formalism, and uses a simple N-body simulation setup to illustrate basic ideas that are of relevance in various different subdomains of scientific computing. Its primary goal is to make theoretical and paradigmatic ideas accessible to undergraduate students and to bring the fascination of the field across.

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.

Parallel and High Performance Computing

Big Data Analysis with Python

Parallel R

A Seamless Approach to Parallel Algorithms and their Implementation

From Algorithms to Programming on State-of-the-Art Platforms

Parallel Computing: Accelerating Computational Science and Engineering (CSE)

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

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 Amdahl 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.

An overview of the most prominent contemporary parallel processing programming models, written in a unique tutorial style. With the coming of the parallel computing era, computer scientists have turned their attention to designing programming models that are suited for high-performance parallel computing and supercomputing systems. Programming parallel systems is complicated by the fact that multiple processing units are simultaneously computing and moving data. This book offers an overview of some of the most prominent parallel programming models used in high-performance computing and supercomputing systems today. The chapters describe the programming models in a unique tutorial style rather than using the formal approach taken in the research literature. The aim is to cover a wide range of parallel programming models, enabling the reader to understand what each has to offer. The book begins with a description of the Message Passing Interface (MPI), the most common parallel programming model for distributed memory computing. It goes on to cover one-sided communication models, ranging from low-level runtime libraries (GASNet, OpenSHMEM) to high-level programming models (UPC, GA, Chapel); task-oriented programming models (Charm++, ADLB, Scioto, Swift, CnC) that allow users to describe their computation and data units as tasks so that the runtime system can manage computation and data movement as necessary; and parallel programming models intended for on-node parallelism in the context of multicore architecture or attached accelerators (OpenMP, Cilk Plus, TBB, CUDA, OpenCL). The book will be a valuable resource for graduate students, researchers, and any scientist who works with data sets and large computations. Contributors Timothy Armstrong, Michael G. Burke, Ralph Butler, Bradford L. Chamberlain, Sunita Chandrasekaran, Barbara Chapman, Jeff Daily, James Dinan, Deepak Eachempati, Ian T. Foster, William D. Gropp, Paul Hargrove, Wen-mei Hwu, Nikhil Jain, Laxmikant Kale, David Kirk, Kath Knoeb, Ariram Krishnamoorthy, Jeffery A. Kuehn, Alexey Kukanov, Charles E. Leiserson, Jonathan Lifflander, Ewing Lusk, Tim Mattson, Bruce Palmer, Steven C. Pieper, Stephen W. Poole, Arch D. Robison, Frank Schlimbach, Rajeev Thakur, Abhinav Vishnu, Justin M. Wozniak, Michael Wilde, Kathy Yelick, Yili Zheng

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.

Parallel Processing for Scientific Computing Handbook of Parallel Computing With Examples in R, C++ and CUDA Parallel Computing Technologies Algorithms and Parallel Computing Data Science with Python and Dask

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

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. This contributed volume highlights two areas of fundamental interest in high-performance computing: core algorithms for important kernels and computationally demanding applications. The first few chapters explore algorithms, numerical techniques, and their parallel formulations for a variety of kernels that arise in applications. The rest of the volume focuses on state-of-the-art applications from diverse domains. By structuring the volume around these two areas, it presents a comprehensive view of the application landscape for high-performance computing, while also enabling readers to develop new applications using the kernels. Readers will learn how to choose the most suitable parallel algorithms for any given application, ensuring that theory and practicality are clearly connected. Applications using these techniques are illustrated in detail, including: Computational materials science and engineering Computational cardiovascular analysis Multiscale analysis of wind turbines and turbomachinery Weather forecasting Machine learning techniques Parallel Algorithms in Computational Science and Engineering will be an ideal reference for applied mathematicians, engineers, computer scientists, and other researchers who utilize high-performance computing in their work.

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

Advances in Edge Computing: Massive Parallel Processing and Applications

Data-parallel Programming on MIMD Computers

Data Intensive Computing Applications for Big Data

16th International Conference, PaCT 2021, Kaliningrad, Russia, September 13–18, 2021, Proceedings

Learn R and Python in Parallel

Deep Learning and Parallel Computing Environment for Bioengineering Systems

Technological improvements continue to push back the frontier of processor speed in modern computers. Unfortunately, the computational intensity demanded by modern research problems grows even faster. Parallel computing has emerged as the most successful bridge to this computational gap, and many popular solutions have emerged based on its concepts

In modern computer science, there exists no truly sequential computing system; and most advanced programming is parallel programming. This is particularly evident in modern application domains like scientific computation, data science, machine intelligence, etc. This lucid introductory textbook will be invaluable to students of computer science and technology, acting as a self-contained primer to parallel programming. It takes the reader from introduction to expertise, addressing a broad gamut of issues. It covers different parallel programming styles, describes parallel architecture, includes parallel programming frameworks and techniques, presents algorithmic and analysis techniques and discusses parallel design and performance issues. With its broad coverage, the book can be useful in a wide range of courses; and can also prove useful as a ready reckoner for professionals in the field.

Advancements in microprocessor architecture, interconnection technology, and software development have fueled rapid growth in parallel and distributed computing. However, this development is only of practical benefit if it is accompanied by progress in the design, analysis and programming of parallel algorithms. This concise textbook provides, in one place, three mainstream parallelization approaches, Open MPP, MPI and OpenCL, for multicore computers, interconnected computers and graphical processing units. An overview of practical parallel computing and principles will enable the reader to design efficient parallel programs for solving various computational problems on state-of-the-art personal computers and computing clusters. Topics covered range from parallel algorithms, programming tools, OpenMP, MPI and OpenCL, followed by experimental measurements of parallel programs' run-times, and by engineering analysis of obtained results for improved parallel execution performances. Many examples and exercises support the exposition.

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.

A First Guide to Numerical Concepts and Programming Methods

IoT Big Data Stream Processing

Programming Models for Parallel Computing

With Examples in R, C++ and Cuda

Combine Spark and Python to unlock the powers of parallel computing and machine learning

Parallel Computing Architectures and APIs

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. Parallel computing has become a 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 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 on the concepts and techniques of parallel computing. The book also provides 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 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

The ability of parallel computing to process large data sets and handle time-consuming operations has resulted in unprecedented advances in biological and scientific computing, modeling, and simulations. Exploring these recent developments, the Handbook of Parallel Computing: Models, Algorithms, and Applications provides comprehensive coverage on a variety of parallel computing architectures and APIs: IoT Big Data Stream Processing comes from the point high-performance uniprocessors were becoming increasingly complex, expensive, and power-hungry. A basic trade-off exists between the use of one or a small number of such complex processors, at one extreme, and a moderate to very large number of simpler processors, at the other. When you use parallel computing, you must understand its architectures so that you will be able to exploit those architectures during programming via the standardized APIs. This book would be useful for analysts, designers and developers of high-throughput computing systems essential for big data stream processing emanating from IoT-driven cyber-physical systems (CPS). This pragmatic book: Devolves uniprocessors in terms of their characteristics at a particular level of abstraction Explains limitations of uniprocessor high performance because of Moore's Law Introduces basics of processors, networks and distributed systems Explains characteristics of parallel systems, parallel computing models and parallel algorithms Explains the three primary categorical representatives of parallel computing architectures, namely, shared memory multiprocessors, distributed memory multiprocessors, and hybrid multiprocessors Provides an overview of Internet of Things (IoT), wireless sensor networks (WSN), sensor data processing, Big Data and stream processing Provides introduction to 5G communications, Edge and Fog computing Parallel Computing Architectures and APIs: IoT Big Data Stream Processing discusses the challenges of processing and analysis of high-volume, heterogeneous, continuous Internet of Things (IoT) big data streams, to extract insights and actionable results in real time. Application domains requiring data stream management include military, homeland security, sensor networks, financial applications, network management, web site performance tracking, real-time credit card fraud detection, etc. 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 to make the book an easy-to-read and enjoyable read. This book 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

Principles of Parallel Scientific Computing

Parallel Computing is Everywhere

Parallel Scientific Computing

Mastering Parallel Programming with R

Introduction to HPC with MPI for Data Science

Scientific Parallel Computing

Numerical algorithms, modern programming techniques, and parallel computing are often taught serially across different courses and different textbooks. The need to integrate concepts and tools usually comes only in employment or in research - after the courses are concluded - forcing the student to synthesise what is perceived to be three independent subfields into one. This book provides a seamless approach to stimulate the student simultaneously through the eyes of multiple disciplines, leading to enhanced understanding of scientific computing as a whole. The book includes both basic as well as advanced topics and places equal emphasis on the discretization of partial differential equations and on solvers. Some of the advanced topics include wavelets, high-order methods, non-symmetric systems, and parallelization of sparse systems. The material covered is suited to students from engineering, computer science, physics and mathematics.

A Tour of Data Science: Learn R and Python in Parallel covers the fundamentals of data science, including programming, statistics, optimization, and machine learning in a single short book. It does not cover everything, but rather, teaches the key concepts and topics in Data Science. It also covers two of the most popular programming languages used in Data Science, R and Python, in one source. Key features: Allows you to learn R and Python in parallel Cover statistics, programming, optimization and predictive modelling, and the popular data manipulation tools - data.table and pandas Provides a concise and accessible presentation Includes machine learning algorithms implemented from scratch, linear regression, lasso, ridge, logistic regression, gradient boosting trees, etc. Appealing to data scientists, statisticians, quantitative analysts, and others who want to learn programming with R and Python from a data science perspective.

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

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.

Parallel Algorithms in Computational Science and Engineering

Parallel Scientific Computing in C++ and MPI

Advances and Applications

Parallel Scientific Computing and Optimization

Parallel Programming

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 searches 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, Benchmarks, 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