

The Grid

The term *Grid* refers to a new infrastructure that builds on today's Internet and Web to enable and exploit large-scale sharing of resources within distributed, often loosely coordinated groups what are sometimes termed *virtual organizations*. Grids provide scalable, secure, and reliable mechanisms for discovering and negotiating access to remote resources including, but certainly not restricted to, clusters. The availability of Grid infrastructure, when combined with ubiquitous Internet connectivity and ever faster networks, enables entirely new, often communication-, compute-, and data-intensive applications.

Grid concepts and technologies emerged first within the scientific computing community, as a means of pooling computers, streaming large amounts of data from databases or instruments to remote computers, linking sensors with each other and with computers and archives, and connecting people, computing, and storage in collaborative environments that avoid a need for travel. Grid technologies are being deployed in a wide range of "e-science" projects and provide foundational elements for national and international "cyberinfrastructures" (a term coined by the U.S. National Science Foundation). Interest within industry is growing as companies realize that the needs of advanced e-business have much in common with those of e-science. Both require systems that span multiple institutions and that execute reliably, delivering consistent performance despite heterogeneous hardware, software, and policies.

The central motivation for Grid computing is that in a world where communication is close to free, we are not restricted, when solving problems, to local resources. For example, I can run interesting computer programs (game, scientific simulation, or business logic) remotely, rather than installing them locally. When analyzing data, I can have the remote code access relevant datasets directly. I can repeat a computation hundreds of times, on different datasets, by calling upon the collective computing power of my company or research collaboration or by purchasing cycles from a cycle provider. And I can review output with remote colleagues in rich collaborative environments.

While high-speed networks are often necessary for scenarios such as these, they are far from sufficient. Remote resources are typically owned by others, exist within different administrative domains, run different software, and are subject to different security and access control policies. These issues characterize a Grid

and historically have made distributed computing difficult. Grid technologies overcome these obstacles by providing standard and uniform mechanisms for such critical tasks as creating and managing services on remote computers, for supporting "single sign on" to distributed resources, for transferring large datasets at high speeds, and for forming large distributed virtual communities and maintaining information about the existence, state, and usage policies of community resources.

Standard mechanisms are critical if we are to avoid balkanization of infrastructure caused by different application developers rolling their own architectures. A decade of research, development, experimentation, and standardization has produced considerable consensus on Grid architecture principles and technologies, with in particular the community based, open source Globus ToolkitÆ (GT) being used by most major Grid projects and seeing significant industrial adoption. In addition, emerging Open Grid Services Architecture (OGSA) standards firmly align Grid computing with broad industry initiatives in service-oriented architecture and Web services. Grid-related standards are being defined within various standards bodies, including the Global Grid Forum, which has emerged as a significant force for standards setting and community development, with close to 1000 people from more than 400 organizations regularly attending its meetings. The NSF Middleware Initiative (NMI) also plays an important integrating role, providing essential support for the productization, packaging, dissemination, and support of open source Grid technologies such as GT.

The NMI-sponsored GRIDS Center has created a searchable database called the Grid Projects and Deployments System (GPDS, at <http://www.gpds.org>), with over 100 examples of the Grid in action. Visitors may search or browse by region, sponsor or user disciplines, and submit their own examples by completing a simple form. The GPDS is a growing resource for anyone seeking information about how and where Grid tools like GT, Condor-G and others are being used. Projects from 22 countries on four continents include sponsored research and commercial applications.

As the GPDS demonstrates, many scientific disciplines are using the Grid to conceive new avenues of research that weren't possible until recently. While the technology has had its earliest impact in fields (e.g., physics) where data and other resources have traditionally been shared freely, the Grid is causing other, less open communities (e.g., medical imaging) to reconsider long-standing norms that have mitigated against broad

collaboration. The same benefits can occur within companies that might not have a culture of sharing, whether internally or externally.

Grid computing is a compelling example of how sustained exponential technology evolution can have revolutionary impacts on the practice of computing. With high-speed networks becoming ubiquitous, and e-business and e-science concepts achieving broad adoption, we can expect Grid technologies and applications to become a major part of the computing landscape.

Ian Foster is the Associate Division Director of the Mathematics and Computer Science Division Argonne National Laboratory and Professor of Computer Science University of Chicago. He can be reached at foster@mcs.anl.gov His web page is <http://www.mcs.anl.gov/~foster>.