



The Grid Grows Up

Fred Douglis
IBM T.J. Watson Research Center

Ian Foster
*Argonne National Laboratory
 and University of Chicago*

The term “the Grid” was coined in the mid-1990s to denote a (then) proposed distributed computing infrastructure for advanced science and engineering. Researchers have since made much progress in constructing such an infrastructure and extending and applying it to a broad range of computing problems. As a result, “grid” has entered the computer science vocabulary to denote middleware infrastructure, tools, and applications concerned with integrating geographically distributed computational resources. The term’s use can create confusion due to the obvious overlap with established fields. However, its popularity also bears witness to the emergence of a vibrant community of researchers and practitioners whose concerns build on, but reach beyond, those of traditional networking and distributed computing.

Researchers first developed grid concepts and technologies to enable resource sharing within scientific collaborations — first in early gigabit test beds, and then on increasingly larger scales. At the root of these collaborations was the need for participants to share not only data sets but also software, computational resources, and even specialized instruments such as

telescopes and microscopes. More generally, scientists needed technologies to support coordinated resource sharing and problem solving in dynamic, multi-institutional collaborations. Similar requirements for sharing resources between cross-organizational collaborations arise within commercial environments, including enterprise application integration, on-demand service provisioning, data center federation, and business-to-business partner collaboration over the Internet. Commercial adoption of grid technologies is thus accelerating along a trajectory similar to the World Wide Web, which began as a technology for scientific collaboration but was adopted for e-business.

The Grid’s success to date owes much to the relatively early emergence of clean architectural principles, de facto standard software (in particular, the Globus Toolkit; www.globus.org), aggressive early adopters with challenging application problems, and a vibrant international community of researchers, developers, and users. This combination of factors created a solid experience base that has more recently driven the Global Grid Forum’s (GGF; www.ggf.org) efforts to define the Open Grid Services Architecture (OGSA), which now forms the basis

To Learn More

We have not included an "introduction to the Grid" article in this issue of *IC*, as that topic is covered in detail elsewhere. Instead, we refer you to two recent articles that introduce Grid computing¹ and the Open Grid Services Architecture,² respectively, as well as two additional articles that provide other perspectives on the surrounding issues.^{3,4} For those seeking more detail, we recommend the second edition of the seminal *The Grid*⁵ and a recent collection of technical papers, entitled *Grid Computing: Making the Global Infrastructure a Reality*.⁶ A recent NSF report also provides a compelling description of the scientific case for Grid infrastructure and tools.⁷

Other sources of information include

- the annual IEEE International Sym-

posium on High Performance Distributed Computing (www.hpdc.org), which is the principal academic conference in the field;

- a forthcoming special issue of the *Journal of Parallel and Distributed Computing* (www.academicpress.com/jpdc);
- the new *Journal of Grid Computing*; (www.kluweronline.com/issn/1570-7873);
- the Global Grid Forum (www.ggf.org) Web site and triannual meetings;
- GlobusWorld (www.globusworld.org), a user group and tutorial meeting for the open-source Globus Toolkit; and
- the *GridToday* online newsletter (www.gridtoday.com).

References

1. I. Foster, "The Grid: Computing without Bounds,"

Scientific Am., vol. 288, no. 4, 2003, pp. 78–85.

2. I. Foster et al., "Grid Services for Distributed Systems Integration," *Computer*, vol. 35, no. 6, June 2002, pp. 37–46.
3. C. Catlett and L. Smarr, "Metacomputing," *Comm. ACM*, vol. 35, no. 6, 1992, pp. 44–52.
4. A.S. Grimshaw and W.A. Wulf, "The Legion Vision of a Worldwide Virtual Computer," *Comm. ACM*, vol. 40, no. 1, 1997, pp. 39–45.
5. I. Foster and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*, 2nd ed., Morgan-Kaufmann, 2004.
6. F. Berman, G. Fox, and T. Hey, eds., *Grid Computing: Making the Global Infrastructure a Reality*, John Wiley & Sons, 2003.
7. *Revolutionizing Science and Engineering Through Cyberinfrastructure*, tech. report, US Nat'l Science Foundation, Blue Ribbon Advisory Panel on Cyberinfrastructure 2003; www.communitytechnology.org/nsf_ci_report.

of both open-source and commercial grid products. Building on Web services principles and technologies, the core OGSA *Grid service* specification defines standard interfaces and behaviors that address critical issues in distributed system integration and management, including the creation, lifetime management, introspection, and grouping of services. Other specifications under development address such issues as security, registry, policy, data access and integration, service management, and workflow.

We cannot hope to do justice to all aspects of this broad topic in a single theme issue, but the four articles that follow provide a good introduction to some of the concerns that drive current work on Grid deployment, applications, and research.

The Articles

In the first article, "Building a Production Grid in Scandinavia," Eerola and colleagues provide a detailed account of the design, deployment, and application of a scientific grid infrastructure within the Nordic countries. The goal of this work is to let a multinational community of high-energy physicists federate computing and storage resources across institutions to support various computation- and data-intensive tasks. The article describes the Globus-Toolkit-based NorduGrid architecture, the software elements developed (or configured) specifically for the NorduGrid, and the

authors' application experiences with it.

In "Service-Centric Globally Distributed Computing," Graupner and colleagues introduce some concepts that could prove useful when applying grid technologies within commercial computing infrastructures. In particular, the authors address the question of how to confederate multiple compute-data centers ("utility data centers") to increase flexibility and utilization for business workloads. They introduce the virtual server abstraction as a means of virtualizing computational resources, and they describe algorithms for managing the mapping of physical resources to virtual servers.

Leff, Rayfeld, and Dias address similar issues in the third article, "Service-Level Agreements and Commercial Grids," but they work from the perspective of the negotiation process that lets a resource consumer establish commitments from a resource provider. The authors describe how to use the Web-Service-Level Agreement notation to express user requirements and how to use dynamic offload mechanisms to deal with excessive load.

As the Grid has evolved, it has moved from merely distributed computation or data to an entire application infrastructure. In the final article, "Prototyping the Workspaces of the Future," Stevens, Papka, and Disz describe a geographically distributed collaborative environment underpinned by video for creating persistent virtual venues. The

article includes history, future directions, and a set of detailed examples of environments in which the Access Grid is being applied. Because this article presents a system that has been in wide use for a few years, and describes a blueprint for the next-generation Access Grid, it provides a useful perspective on the Grid's evolution.

The Future of Grid Computing

Although these articles provide just a sampling of the work under way in the Grid community, they suggest the breadth of the modalities that are being addressed (storage, data, computers, services, and collaboration), the wide range of participants (application scientists, industrial researchers, and academic computer scientists), and the varying degrees of maturity in the ideas and technologies.

The articles also suggest areas in which future deployment, application, research, and development will likely occur. For many application communities, the next few years will focus on achieving robust and sophisticated operational deployments to support increasingly ambitious application scenarios. Some communities are already scrambling to use grid technologies to handle overwhelming quantities of data; others are still considering how to reengineer their work modes to exploit large-scale resource sharing.

Within industry, we see tremendous effort and progress on the commercial application of grid technologies. Initial deployments from companies such as Avaki, Data Synapse, Fujitsu, Hitachi, HP, IBM, NEC, Oracle, Platform, Sun, and United Devices have focused primarily on the intra-enterprise federation of computation and data resources. Future products, some of which are already in field trials, will address issues such as service virtualization, distributed system management, and data integration.

For computer scientists, this growing interest in large-scale resource sharing and distributed system integration represents a tremendous opportunity to discover new technical challenges, reassess existing approaches and understandings, and deliver new capabilities. We can exploit as unique test beds for advanced grid technologies the extensive grid deployments that are now under way or planned in various scientific communities. Careful study of the challenges that arise within such systems already suggests many directions for new research, in areas ranging from trust, security, and policy to distributed system management.

While the participants in this work are still largely from "the Grid community," we see hopeful signs of increasing engagement by researchers in other fields, including networking, distributed systems, databases, algorithms, system management, security, human factors, artificial intelligence, and computer-assisted collaborative work. □

Fred Douglass is a research staff member at the IBM T.J. Watson Research Center. His research interests include storage systems and distributed computing. He received a PhD in computer science from the University of California, Berkeley. He is a senior member of IEEE and a member of *IEEE Internet Computing's* editorial board. Contact him at f.douglass@computer.org.

Ian Foster is associate director of the mathematics and computer science division at Argonne National Laboratory, and professor of computer science at the University of Chicago. His research interests include distributed computing and computational science. He received a PhD in computer science from Imperial College, London. A second edition of his book, *The Grid: Blueprint for a New Computing Infrastructure* (Morgan Kaufmann), will appear in November 2003. Contact him at foster@mcs.anl.gov.

Eleven good reasons why more than 100,000 computing professionals join the IEEE Computer Society

Transactions on

- **Computers**
- **Information Technology in Biomedicine**
- **Knowledge and Data Engineering**
- **Mobile Computing**
- **Multimedia**

- **Networking**
- **Parallel and Distributed Systems**
- **Pattern Analysis and Machine Intelligence**
- **Software Engineering**
- **Very Large Scale Integration Systems**
- **Visualization and Computer Graphics**



computer.org/publications/