Workshop Overview

Workshop on Exascale Operating Systems and Runtime Software

Dates: October 4-5, Washington, DC.

Location: The facilities of the American Geophysical Union (http://www.agu.org/)

Registration:

Please register online prior to the workshop. Only registered participants will be allowed to attend.

If you have registration questions or need assistance, please contact Julie Webber at Julie.webber@orise.orau.gov or Jeannie Robinson J.robinson@orise.orau.gov.

The agenda for the meeting is still being edited. There is a Agenda that can be used for planning.

Workshop Goal:

This workshop is for the CS community, organized by a group of experts from the DOE community. The goal of this workshop is to explore fundamental computer science research directions in operating systems and runtime (OS/R) software, engaging the research community on the Exascale OS and runtime software roadmap.

During the workshop we will briefly review key challenges that have been previously identified. The workshop will focus on understanding potentially disruptive solutions and research directions that could overcome the identified challenges. We will identify and discuss revolutionary approaches for Exascale OS and runtime systems, articulating research priorities. We will also discuss an integration plan that involves the research community and vendor developers. The workshop discussions will enable a DOE roadmap for the research and development of Exascale platforms OS and runtime software, including prioritized areas of investment, timelines, and scale of investment.

Challenges to be addressed:

The gigahertz race in computing is over. Because of power constraints resulting from current leakage, individual processing units in CPUs have stopped getting significantly faster across subsequent generations. Future HPC nodes will have hundreds or thousands of cores. Extreme-scale systems could harness up to a billion parallel threads of control. In addition, the drive to reduce power consumption will also move chips toward threshold voltages, increasing the frequency of single-bit upsets and potentially decreasing reliability. Another power-driven change to chip architecture will be designs to reduce data movement, which can dominate energy consumption across the system. Complex memory hierarchies and memory coherence domains will be introduced to move data closer to arithmetic functional units. There are three key changes: increased parallelism, reduced reliability, and deeper memory hierarchies, which require explicit runtime power management of hardware resources that will revolutionize both the low-level operating system and runtime. Existing node-level operating systems are not capable of handling these enormous changes. New designs, methods, and interfaces will be required.

For applications, the need to hide latency across deep memory hierarchies and the dynamic management of faults and power will make bulk synchronous programming models, which assume equal sized work items will execute in equal time, too inefficient for extreme scale. Instead, dynamic, asynchronous programming models will hide latency and respond to the changing system. Likewise, the messaging layer will break free from the legacy of bulk synchronous data movement and become more parallel, supporting asynchronous data delivery to/from thousands of lightweight threads. The core operating system, messaging layers, and lightweight threads must be designed hand-in-hand to efficiently support these new, dynamic and adaptive programming models.

Moreover, responding to faults and managing power requires a broad perspective, outside the traditional view of a local-node OS/R. Extreme-scale systems must manage power and bandwidth globally, across the entire system. With power a first-class resource, a new set of global-view interfaces and OS/R functions are required. The file system, compute partitions, and interconnect fabric must all collect real-time data and respond quickly to control messages. Further, many global-view system software components, such as resource managers, are already used in large-scale systems. However, the interactions of these components with each other and the node operating system are typically managed in an ad hoc fashion. Wiring the node OS/R to global-view system software will require new ideas and directions.

Finally, extreme-scale OS/R software must be designed and prototyped along with an integrated plan for deployment and support. Three key stakeholders must be involved. For vendors building and integrating extreme-scale platforms, the OS/R is a core component. There must be a clear business path, based on open source and coordinated with the DOE-funded research, for advanced prototypes to be shared with the community and integrated with products if the technology provides key new functionality and capability. The application community is another key partner in the design and deployment of the OS/R. There must be a clear path, based on the principles of co-design, to the definition (and continued evolution) of the APIs that are available to application developers. Finally, the computer centers that deploy extreme-scale systems have unique needs. Local cybersecurity policies, patching and upgrading routines, and job monitoring and fault responses can all impact the design of extreme-scale operating system and runtime software.

During the first nine months of 2012, DOE is leading a series of discussions in order to build a roadmap for the research and development of extreme-scale OS/R software. The series of discussions has involved the stakeholders identified earlier (vendors, application developers, and facility managers) and academic OS/R researchers who will bring new perspectives. These discussions will inform the roadmap and identify promising research strategies for OS and runtime systems software. The Workshop on Exascale Operating Systems and Runtime Software will further explore key research challenges and engage the research community on the software roadmap.

We will survey the state-of-the-art in OS and runtime systems for high performance computing platforms. We will discuss the relevant challenges posed by massive parallelism, energy efficiency and active power management, heterogeneous architectures, whole-system management, fault tolerance, and embedded interconnects.

We anticipate technically provocative statements that instigate a lot of discussion, new thinking and revolutionary new approaches to the problems.

Specifically, in the two days of this event we plan to:

1. Survey the state-of-the-art in OS and runtime systems
2. Engage in in-depth technical discussions concerning energy and power management, OS/R resilience, lightweight messaging and thread runtime layers, and heterogeneous platforms.
3. Discuss the architectural and application trends that need to be considered in the co-design of OS and runtime systems
4. Brainstorm out-of-the-box alternatives and techniques for Exascale
5. Identify the areas of R&D in this arena that require pointed and/or joint efforts, both medium and long term
6. Identify where investment is needed to meet DOE mission requirements, and outline a roadmap for this research area of Exascale computing

A report that leverages the workshop discussions and presents a final version of the exascale operating system and runtime software challenges and roadmap will be delivered in October 2012.

Attendance

There is very limited seating space at the workshop location. Open registration requests closed on September 21. People from the waiting list will be notified by September 28th with final registration information if there is space for their attendance.