G

Grid Computing for Social Science

Kenneth J. Turner
University of Stirling, UK

Paul Lambert
University of Stirling, UK

K. L. Tan
University of Stirling, UK

Vernon Gayle
University of Stirling, UK

Richard O. Sinnott
University of Glasgow, UK

Ken Prandy
University of Cardiff, UK

Erik Bihagen
Swedish Institute for Social Research, Sweden

Marco H. D. van Leeuwen
International Institute for Social History, The Netherlands

INTRODUCTION

Grid computing is named by analogy with the electrical power grid. Power stations are linked into a universal supply that delivers electricity on demand to consumers. Similarly, computational resources can be linked into a grid that delivers computing or data on demand to the user's desktop. The origins of grid computing lie in networked computing, distributed computing, and parallel computing. Grid computing coordinates distributed resources that are not subject to central control, using standard protocols and interfaces to meet the required levels of service (Foster, 2002).

International standards have been defined for open and vendor-independent networking. However, this simply allows exchange of information among computers that operate largely independently. Distributed computing was developed to use networks for effective sharing of resources: information, processing, and devices. This allows multiple computers to cooperate in manipulating information, no matter where it is stored or processed.

Parallel computing evolved separately as a means of allowing closely coupled processors to work on the same tasks. However, it is often more cost-effective to harness the power of many processors through distributed computing than to build specialized parallel processing systems.

Grid computing continues these trends by offering a standard architecture for large-scale distributed and parallel computing. Grid computing has proven particularly attractive where substantial computation is needed to analyze the rapidly growing amounts of data used by many collaborators across a wide range of disciplines. For example, it is envisaged that the Large Hadron Collider will regularly produce many petabytes of data (over 10^{15} bytes per second from small particle collisions).

In the context of the grid, a virtual organization is a set of individuals or institutions that agree to share distributed computing facilities. Such an organization is virtual because it crosses conventional organizational boundaries. A virtual organization is typically closed (the membership is specified) and may be dynamic (membership may change over its lifetime).

Grid computing provides facilities to support virtual organizations, notably authentication and security. This chapter describes a case study that demonstrates the value of virtual organizations for collaboration among social scientists.

BACKGROUND

Distributed Computing

A wide range of standards was developed for networked computing under the banner of open systems interconnection (ISO/IEC, 1994). Although this architecture remains an important framework for networking, both it and the proprietary solutions it superseded have largely faded from practical use. Instead, Internet-related standards have become the predominant means of linking networked systems.

Remote procedure calls (Birrell, 1984; ISO/IEC, 1988) were developed as a simple but effective way of linking computation in different systems. They allow procedures (methods) to be invoked in the same way, irrespective of whether they are on local or networked processors. The same approach is supported by modern programming languages such as C# or Java (Downing, 1998).

Initially, proprietary standards were developed for distributed computing, for example Microsoft distributed component object model (Abernethy,Morin, Chahin, & Morin, 1999). The need for vendor-independent standards led to the common object request broker architecture (Bolton, 2001). International standards were also defined for open distributed processing (ISO/IEC, 1995).

Distributed architectures have been defined in specialized areas such as telephony. For example the architecture of the intelligent network (ITU, 2000) supports distributed call processing, while the telecommunications intelligent network architecture (Dupuy, Nilsson, & Inoue, 1995) focuses on telephony services.

Grid Computing

Grid computing has grown out of web computing. The growth of the web as a means of providing information led to strong industrial interest in using a similar approach for distributed computing. Web services are supported by many commercial solutions such as BEA Weblogic (Mountjoy & Chugh, 2004), IBM Websphere (Francis, Herness, Knutson et al., 2002), Microsoft .NET (Platt, 2003), and Sun Open Net Environment (Mogha & Bhargava, 2002).

Web services allow networked applications to cooperate. Extensible markup language (XML) is a common foundation in web service standards for representation of structured data. The interface to a web service is defined by the web services description language (World Wide Web Consortium, 2001). Communication between web services is supported by the simple object access protocol (World Wide Web Consortium, 2003). Repositories of web services are created and interrogated through universal description, discovery and integration (OASIS, 2002).

Web services were adopted as the basis for grid computing, allowing grid applications to support distributed resource sharing and computation. Grid applications often support long-lived services. This requires stateful services that preserve state across different invocations. To meet this requirement, specialized solutions were initially developed for the grid. However, web services can also benefit from being stateful. The web Services Resource Framework (Graham, Marmakar, Mischinsky et al., 2006) was therefore developed as a collection of standards that allow Web services to expose the resources they use. This has allowed a much closer convergence between web services and grid services. However, the grid also has specialized standards in areas such as data access, file transfer, resource management and security. A grid service is said to run in a *container* that provides the infrastructure required.

Berman, Fox, and Hey (2003) provide a high-level overview of what the grid aims to achieve. Grid computing offers a number of distinctive advantages that draw on features for:

- *Virtual organizations* that transcend conventional boundaries, and may come together for only a particular task or to share resources
- *Grid portals* that provide simplified, user-oriented access to grid-enabled resources
- *Single sign-on* that requires authenticating once to use multiple resources such as data repositories or computational servers
- *Virtualized resources* that allow platform-independent and location-independent access to heterogeneous computing and data resources, or to specialized equipment