International Journal of Distributed Systems and Technologies, 1(2), 1-22, April-June 2010

On the Path to Exascale

Ken Alvin, Sandia National Laboratories, USA

Brian Barrett, Sandia National Laboratories, USA

Ron Brightwell, Sandia National Laboratories, USA

Sudip Dosanjh, Sandia National Laboratories, USA

Al Geist, Oak Ridge National Laboratory, USA

Scott Hemmert, Sandia National Laboratories, USA

Michael Heroux, Sandia National Laboratories, USA

Doug Kothe, Oak Ridge National Laboratory, USA

Richard Murphy, Sandia National Laboratories, USA

Jeff Nichols, Oak Ridge National Laboratory, USA

Ron Oldfield, Sandia National Laboratories, USA

Arun Rodrigues, Sandia National Laboratories, USA

Jeffrey S. Vetter, Oak Ridge National Laboratory, USA

ABSTRACT

There is considerable interest in achieving a 1000 fold increase in supercomputing power in the next decade, but the challenges are formidable. In this paper, the authors discuss some of the driving science and security applications that require Exascale computing (a million, trillion operations per second). Key architectural challenges include power, memory, interconnection networks and resilience. The paper summarizes ongoing research aimed at overcoming these hurdles. Topics of interest are architecture aware and scalable algorithms, system simulation, 3D integration, new approaches to system-directed resilience and new benchmarks. Although significant progress is being made, a broader international program is needed.

Keywords: *Architectural Features, Architecture Types, Computer Architecture, Computer Size, Networking Technology, Operating Systems, Processor Architecture, Supercomputers*

INTRODUCTION

In 1997 Intel's ASCI Red broke the Teraflops barrier, achieving over 1 trillion floating point operations per second (Heermann, 1998). Last year both IBM's Roadrunner and Cray's Jaguar system surpassed 1 PetaFLOPS or 1,000 TeraFLOPS (Feldman, 2008). The next factor of 1000 improvement in supercomputing performance (Exascale) will be even more chal-

DOI: 10.4018/jdst.2010040101

lenging. The primary driver for the architectural change underway is that clock speeds are increasingly constrained by power and cooling limits. All of the major chip manufactures are moving to multi-core architectures, which results in the addition of hierarchical parallelism to supercomputers. Oak Ridge National Laboratory's Cray Jaguar system has 18,688 nodes, each comprised of two quad core AMD Opterons. Roadrunner is a one of a kind supercomputer composed of 6480 dual core AMD Opterons, each connected to two IBM PowerXCell 8i processors, which are similar to the processor used in Sony's Playstation 3 (Kahle, 2005). Both of these systems demonstrate that the transition to multi-core is already a key design challenge for supercomputer architectures.

The memory wall is defined as the mismatch between CPU and memory performance (latency, chip I/O capabilities, etc.), and will continue to plague processor design. Today's applications are primarily limited by data movement, represented by the statement "FLOPS are free" (Shiva, 2005). Power budgets continue to increase, which will result in power requirements in excess of 100 MW if existing Petascale design methodologies are used for an Exascale system. It is unlikely that such a system would be built unless power demands can be decreased significantly. Additionally, resilience will limit the availability of such systems. The current method for recovering from faults is check-pointing – at various times during a calculation a restart file is written to disk. Given weak scaling, I/O bandwidth requirements scale with the memory size and inversely with the mean time between interrupts (which decreases with increasing parts counts). Since Exascale may require millions of nodes, the mean time between interrupts will decrease and check-pointing will become an impractical mechanism for fault recovery.

These challenges require new approaches to applications, algorithms, system software, and computer architecture, which has been noted by the numerous workshops and reports devoted to the problem (Kogge et al., 2008; Simon et al., 2007). The remainder of the paper discusses science and security applications that require Exascale computing, architectural challenges, and ongoing work aimed at overcoming some of these obstacles.

APPLICATIONS

Scientific computing is essential to the advancement of numerous fields of study. Never before have we been able to accurately anticipate, analyze, and plan for complex events that have not occurred—from the operation of a reactor running at 100 million degrees to future changes in climate. Combined with the more traditional approaches of theory and experiment, scientific computing provides a profound tool for insight as we look at complex systems containing billions of components.

Science

Key areas of scientific research, including materials science, Earth science, energy assurance, fundamental science, biology and medicine, engineering design, and security can benefit from continued growth in high performance computing (to Exascale and beyond). Table 1 summarizes scientific opportunities that can be enabled by Exascale computing, key application areas, and the goals and associated benefits.

The system and application-wide advances required to reach Exascale are not inevitable, and require a fundamental rethinking across all aspects of High Performance Computing (HPC).

Material Science

Materials science drivers, objectives, and impacts that are enabled by Exascale leadership platforms have been identified in Table 2 (Department of Energy, 2007).

Earth Science

Earth science and climate change research will focus on two principal activities in the decade ahead:

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/path-exascale/42973

Related Content

A Synchronized Test Control Execution Model of Distributed Systems
Salma Azzouzi, Sara Hsainiand My El Hassan Charaf (2020). *International Journal of Grid and High Performance Computing (pp. 1-17).*
www.irma-international.org/article/a-synchronized-test-control-execution-model-of-distributed-systems/240602

Unified Data Access/Query over Integrated Data-views for Decision Making in Geographic Information Systems
Ahmet Sayar, Geoffrey C. Foxand Marlon E. Pierce (2009). *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations (pp. 276-298).*
www.irma-international.org/chapter/unified-data-access-query-over/19349

sl-LSTM: A Bi-Directional LSTM With Stochastic Gradient Descent Optimization for Sequence Labeling Tasks in Big Data
Nancy Victorand Daphne Lopez (2020). *International Journal of Grid and High Performance Computing (pp. 1-16).*
www.irma-international.org/article/sl-lstm/257221

Resource Co-Allocation in Grid Computing Environments
Marco A.S. Nettoand Rajkumar Buyya (2010). *Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications (pp. 476-494).*
www.irma-international.org/chapter/resource-allocation-grid-computing-environments/40814

OCEDS: Optimal Cost-Effective Data Storage in Cloud Data Centers
Arunambika T.and Senthil Vadivu P. (2021). *International Journal of Distributed Systems and Technologies (pp. 48-63).*
www.irma-international.org/article/oceds/284433