Evaluation Of Hpc Applications On Cloud

This book provides a multidisciplinary view of smart infrastructure through a range of diverse introductory and advanced topics. The book features an array of subjects that include: smart cities and infrastructure, e-healthcare, emergency and disaster management, Internet of Vehicles, supply chain management, eGovernance, and high performance computing. The book is divided into five parts: Smart Transportation, Smart Healthcare, Miscellaneous Applications, Big Data and High Performance Computing, and Internet of Things (IoT). Contributions are from academics, researchers, and industry professionals around the world. Features a broad mix of topics related to smart infrastructure and smart applications, particularly high performance computing, big data, and artificial intelligence; Includes a strong emphasis on methodological aspects of infrastructure, technology and application development; Presents a substantial overview of research and development on key economic sectors including healthcare and transportation.
This book constitutes the refereed proceedings of the 15th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems, DAIS 2015, held as part of the 10th International Federated Conference on Distributed Computing Techniques, DisCoTec 2015, in Grenoble, France, in June 2015. The 14 full papers and 3 short papers presented in this volume were carefully reviewed and selected from 44 submissions. They deal with topics such as fault tolerance, privacy, resource management, social recommenders and cloud systems.
This book constitutes the proceedings of the 13th International Conference on Web Information Systems Engineering, WISE 2012, held in Paphos, Cyprus, in November 2012. The 44 full papers, 13 short papers, 9 demonstrations papers and 9 "challenge" papers were carefully reviewed and selected from 194 submissions. The papers cover various topics in the field of Web Information Systems Engineering.
Traditional computing concepts are maturing into a new generation of cloud computing systems with wide-spread global applications. However, even as these systems continue to expand, they are accompanied by overall performance degradation and wasted resources. Emerging Research in Cloud Distributed Computing Systems covers the latest innovations in resource management, control and monitoring applications, and security of cloud technology. Compiling and analyzing current trends, technological concepts, and future directions of computing systems, this publication is a timely resource for practicing engineers, technologists, researchers, and advanced students interested in the domain of cloud computing.
15th IFIP WG 6.1 International Conference, DAIS 2015, Held as Part of the 10th International Federated Conference on Distributed Computing Techniques, DisCoTec 2015, Grenoble, France, June 2-4, 2015, Proceedings
A Cross-layer Approach
Applications and Developments in Grid, Cloud, and High Performance Computing
Evaluation of Low-power Architectures in a Scientific Computing Environment
Cloud Computing and Big Data
Paravirtualization for HPC Systems
Evaluation and Optimization of Turnaround Time and Cost of HPC Applications on the Cloud
This book constitutes the refereed proceedings of the 16th International Conference on Economics of Grids, Clouds, Systems, and Services, GECON 2019, held in Leeds, UK, in September 2019. The 12 full papers and 10 short papers presented in this book were carefully reviewed and selected from 48 submissions. This GECON 2019 proceedings was structured in selected topics, namely: blockchain technology and smart contracts; cost-based computing allocation; resource, service and communication federations; economic assessment, business and pricing models; blockchain and network function virtualization technologies; economic models for cyber-physical systems, industry 4.0 and sustainable systems; resource management; and emerging ideas.
Modern HPC applications pose high demands on I/O performance and storage capability. The emerging non-volatile memory (NVM) techniques, such as Phase Change Memory and STT-RAM, offer low-latency, high bandwidth, and persistence for HPC applications. However, the existing I/O stack, including OS, high level library, I/O middleware, and applications, are designed and optimized based on an assumption of disk-based storage. To effectively use NVM, we must re-examine the existing I/O sub-system to properly integrate NVM into it. Using NVM as a fast storage, the previous assumption on the inferior performance of storage (e.g., hard drive) is not valid any more. The performance problem caused by slow storage may be mitigated; the existing mechanisms to narrow the performance gap between storage and CPU may be unnecessary and result in large overhead. Thus fully understanding of the impact of introducing NVM into the HPC software stack demands a thorough performance study. In this paper, we analyze and model the performance of I/O intensive HPC applications with NVM as a block device. We study the performance from three perspectives: (1) the impact of NVM on the performance of traditional page caches; (2) a performance comparison between MPI individual I/O and POSIX I/O; and (3) the impact of NVM on the performance of collective I/O. We reveal the diminishing effects of page caches, ignorable performance difference between MPI individual I/O and POSIX I/O, and performance disadvantage of collective I/O on NVM due to unnecessary data shuffling. We model the performance of MPI collective I/O and study the complex interaction between data shuffling, storage performance, and I/O access patterns. Extensive experiments have been conducted to verify our analysis.
Current advances in High Performance Computing (HPC) increasingly impact efficient software development workflows. Programmers for HPC applications need to consider trends such as increased core counts, multiple levels of parallelism, reduced memory per core, and I/O system challenges in order to derive well performing and highly scalable codes. At the same time, the increasing complexity adds further sources of program defects. While novel programming paradigms and advanced system libraries provide solutions for some of these challenges, appropriate supporting tools are indispensable. Such tools aid application developers in debugging, performance analysis, or code optimization and therefore make a major contribution to the development of robust and efficient parallel software. This book introduces a selection of the tools presented and discussed at the 7th International Parallel Tools Workshop, held in Dresden, Germany, September 3-4, 2013.
Proceedings of ICBDCC 2019
Performance Evaluation Models for Distributed Service Networks
ERCICA 2020
Cloud Computing for Scientific Research
Foundations for Smarter Cities and Societies
35th International Conference, ISC High Performance 2020, Frankfurt/Main, Germany, June 22-25, 2020, Proceedings
This book constitutes the refereed proceedings of the 36th International Conference on High Performance Computing, ISC High Performance 2021, held virtually in June/July 2021. The 24 full papers presented were carefully reviewed and selected from 74 submissions. The papers cover a broad range of topics such as architecture, networks, and storage; machine learning, AI, and emerging technologies; HPC algorithms and applications; performance modeling, evaluation, and analysis; and programming environments and systems software.
The popularity of Amazon's EC2 cloud platform has increased in commercial and scientific high-performance computing (HPC) applications domain in recent years. However, many HPC users consider dedicated high-performance clusters, typically found in large compute centers such as those in national laboratories, to be far superior to EC2 because of significant communication overhead of the latter. We find this view to be quite narrow and the proper metrics for comparing high-performance clusters to EC2 is turnaround time and cost. In this work, we first compare the HPC EC2 cluster to top-of-the-line HPC clusters based on turnaround time and total cost of execution. When measuring turnaround time, we include expected queue wait time on HPC clusters. Our results show that although as expected, standard HPC clusters are superior in raw performance, they suffer from potentially significant queue wait times. We show that EC2 clusters may produce better turnaround times due to typically lower wait queue times. To estimate cost, we developed a pricing model---relative to EC2's node-hour prices---to set node-hour prices for (currently free) clusters. We observe that the cost-effectiveness of running an application on a cluster depends on raw performance and application scalability. However, despite the potentially lower queue wait and turnaround times, the primary barrier to using clouds for many HPC users is the cost. Amazon EC2 provides a fixed-cost option (called on-demand) and a variable-cost, auction-based option (called the spot market). The spot market trades lower cost for potential interruptions that necessitate checkpointing; if the market price exceeds the bid price, a node is taken away from the user without warning. We explore techniques to maximize performance per dollar given a time constraint within which an application must complete. Specifically, we design and implement multiple techniques to reduce expected cost by exploiting redundancy in the EC2 spot market. We then design an adaptive algorithm that selects a scheduling algorithm and determines the bid price. We show that our adaptive algorithm executes programs up to 7x cheaper than using the on-demand market and up to 44% cheaper than the best non-redundant, spot-market algorithm. Finally, we extend our adaptive algorithm to exploit several opportunities for cost-savings on the EC2 spot market. First, we incorporate application scalability characteristics into our adaptive policy. We show that the adaptive algorithm informed with scalability characteristics of applications achieves up to 56% cost-savings compared to the expected cost for the base adaptive algorithm run at a fixed, user-defined scale. Second, we demonstrate potential for obtaining considerable free computationtime on the spot market enabled by its hour-boundary pricing model.
I introduce the cloud computing fundamentals, architecture of layers, and scientific services on the cloud firstly. Then, I introduce several typical commercial cloud computing platforms, such as Amazon Cloud Computing, Microsoft Azure, and Google Cloud Platform. Lastly, I discuss the scientific cloud computing based on these three commercial cloud computing platforms.
High Performance Computing (HPC) is used for running advanced application programs efficiently, reliably, and quickly. In earlier decades, performance analysis of HPC applications was evaluated based on speed, scalability of threads, memory hierarchy. Now, it is essential to consider the energy or the power consumed by the system while executing an application. In fact, the High Power Consumption (HPC) is one of biggest problems for the High Performance Computing (HPC) community and one of the major obstacles for exascale systems design. The new generations systems intend to achieve exaflop performances and will demand even more energy to processing and cooling. Nowadays, the growth of HPC systems is limited by energy issues Recently, many research centers have focused the attention on doing an automatic tuning of HPC applications which require a wide study of HPC applications in terms of power efficiency. In this context, this paper aims to propose the study of an oceanographic application, named OceanVar, that implements Domain Decomposition based 4D Variational model (DD-4DVar), one of the most commonly applications, going to evaluate not only the classic aspects of performance but also aspects related to power efficiency in different case of studies. These work were realized at Bsc (Barcelona Supercomputing Center), Spain within the Mont-Blanc project, performing the test first on HCA server with Intel technology and then on a mini-cluster Thunder with ARM technology. In this work of thesis it was initially explained the concept of assimilation date, the context in which it is developed, and a brief description of the mathematical model 4DVAR. After this problem's close examin was performed a porting from Matlab description of the problem of data-assimilation to its sequential version in C language. Secondly, after identifying the most onerous computational kernels in order of time, it has been developed a parallel version of the application with a parallel multiprocessor programming style, using the MPI (Message Passing Interface) protocol. The experiments results, in terms of performance, have shown that, in the case of running on HCA server, an Intel architecture, values of efficiency of the two most onerous functions obtained, growing the numb process, are approximately equal to 80%. In the case of running on ARM architecture, specifically on Thunder mini-cluster, instead, the trend obtained is labeled as "SuperLinear Speedup" and, in our case, it can be explained by a more efficient use of resources (cache memory access) compared with the sequential case. In the second part of this paper was presented an analysis of the some issues of this application that has impact in the energy efficiency. After a brief discussion about the energy consumption characteristics of the Thunder chip in technological landscape, th of a power consumption detector, the Yokogawa Power Meter, values of energy consumption of mini-cluster Thunder were evaluated in order to determine an overview on the power-to-solution of this application to use as the basic standard for successive analysis with other parallel styles. Finally, a comprehensive performance evaluation, targeted to estimate the goodness of MPI parallelization, is conducted using a suitable performance tool named Paraver, developed by BSC. Paraver is such a performance analysis and visualisation tool which can be used to analyse MPI, t mixed mode programmes and represents the key to perform a parallel profiling and to optimise the code for High Performance Computing. A set of graphical representation of these statistics make it easy for a developer to identify performance problems. Some of the problems that can be easily identified are load imbalanced decompositions, excessive communication overheads and poor average floating operations per second achieved. Paraver can also report statistics based on hardware counters, which are provided by the underlying hardware. This project aimed to use Pa configuration files to allow certain metrics to be analysed for this application. To explain in some way the performance trend obtained in the case of analysis on the mini-cluster Thunder, the tracks were extracted from various case of studies and the results achieved is what expected, that is a drastic drop of cache misses by the case ppn (process per node) = 1 to case ppn = 16. This in some way explains a more efficient use of cluster resources with an increase of the number of processes.
Performance Evaluation: Metrics, Models and Benchmarks
Performance Analysis, Modeling and Scaling of HPC Applications and Tools
Harnessing Performance Variability in Embedded and High-performance Many/Multi-core Platforms
ISPA 2006 Workshops : ISPA 2006 International Workshops, FHPCN, XHPC, S-GRACE, GridGIS, HPC-GTP, PDCE, ParDMCom, WOMP, ISDF, and UPWN, Sorrento, Italy, December 4-7, 2006 : Proceedings
13th TPC Technology Conference, TPCTC 2021, Copenhagen, Denmark, August 20, 2021, Revised Selected Papers
Service-Oriented Computing
This book is a compendium of the proceedings of the International Conference on Big-Data and Cloud Computing. The papers discuss the recent advances in the areas of big data analytics, data analytics in cloud, smart cities and grid, etc. This volume primarily focuses on the application of knowledge which promotes ideas for solving problems of the society through cutting-edge big-data technologies. The essays featured in this proceeding provide novel ideas that contribute for the growth of world class research and development. It will be useful to researchers in the area of advanced engineering sciences.
Cloud computing offers many advantages to researchers and engineers who need access to high performance computing facilities for solving particular compute-intensive and/or large-scale problems, but whose overall high performance computing (HPC) needs do not justify the acquisition and operation of dedicated HPC facilities. There are, however, a number of fundamental problems which must be addressed, such as the limitations imposed by accessibility, security and communication speed, before these advantages can be exploited to the full. This book presents 14 contributions selected from the International Research Workshop on Advanced High Performance Computing Systems, held in Cetraro, Italy, in June 2012. The papers are arranged in three chapters. Chapter 1 includes five papers on cloud infrastructures, while Chapter 2 discusses cloud applications. The third chapter in the book deals with big data, which is nothing new - large scientific organizations have been collecting large amounts of data for decades - but what is new is that the focus has now broadened to include sectors such as business analytics, financial analyses, Internet service providers, oil and gas, medicine, automotive and a host of others. This book will be of interest to all those whose work involves them with aspects of cloud computing and big data applications.
This book constitutes the refereed proceedings of the 20th International Conference on Parallel and Distributed Computing, Euro-Par 2014, held in Porto, Portugal, in August 2014. The 68 revised full papers presented were carefully reviewed and selected from 267 submissions. The papers are organized in 15 topical sections: support tools environments; performance prediction and evaluation; scheduling and load balancing; high-performance architectures and compilers; parallel and distributed data management; grid, cluster and cloud computing; green high performance computing; distributed systems and algorithms; parallel and distributed programming; parallel numerical algorithms; multicore and manycore programming; theory and algorithms for parallel computation; high performance networks and communication; high performance and scientific applications; and GPU and accelerator computing.
A major contributor to the deployment and operational costs of a large-scale high-performance computing (HPC) clusters is the memory system. In terms of system performance it is one of the most critical aspects of the system's design. However, next generation of HPC systems poses significant challenges for the main memory, and it is questionable whether current memory technologies will meet the required goals. In this thesis we focus on HPC performance aspects of the memory system design, covering memory bandwidth and latency. We start our study by evaluating and comparing three mainstream and five alternative HPC architectures, regarding memory bandwidth and latency aspects. Increasing diversity of HPC systems in the market causes their evaluation and comparison in terms of HPC features to become complex. There is as yet no well established methodology for a unified evaluation of HPC systems and workloads that quantifies the main performance bottlenecks. Our work provides a significant body of useful information and emphasizes four usually overlooked aspects of HPC systems' evaluation. Understanding the dominant performance bottlenecks of HPC applications is essential for designing a balanced HPC system. In our study, we execute a set of real HPC applications from diverse scientific fields, quantifying FLOPS performance and memory bandwidth congestion. We show that the results depend significantly on the number of execution processes, and argue for guidance on selecting the representative scale of the experiments. Also, we find that average measurements of performance metrics and bottlenecks can be highly misleading, and suggest reporting as the percentage of execution time in which applications use certain portions of maximum sustained values. Innovations in 3D-stacking technology enable DRAM devices with much higher bandwidths than traditional DIO6s. The first such products hit the market, and some of the publicity claims that they will break through the memory wall. We summarize our preliminary analysis and expectations of how such 3D-stacked DRAMs will affect the memory wall for a set of representative HPC applications. We conclude that although 3D-stacked DRAM is a major technological innovation, it is unlikely to break through the memory wall. Novel memory systems are typically explored by hardware simulators that are slow and often have a simplified or obsolete model of the CPU. We propose an analytical model that quantifies the impact of the main memory on application performance and system power and energy consumption, based on the memory system and application profiles. The model is evaluated on a mainstream platform, comprising various 18R3 memory configurations, and an alternative platform comprising 18R4 and 3D-stacked high-bandwidth memory. The evaluation results show that the model predictions are accurate, typically with only 2% difference from the values measured on actual hardware. Additionally, we compare the model performance estimation with simulation results, and our model shows significantly better accuracy over the simulator, while being faster by three orders of magnitude. Overall, we believe our study provides valuable insights on the importance of memory bandwidth and latency in HPC: their role in evaluation and comparison of HPC platforms, guidelines on measuring and presenting the related performance bottlenecks, and understanding and modeling of their performance, power and energy impact.
Co-Scheduling of HPC Applications
Proceedings of the 7th International Workshop on Parallel Tools for High Performance Computing, September 2013, ZIH, Dresden, Germany
SPEC International Performance Evaluation Workshop, SIPEW 2008, Darmstadt, Germany, June 27-28, 2008, Proceedings
High Performance Computing
A Holistic Method for Evaluating High Performance Computing Systems
Euro-Par 2014: Parallel Processing
This book constitutes the thoroughly refereed post-conference proceedings of the Third International Conference on High Performance Computing in Science and Engineering, HPCSE 2017, held in Karolinka, Czech Republic, in May 2017. The 15 papers presented in this volume were carefully reviewed and selected from 20 submissions. The conference provides an international forum for exchanging ideas among researchers involved in scientific and parallel computing, including theory and applications, as well as applied and computational mathematics. The focus of HPCSE 2017 was on models, algorithms, and software and convenient utilization of modern parallel and distributed computing architectures, as well as on large-scale applications.
This book describes the state-of-the art of industrial and academic research in the architectural design of heterogeneous, multi/many-core processors. The authors describe methods and tools to enable next-generation embedded and high-performance heterogeneous processors to confront cost-effectively the inevitable variations by providing Dependable-Performance: correct functionality and timing guarantees throughout the expected lifetime of a platform under thermal, power, and energy constraints. Various aspects of the reliability problem are discussed, at both the circuit and architecture level, the intelligent selection platforms, and systematic design methodologies. The authors demonstrate how new techniques have been applied in real case studies from different applications domain and report on results and conclusions of those experiments. Enables readers to develop performance-dependable heterogeneous multi/many-core architectures Describes system software designs that support high performance dependability requirements Discusses and analyzes low level methodologies to tradeoff conflicting metrics, i.e. power, performance, reliability and thermal management Includes new application design guidelines to improve perfor
"This book provides insight into the current trends and emerging issues by investigating grid and cloud evolution, workflow management, and the impact new computing systems have on the education fields as well as the industries"--Provided by publisher.
High-performance computing (HPC) has become an essential tool in the modern world. However, systems frequently run well below theoretical peak performance, with only 5% being reached in many cases. In addition, costly components often remain idle when not required for specific programs, as parts of the HPC systems are reserved and used exclusively for applications. A project was started in 2013, funded by the German Ministry of Education and Research (BMBF), to find ways of improving system utilization by compromising on dedicated reservations for HPC codes and applying co-scheduling of applications inste international discussion to find the best solutions to this HPC utilization issue, and a workshop on co-scheduling in HPC, open to international participants – the COSH workshop – was held for the first time at the European HiPEAC conference, in Prague, Czech Republic, in January 2016. This book presents extended versions of papers submitted to the workshop, reviewed for the second time to ensure scientific quality. It also includes an introduction to the main challenges of co-scheduling and a foreword by Arndt Bode, head of LRZ, one of Europe's leading computer centers, as well as a chapter corresponding to the invit recent extensions to their processors allow for better control of co-scheduling.
13th International Conference, Paphos, Cyprus, November 28-30, 2012, Proceedings
21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part III
Emerging Research in Cloud Distributed Computing Systems
Emerging Research in Computing, Information, Communication and Applications
Performance Evaluation of Darshan 3.0.0 on the Cray XC30
Web Information Systems Engineering - WISE 2012
Because of the growing popularity of parallel programming in multi-core/multiprocessor and multithreaded hardware, more and more applications are implemented in the well-written concurrent programming model. These programming models are MPI, OpenMP and Hybrid MPI/OpenMP. However, developing concurrent programs is extremely difficult. Concurrency introduces the possibility of errors that do not happen in traditional sequential programs, such as data race, deadlock and thread-safety issues. In addition, the performance issue of concurrent programs is another research area. This dissertation presents an integrated static and dynamic program analysis framework to address these concurrent issues in the OpenMP multithreaded application and hybrid OpenMP/MPI programming model. This dissertation also introduces the approach to reallocating the computing resources to improve the performance of MPI parallel programs in the container-based virtual cloud. First, we present the OpenMP Analysis Toolkit (OAT), which uses Satisfiability Modulo Theories (SMT) solver based symbolic analysis to detect data races and deadlocks in OpenMP applications. Our approach approximately simulates the real execution schedule of an OpenMP program through schedule permutation with partial order reduction to improve the analysis efficiency. We conducted experiments on real-world OpenMP benchmarks by comparing our OAT tool with two commercial dynamic analysis tools: Intel Thread Checker and Sun Thread Analyzer, and one commercial static analysis tool: Viva64 PVS Studio. The experiments show that our symbolic analysis approach is more accurate than static analysis and more efficient and scalable than dynamic analysis tools with less false positives and negatives. The second part of the dissertation proposes an approach by integrating static and dynamic program analyses to check thread-safety violations in hybrid MPI/OpenMP programs. We use an innovative method to transform the thread-safety violation problems in race conditions. In our approach, the static analysis identifies a list of MPI calls related to thread-safety violations, then replaces them with our own MPI wrappers, which access specific shared variables. The static analysis avoids instrumenting unrelated code, which significantly reduces runtime overhead. In the dynamic analysis, both happen-before and lockset-based race detection algorithms are used to check races on these aforementioned shared variables. By checking races, we can identify thread-safety violations according to their specifications. Our experimental evaluation over real-world applications shows that our approach is both accurate and efficient. Finally, the dissertation describes an approach that uses adaptive resource management enabled by container-based virtualization techniques to automatically tune performance of MPI programs in the cloud. Specifically, the containers running on physical hosts can dynamically allocate CPU resources to MPI processes according to the current program execution state and system resource status. High Performance Computing (HPC) in the cloud has great potential as an effective and convenient option for users to launch HPC applications. However, there are still many open problems to be solved in order for the cloud to be more amenable to HPC applications. In order to tune the performance of MPI applications during runtime, many traditional techniques try to balance the workloads by distributing datasets approximately equally to all computing nodes. However, the computing resource imbalance may still arise from data skew, and it is nontrivial to foresee such imbalances beforehand. The resource allocation among MPI processes are adjusted in two ways: the intra-host level, which dynamically adjusts resources within a host; and the inter-host level, which migrates containers together with MPI processes from one host to another host. We have implemented and evaluated our approach on the Amazon EC2 platform using real-world scientific benchmarks and applications, which demonstrates that the performance can be improved up to 31.1% (with an average of 15.6%) when comparing with the baseline of application runtime.
This book constitutes the proceedings of the 13th International Conference on Service-Oriented Computing, ICSOC 2015, held in Goa, India, in November 2015. The 23 full, 9 short, and 5 demo track papers presented in this volume were carefully reviewed and selected from 132 submissions. The research track papers are organized in topical sections named: internet of services/things; data services and cloud platform management; cloud services management; service composition; business process management; cloud services; QoS and trust; service composition.
This book constitutes the proceedings of the 24th International Conference on Parallel and Distributed Computing, Euro-Par 2018, held in Turin, Italy, in August 2018. The 57 full papers presented in this volume were carefully reviewed and selected from 194 submissions. They were organized in topical sections named: support tools and environments; performance and power modeling, prediction and evaluation; scheduling and load balancing; high performance architecutres and compilers; parallel and distributed data management and analytics; cluster and cloud computing; distributed systems and algorithms; parallel and distributed programming, interfaces, and languages; multicore and manycore methods and tools; theory and algorithms for parallel computation and networking; parallel numerical methods and applications; and accelerator computing for advanced applications.
The six-volume set LNCS 12742, 12743, 12744, 12745, 12746, and 12747 constitutes the proceedings of the 21st International Conference on Computational Science, ICCS 2021, held in Krakow, Poland, in June 2021.* The total of 260 full papers and 57 short papers presented in this book set were carefully reviewed and selected from 635 submissions. 48 full and 14 short papers were accepted to the main track from 156 submissions; 212 full and 43 short papers were accepted to the workshops/ thematic tracks from 479 submissions. The papers were organized in topical sections named: Part I: ICCS Main Track Part II: Advances in High-Performance Computational Earth Sciences: Applications and Frameworks; Applications of Computational Methods in Artificial Intelligence and Machine Learning; Artificial Intelligence and High-Performance Computing for Advanced Simulations; Biomedical and Bioinformatics Challenges for Computer Science Part III: Classifier Learning from Difficult Data; Computational Analysis of Complex Social Systems; Computational Collective Intelligence; Computational Health Part IV: Computational Methods for Emerging Problems in (dis-)Information Analysis; Computational Methods in Smart Agriculture; Computational Optimization, Modelling and Simulation; Computational Science in IoT and Smart Systems Part V: Computer Graphics, Image Processing and Artificial Intelligence; Data-Driven Computational Sciences; Machine Learning and Data Assimilation for Dynamical Systems; MeshFree Methods and Radial Basis Functions in Computational Sciences; Multiscale Modelling and Simulation Part VI: Quantum Computing Workshop; Simulations of Flow and Transport: Modeling, Algorithms and Computation; Smart Systems: Bringing Together Computer Vision, Sensor Networks and Machine Learning; Software Engineering for Computational Science; Solving Problems with Uncertainty; Teaching Computational Science; Uncertainty Quantification for Computational Models *The conference was held virtually.
Evaluation of the Actor Model for the Parallelization of Block-Structured Adaptive HPC Applications
Performance Evaluation and Benchmarking
Evaluation and Optimization of Turnaround Time and Cost of HPC Applications on the Cloud
Improving Reliability and Performance of High Performance Computing Applications
Supercomputing Frontiers
Performance-aware Energy Optimizations in Networks for HPC.
HPC (High Performance Computing) represents, together with theory and experiments, the third pillar of science. Through HPC, scientists can simulate phenomena otherwise impossible to study. The need of performing larger and more accurate simulations requires to HPC to improve every day. HPC is constantly looking for new computational platforms that can improve cost and power efficiency. The Mont-Blanc project is a EU funded research project that targets to study new hardware and software solutions that can improve efficiency of HPC systems. The vision of the project is to leverage the fast growing market of mobile devices to develop the next generation supercomputers. In this work we contribute to the objectives of the Mont-Blanc project by evaluating performance of production scientific applications on innovative low power architectures. In order to do so, we describe our experiences porting and evaluating sate of the art scientific applications on the Mont-Blanc prototype, the first HPC system built with commodity low power embedded technology. We then extend our study to compare off-the-shelves ARMv8 platforms. We finally discuss the most impacting issues encountered during the development of the Mont-Blanc prototype system.
Energy efficiency is an important challenge in the field of High Performance Computing (HPC). High energy requirements not only limit the potential to realize next-generation machines but are also an increasing part of the total cost of ownership of an HPC system. While at large HPC systems are becoming increasingly energy proportional in an effort to reduce energy costs, interconnect links stand out for their inefficiency. Commodity interconnect links remain ¿always-on¿, consuming full power even when no data is being transmitted. Although various techniques have been proposed towards energy- proportional interconnects, they are often too conservative or are not focused toward HPC. Aggressive techniques for interconnect energy savings are often not applied to HPC, in particular, because they may incur excessive performance overheads. Any energy-saving technique will only be adopted in HPC if there is no significant impact on performance, which is still the primary design objective. This thesis explores interconnect energy proportionality from a performance perspective. In this thesis, first a characterization of HPC applications is presented, making a case for the enormous potential for interconnect energy proportionality with HPC applications. Next, an HPC interconnect with on/off based links, modeled after the IEEE Energy Efficient Ethernet protocol, is evaluated. This evaluation while presenting a relationship between performance impact and energy over HPC applications also emphasizes the need for performance focused designs in energy efficient interconnects. Next, an adaptive mechanism, PerfBound, is presented that saves link energy subject to a bound on application performance overheads. Finally this evaluation structure is applied into an intermediate link power state, in addition to the traditional on and off states. Results of this study, over 15 production HPC applications show that, compared to current day always-on HPC interconnects, link energy can be reduced by unto 70%, while application performance overhead is bounded to only 1%.
It constitutes the refereed proceedings of the 4th Asian Supercomputing Conference, SCFA 2018, held in Singapore in March 2018. Supercomputing Frontiers will be rebranded as Supercomputing Frontiers Asia (SCFA), which serves as the technical programme for SCA18. The technical programme for SCA18 consists of four tracks: Application, Algorithms & Libraries Programming System Software Architecture, Network/Communications & Management Data, Storage & Visualisation The 20 papers presented in this volume were carefully reviewed nd selected from 60 submissions.
Virtualization has become increasingly popular for enabling full system isolation, load balancing, and hardware multiplexing. This wide-spread use is the result of novel techniques such as paravirtualization that make virtualization systems practical and efficient. Paravirtualizing systems export an interface that is slightly different from the underlying hardware but that significantly streamlines and simplifies the virtualization process. In this work, we investigate the efficacy of using paravirtualizing software for performance-critical HPC kernels and applications. Such systems are not currently employed in HPC environments due to their perceived overhead. However, virtualization systems offer tremendous potential for benefiting HPC systems by facilitating application isolation, portability, operating system customization, and program migration. We present a comprehensive performance evaluation of Xen, a low-overhead, Linux-based, virtual machine monitor (VMM), for paravirtualization of HPC cluster systems at Lawrence Livermore National Lab (LLNL). We consider four categories of micro-benchmarks from the HPC Challenge (HPCC) and LLNL ASCI Purple suites to evaluate a wide range of subsystem-specific behaviors. In addition, we employ macro-benchmarks and HPC application to evaluate overall performance in a real setting. We also employ statistically sound methods to compare the performance of a paravirtualized kernel against three popular Linux operating systems: RedHat Enterprise 4 (RHEL4) for build versions 2.6.9 and 2.6.12 and the LLNL CHAOS kernel, a specialized version of RHEL4. Our results indicate that Xen is very efficient and practical for HPC systems.
36th International Conference, ISC High Performance 2021, Virtual Event, June 24 – July 2, 2021, Proceedings
High Performance Computing in Science and Engineering
Smart Infrastructure and Applications

Application Performance Evaluation Using Deep Learning
Distributed Applications and Interoperable Systems
This book constitutes the refereed joint proceedings of ten international workshops held in conjunction with the 4th International Symposium on Parallel and Distributed Processing and Applications, ISPA 2006, held in Sorrento, Italy in December 2006. It contains 116 papers that contribute to enlarging the spectrum of the more general topics treated in the ISPA 2006 main conference.
This book constitutes the refereed proceedings of the SPEC International Performance Evaluation Workshop, SIPEW 2008, held in Darmstadt, Germany, in June 2008. The 17 revised full papers presented together with 3 keynote talks were carefully reviewed and selected out of 39 submissions for inclusion in the book. The papers are organized in topical sections on models for software performance engineering; benchmarks and workload characterization; Web services and service-oriented architectures; power and performance; and profiling, monitoring and optimization.
This book presents novel approaches to formulate, analyze, and solve problems in the area of distributed service networks, notably based on AI-related methods (parallel/cloud computing, declarative modeling, fuzzy methods). Distributed service networks are an important area of research and applications. The methods presented are meant to integrate both emerging and existing concepts and approaches for different types of production flows through synchronizations. An integration of logistics services (e.g., supply chains and projects portfolios), public and multimodal transport, traffic flow congestion management in ad hoc networks, design of high-performance cloud data centers, and milk-run distribution networks are shown as illustrations for the methods proposed. The book is of interest to researchers and practitioners in computer science, operations management, production control, and related fields.
This book constitutes the refereed proceedings of the 35th International Conference on High Performance Computing, ISC High Performance 2020, held in Frankfurt/Main, Germany, in June 2020.* The 27 revised full papers presented were carefully reviewed and selected from 87 submissions. The papers cover a broad range of topics such as architectures, networks & infrastructure; artificial intelligence and machine learning; data, storage & visualization; emerging technologies; HPC algorithms; HPC applications; performance modeling & measurement; programming models & systems software. *The conference was held virtually due to the COVID-19 pandemic. Chapters "Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) Streaming-Aggregation Hardware Design and Evaluation", "Solving Acoustic Boundary Integral Equations Using High Performance Tile Low-Rank LU Factorization", "Scaling Genomics Data Processing with Memory-Driven Computing to Accelerate Computational Biology", "Footprint-Aware Power Capping for Hybrid Memory Based Systems", and "Pattern-Aware Staging for Hybrid Memory Systems" are available open access under a Creative Commons Attribution 4.0 International License via link.springer.com.
Development of an Oceanographic Application in HPC.
Economics of Grids, Clouds, Systems, and Services
Workload Modeling for Computer Systems Performance Evaluation
Third International Conference, HPCSE 2017, Karolinka, Czech Republic, May 22–25, 2017, Revised Selected Papers
4th Asian Conference, SCFA 2018, Singapore, March 26-29, 2018, Proceedings
20th International Conference, Porto, Portugal, August 25-29, 2014, Proceedings
This book presents the proceedings of International Conference on Emerging Research in Computing, Information, Communication and Applications, ERCICA 2020. The conference provides an interdisciplinary forum for researchers, professional engineers and scientists, educators and technologists to discuss, debate and promote research and technology in the upcoming areas of computing, information, communication and their applications. The book discusses these emerging research areas, providing a valuable resource for researchers and practicing engineers alike.
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.

Abstract--Darshan is a lightweight I/O characterization tool used to gather and summarize salient I/O workload statistics from HPC applications. Darshan was designed to minimize any possible perturbations of an application's performance, leading it to be enabled by default on a number of production HPC systems. For each file accessed by a given application, Darshan records the count and types of I/O operations performed, histograms of access sizes, cumulative timers on the amount of time spent doing I/O, and other statistical data. This type of data has proved invaluable in understanding and improving the I/O performance of HPC applications. Darshan 3.0.0 is the new modularized version of the traditional Darshan library and file format, allowing users to easily add more in-depth I/O characterization data to Darshan logs. In this work we perform an empirical evaluation of Darshan 3.0.0 to ensure that it continues to meet performance expectations for broad deployment. In particular, we evaluate the imposed overhead on instrumented I/O operations, time taken to shut down and generate corresponding Darshan log files, and resultant log file sizes for different workloads. These performance results are compared to results of Darshan 2.3.0 on the Edison XC30 system at NERSC to determine whether the new version is lightweight enough to run full-time on production HPC systems. Our evaluation shows that Darshan has limited impact on application I/O performance and can fully generate a corresponding log file for most application workloads in under two seconds.

E cient use of supercomputers at DOE centers is vital for maximizing system throughput, mini- mizing energy costs and enabling science breakthroughs faster. This requires complementary e orts along several directions to optimize the performance of scienti c simulation codes and the under- lying runtimes and software stacks. This in turn requires providing scalable performance analysis tools and modeling techniques that can provide feedback to physicists and computer scientists developing the simulation codes and runtimes respectively. The PAMS project is using time allocations on supercomputers at ALCF, NERSC and OLCF to further the goals described above by performing research along the following fronts: 1. Scaling Study of HPC applications; 2. Evaluation of Programming Models; 3. Hardening of Performance Tools; 4. Performance Modeling of Irregular Codes; and 5. Statistical Analysis of Historical Performance Data. We are a team of computer and computational scientists funded by both DOE/NNSA and DOE/ ASCR programs such as ECRP, XStack (Traleika Glacier, PIPER), ExaOSR (ARGO), SDMAV II (MONA) and PSAAP II (XPACC). This allocation will enable us to study big data issues when analyzing performance on leadership computing class systems and to assist the HPC community in making the most e ective use of these resources.

16th International Conference, GECON 2019, Leeds, UK, September 17-19, 2019, Proceedings

24th International Conference on Parallel and Distributed Computing, Turin, Italy, August 27 - 31, 2018, Proceedings

Intelligence in Big Data Technologies—Beyond the Hype

Tools for High Performance Computing 2013

PERCU

Euro-Par 2018: Parallel Processing

This book constitutes the refereed post-conference proceedings of the 13th TPC Technology Conference on Performance Evaluation and Benchmarking, TPCTC 2021, held in August 2021. The 9 papers presented were carefully reviewed and selected from numerous submissions. The TPC encourages researchers and industry experts to present and debate novel ideas and methodologies in performance evaluation, measurement, and characterization.

A book for experts and practitioners, emphasizing the intuition and reasoning behind definitions and derivations related to evaluating computer systems performance.

Developing software for exascale systems will become even more challenging than for today's systems. Methods for evaluating the performance of applications and identifying potential weaknesses are essential for reaching optimal performance. Though the tools available today are not widely used, and generally require some expert knowledge. In recent years different deep learning techniques have enjoyed great success in various fields, and especially in image recognition. Though it is still to find its way in to the area of application performance evaluation. This work will take the first step towards introducing deep learning to the area of HPC performance evaluation, opening the door for others. Convolutional neural networks will be fed images of timeline views of HPC applications and will identify the intrinsic behavior of the application and return some principal performance metrics. The results show that deep learning techniques indeed can be utilized for evaluating the performance of parallel applications, with the main limitation for its success being the sizes of the data sets available. Furthermore a number of exciting directions for taking the next step utilizing deep learning techniques with performance evaluation are suggested.

Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing

Frontiers of High Performance Computing and Networking

13th International Conference, ICSOC 2015, Goa, India, November 16-19, 2015, Proceedings

Performance Evaluation and Modeling of HPC I/O on Non-volatile Memory

Memory Bandwidth and Latency in HPC: System Requirements and Performance Impact

Computational Science - ICCS 2021