

Scalable Systems Software for Terascale Computer Centers

www.scidac.org/ScalableSystems

Coordinator: Al Geist ORNL

Participating Organizations:

DOE Labs – ORNL, ANL, Ames, LBNL, PNNL, SNL, LANL

NSF Supercomputer Centers – NCSA, PSC, SDSC

Vendors – IBM, Cray, Unlimited Scale, Intel, SGI, HP

Summary

The nation's premiere scientific computing centers are facing a crisis where they are having to rewrite all their home-grown systems software to scale to the multi-teraflops systems that are being installed in their centers. The goal of the Scalable Systems Software project is to fundamentally change the way future high-end systems software is developed to make it more cost effective and robust. The research involves two efforts: Collectively getting the DOE centers, NSF centers, and industry to agree on standardized interfaces between system components. Secondly, producing a compliant, fully integrated suite of systems software that can be used across all the terascale computer centers for the cost effective management and utilization of their computational resources.

System administrators and managers of terascale computer centers are facing a crisis. The nation's premiere scientific computing centers all use incompatible, ad hoc sets of systems tools (See Figure 1) and these tools were not designed to scale to the multi-teraflop systems that are being installed in these centers today. One solution would be for each computer center to take their home-grown software and rewrite it to be scalable. But this would incur a tremendous duplication of effort and delay the availability of terascale computers for scientific discovery.

The purpose of the Scalable Systems Software project is to provide a much more timely and cost effective solution by pulling together representatives from the major computer centers and industry and collectively defining standardized interfaces between system components. At the same time this group will produce a fully integrated suite of systems software and tools that can be used by the nation's largest scientific computing centers.



Figure 1. Systems software areas that are being standardized, integrated, and made scalable to promote scientific discovery.

The scalable systems software suite is being designed to support computers that scale to very large physical sizes without requiring that the number of support staff scale along with the machine. But this research goes beyond just creating a collection of separate scalable components. By defining a software architecture and

interfaces between system components, the Scalable Systems Software research is creating an interoperable framework for the components. This makes it much easier and cost effective for supercomputer centers to adapt, update, and maintain the components in order to keep up with new hardware and software. Publicly documented interfaces are a requirement because it is unlikely that any package or vendor can provide the flexibility to meet the needs of every site. A well-defined interface allows a site to replace or customize individual components as needed. Defining the interfaces between components across the entire system software architecture provides an integrating force between the system components as a whole and improves the long-term usability and manageability of terascale systems at supercomputer centers across the country.

The standardization of the systems interfaces is being done using a process similar to that used to successfully define the message passing standard. It is an open forum of university, lab, and industry representatives who meet regularly to propose and vote on pieces of the standard.

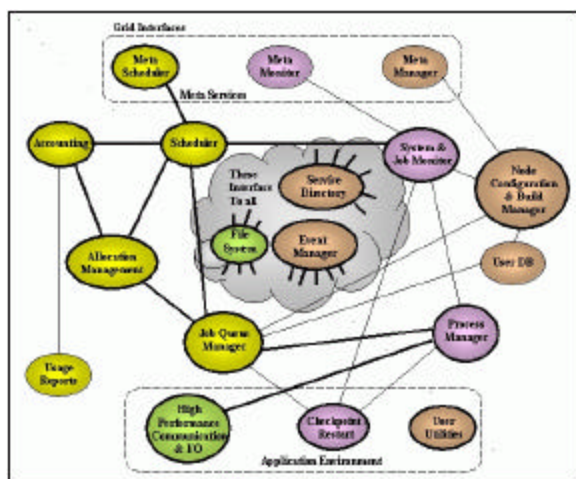


Figure 2. System components presently under development and their interfaces. Dark lines represent working interfaces.

Figure 2 represents the significant progress to date on producing scalable components and defining standardized interfaces between them. The bold lines represent working interfaces. The light lines represent interfaces in progress. The colors of the components just represent which of the four multi-lab working groups inside the project is responsible for it.

The research has developed a Service Directory component, which allows components to find each other and determine what interface they understand, an Event Manager that keeps track of the entire integrated suite, and software to provide communication service between components as well as a flexible authentication scheme to provide security to the overall system. The research has produced working prototypes of scalable scheduler, allocation manager, job manager, system monitor, checkpoint, and process manager components.

Impact: The Scalable Systems Software project is a catalyst for fundamentally changing the way future high-end systems software is developed and distributed. It will reduce facility management costs by: reducing the need to support home-grown software, making higher quality systems tools available, and being able to get new machines up and running faster and keep them running. The project will also facilitate more effective use of machines by scientific applications by providing scalable job launch, standardized job monitoring and management software, and allocation tools for the cost effective management and utilization of terascale computational resources.

For further information on this subject contact:

Al Geist, Project Coordinator

Oak Ridge National Laboratory

Phone: 865-574-3153

gst@ornl.gov

See also: www.scidac.org/ScalableSystems