Titles and Abstracts

Workshop on "Research for Petascale Computing in Science and Engineering"

April 20, 2006
NCSA Auditorium

Software Problems on the Road to Petascale Computing

Marc Snir, University of Illinois at Urbana-Champaign

The software model provided by large-scale parallel systems has not changed for over a decade. At the system level, individual system services at each node are complemented by a thin layer of cluster services for parallel joba scheduling, parallel I/O and a few other similar functions; at the application level, the message-passing model has become dominant. Arguments are made that this model can not carry us into petascale, as systems with hundreds of thousands of nodes become the norm, as failures become more frequent and as the complexity of application codes increase. On the other hand, the experience with systems such as Blue Gene seem to indicate that the current models can be stretched to petascale levels.

Compilers, Libraries, and Infrastructure Software for Petascale Computing

Wen-mei W. Hwu

With the scaling of microprocessor clock frequency and instruction-level parallelism coming to an end, the hardware path to petascale computing will likely be built upon multi-core microprocessor, application-specific processors, reconfigurable logic fabric, and distributed memory blocks. While systems of such mixture have been employed in the embedded computing domain for years, major roadblocks exist in building a high-performance computing system out of these hardware ingredients. One major difference is that one must provide appropriate programming interface to a high-performance system so that application developers do not have to deal with the intricate details of these hardware ingredients. The purpose of this presentation is to help seed the discussions on our approach to petascale computing with a critical look at the trajectory of each hardware ingredient and how one might build a high-perfomance system from these ingredients while maintaining a productive programming interface. Much of the talk will draw on the information from our potential industry partners as well as UIUC research in compilers and tools for

Scientific Data Mining – Challenges at the Petascale

Chandrika Kamath, Lawrence Livermore National Laboratory

The data from scientific and engineering simulations, observations, and experiments is now being measured in terabytes and will soon reach the petabyte regime. Due to the size of the data and the complexity of the problems, much of the data is never even looked at once it has been obtained. This is of course disconcerting to scientists who wonder about the science still undiscovered in the data. In this talk, I will describe my experiences in the Sapphire scientific data mining project at Lawrence Livermore National Laboratory (http://www.llnl.gov/casc/sapphire). Using example problems from domains including fluid mixing, molecular dynamics, astronomy, remote sensing, and experimental physics, I will discuss some of the challenges we have encountered in mining datasets ranging from a few megabytes to a hundred terabytes. I will then discuss what the future holds for data mining as we move to petascale computing.

Petascale Applications and Site Requirements – How Hard Can It Be?

Rob Pennington, University of Illinois at Urbana-Champaign

The two common views on the future of petascale systems are that these systems may be scaled up and evolved versions of currently extant versions of machines or they may be radically new systems that extend into technical domains that are only now being explored with enthusiasm. Each of these approaches has technical support but as these technologies are being considered, it is critical to evaluate them by two basic criteria: The first criteria is which research areas can provide a well defined set of requirements to help define the envelope for the early petascale systems? The second question is what requirements will these first petascale systems place on those who have to support them?

Adaptive Algorithms and Software for Partial Differential Equations

Phillip P. Colella, Lawrence Berkeley National Laboratory

Adaptive algorithms for solving partial differential equations, such as adaptive grid methods (on structured and unstructured grids) and particle methods, present some of the most significant challenges to parallel scalability, particularly in the case where they are being used to solve problems involving multiple physical processes. In this talk, we will discuss the barriers to scaling adaptive algorithms to the petascale, as well as new algorithmic and software ideas that might overcome those barriers

If You're Not scared, You're Not Thinking Big Enough

Robert J. Harrison, Oak Ridge National Laboratory & University of Tennessee

Multiteraflops computers are now commonplace and the era of petaflops computing is rushing towards us. Such resources are scientifically very exciting and in chemistry promise breakthroughs in such nationally important topics as catalysis, biochemical processes, and clean energy production. However, this rush of technology is leaving chemistry and possibly many other disciplines far behind in its wake. Few chemical applications can presently make effective use of the 100-1000s of processors in current teraflop computers, and the anticipated 100,000+ processors of petaflops computers will require an entire new generation of codes. I will discuss issues involved in the development scalable codes and algorithms, with emphasis on chemical applications and with reference to NWChem and MADNESS.

Biosketches

Workshop on "Research for Petascale Computing in Science and Engineering"

April 20, 2006
NCSA Auditorium

Phillip Colella received his A.B. (1974), M.A. (1976) and Ph.D. (1979) degrees from the University of California at Berkeley, all in applied mathematics. He is currently a Senior Staff Scientist and Group Leader for the Applied Numerical Algorithms Group in the Computing Sciences Directorate at the Lawrence Berkeley National Laboratory. His research has been in the area of high-resolution and adaptive methods for partial differential equations. He has also applied numerical methods in a variety of scientific and engineering fields, including shock dynamics, low-Mach number and incompressible flows, combustion, porous media flows, and astrophysical flows. Honors and awards include the IEEE Sidney Fernbach Award for high-performance computing in 1998, the SIAM/ACM prize (with John Bell) for computational science and engineering in 2003, and election to the US National Academy of Sciences in 2004.

Robert J. Harrison holds a joint appointment between Oak Ridge National Laboratory (ORNL) and the chemistry department of the University of Tennessee, Knoxville. He has been at ORNL for two years and is leader of the Computational Chemical Sciences Group in the Computer Science and Mathematics Division. Dr. Harrison has over 75 publications in peer-reviewed journals in the areas of theoretical and computational chemistry, and high-performance computing. His undergraduate (1981) and post-graduate (1984) degrees were obtained at Cambridge University, England. Subsequently, he worked as a postdoctoral research fellow at the Quantum Theory Project, University of Florida, and the Daresbury Laboratory, England, before joining the staff of the theoretical chemistry group at Argonne National Laboratory in 1988. In 1992, he moved to the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory, conducting research in theoretical chemistry and leading the development of NWChem, a computational chemistry code for massively parallel computers. In August 2002, he started the joint faculty appointment with UT/ORNL. In addition to his SciDAC research into efficient and accurate calculations on large systems, he has been pursuing applications in molecular electronics and chemistry at the nanoscale. In 1999, the NWChem team received an R&D Magazine R&D100 award, and, in 2002, Dr. Harrison received the IEEE Computer Society Sydney Fernbach award.

Wen-mei W. Hwu holds the Sanders-AMD Endowed Chair in the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. His research interests are in the area of architecture, implementation, and software for high performance computer systems. He is the director of the IMPACT research group (www.crhc.uiuc.edu/Impact). For his contributions in research and teaching, he received the 1993 Eta Kappa Nu Outstanding Young Electrical Engineer Award, the 1994 Xerox Award for Faculty Research, the 1994 University Scholar Award of the University of Illinois, the 1997 Eta Kappa Nu Holmes MacDonald Outstanding Teaching Award, the 1998 ACM SigArch Maurice Wilkes Award, the 1999 ACM Grace Murray Hopper Award, and the 2001 Tau Beta Pi Daniel C. Drucker Eminent Faculty Award. He is a fellow of IEEE and ACM. Hwu serves on the Executive Committee of the MARCO/DARPA C2S2 (www.c2s2.org) and GSRC (www.gigascale.org) Focus Research Centers. He co-leads the GSRC Soft Systems Theme with Kurt Kuetzer. He also serves on the GELATO Strategy Council (www.gelato.org). Dr. Hwu received his Ph.D. degree in Computer Science from the University of California, Berkeley.

Chandrika Kamath is a computer scientist in the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory, where she leads the Sapphire Project in scientific data mining. Dr. Kamath earned her Ph.D. in 1986 and her M.S. in 1984, both in Computer Science from the University of Illinois at Urbana-Champaign. She received her B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Bombay, in 1981.

Dr. Kamath's current research interests are in the area of scientific data mining. Her interests span all aspects of the field, including image and video processing, feature extraction, dimension reduction, and classification and clustering algorithms. She is especially interested in the practical application of these techniques to scientific data sets that result from experiments, observations, and simulations. She holds five patents in data mining. Her activities in the field of scientific data mining include the organization of the series of Workshops on Mining Scientific Data and the week-long short program at the Institute for Pure and Applied Mathematics on Mathematical Challenges in Scientific Data Mining. She is also the co-editor of the book *Data Mining for Scientific and Engineering Applications.*

Prior to joining LLNL in 1997, Dr. Kamath was a Consulting Software Engineer at Digital Equipment Corporation (DEC), developing high performance mathematical software. Her early career interests included the solution of sparse linear systems of equations and high-performance computing. Since January 1998, Chandrika has been the project lead and an individual contributor for Sapphire, a project in scientific data mining.

Rob Pennington is Chief Technology Officer and leads the Innovative Systems Laboratory at the National Center for Supercomputing Laboratory. He drives the NCSA's efforts to realize the future of computing by researching and deploying innovative systems that decrease the cost and/or extend the range of computational science and engineering. Dr. Pennington also acts as NCSA's TeraGrid site lead and the TeraGrid cluster lead. He is a member of the Open Source Cluster group and the IEEE Task Force for Cluster Computing. In addition, he was recognized for his service to society by inclusion in the Computerworld Honors Collection, a library of cutting-edge information technology projects that is distributed to museums, libraries, and institutions around the world.

Marc Snir is Michael Faiman and Saburo Muroga Professor of Computer Science and Head of the Computer Science Department at the University of Illinois at Urbana-Champaign since Fall 2001. He is pursuing research on architectures and programming environments for High-Performance Computing systems. Before joining UIUC he spent fifteen years at IBM Research. He led the research that resulted in the IBM SP scalable parallel system product - the leading IBM supercomputer product of the last decade – and initiated the IBM Blue Gene project. He was member of the IBM Academy of Technology and received an IBM Corporate award and two IBM Outstanding Technical awards.

Marc Snir received a Ph.D. in Mathematics from the Hebrew University of Jerusalem in 1979, worked at NYU on the NYU Ultracomputer project in 1980-1982 (a leading parallel computer research project at the time), and worked at the Hebrew University of Jerusalem in 1982-1986, before joining IBM. Prof. Snir was a major contributor to the design of the Message Passing Interface that is now almost universally used to code scalable parallel applications. He has co-authored two books and has published over 100 papers. His research has spanned over the years theoretical computer science, the design of parallel computers, parallel algorithms and tools, libraries and programming environments forparallel processing.

Marc Snir is an ACM Fellow and an IEEE Fellow. He was co-chair of a recent NRC study on the Future of Supercomputing, member of an NRC panel on NASA's PRT program, and member of a panel on High Performance Computing for the National Security Community. He is on the Board of the Computing Research Association.