

1

An Introduction to the Grid

1.1 INTRODUCTION

The Grid concepts and technologies are all very new, first expressed by Foster and Kesselman in 1998 [1]. Before this, efforts to orchestrate wide-area distributed resources were known as metacomputing [2]. Even so, whichever date we use to identify when efforts in this area started, compared to general distributed computing, the Grid is a very new discipline and its exact focus and the core components that make up its infrastructure are still being investigated and have yet to be determined. Generally it can be said that the Grid has evolved from a carefully configured infrastructure that supported a limited number of grand challenge applications executing on high-performance hardware between a number of US national centres [3], to what we are aiming at today, which can be seen as a seamless and dynamic virtual environment. In this book we take a step-by-step approach to describe the middleware components that make up this virtual environment which is now called the Grid.

1.2 CHARACTERIZATION OF THE GRID

Before we go any further we need to somehow define and characterize what can be seen as a Grid infrastructure. To start with, let us think about the execution of a distributed application. Here

we usually visualize running such an application “on top” of a software layer called middleware that unifies the resources being used by the application into a single coherent virtual machine. To help understand this view of a distributed application and its accompanying middleware, consider Figure 1.1, which shows the hardware and software components that would be typically found on a PC-based cluster. This view then raises the question, what is the difference between a distributed system and the Grid? Obviously the Grid is a type of distributed system, but this does not really answer the question. So, perhaps we should try and establish “What is a Grid?”

In 1998, Ian Foster and Carl Kesselman provided an initial definition in their book *The Grid: Blueprint for a New Computing Infrastructure* [1]: “A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.” This particular definition stems from the earlier roots of the Grid, that of interconnecting high-performance facilities at various US laboratories and universities.

Since this early definition there have been a number of other attempts to define what a Grid is. For example, “A grid is a software framework providing layers of services to access and manage distributed hardware and software resources” [4] or a “widely

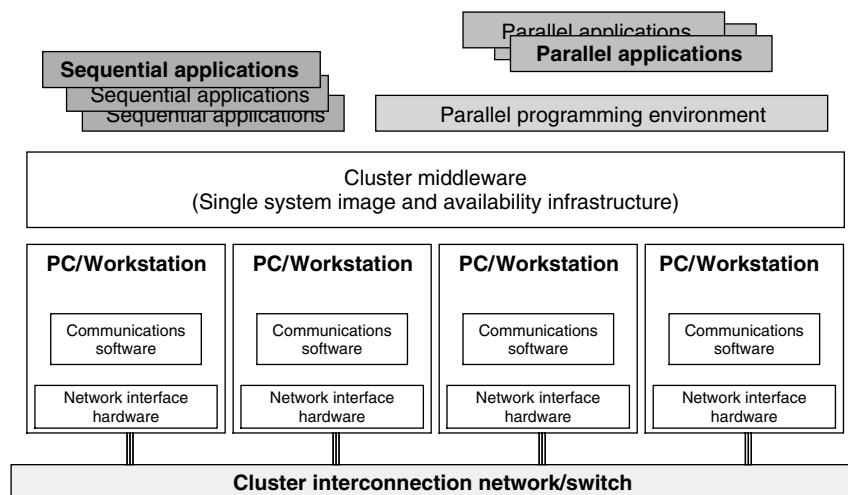


Figure 1.1 The hardware and software components of a typical cluster

distributed network of high-performance computers, stored data, instruments, and collaboration environments shared across institutional boundaries” [5]. In 2001, Foster, Kesselman and Tuecke refined their definition of a Grid to “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” [6]. This latest definition is the one most commonly used today to abstractly define a Grid.

Foster later produced a checklist [7] that could be used to help understand exactly what can be identified as a Grid system. He suggested that the checklist should have three parts to it. (The first part to check off is that there is coordinated resource sharing with no centralized point of control that the users reside within different administrative domains.) If this is not true, it is probably the case that this is not a Grid system. The second part to check off is the use of standard, open, general-purpose protocols and interfaces. If this is not the case it is unlikely that system components will be able to communicate or interoperate, and it is likely that we are dealing with an application-specific system, and not the Grid. The final part to check off is that of delivering non-trivial qualities of service. Here we are considering how the components that make up a Grid can be used in a coordinated way to deliver combined services, which are appreciably greater than the sum of the individual components. These services may be associated with throughput, response time, meantime between failure, security or many other facets.

From a commercial view point, IBM define a grid as “a standards-based application/resource sharing architecture that makes it possible for heterogeneous systems and applications to share, compute and storage resources transparently” [8].

So, overall, we can say that the Grid is about resource sharing; this includes computers, storage, sensors and networks. Sharing is obviously always conditional and based on factors like trust, resource-based policies, negotiation and how payment should be considered. The Grid also includes coordinated problem solving, which is beyond simple client-server paradigm, where we may be interested in combinations of distributed data analysis, computation and collaboration. The Grid also involves dynamic, multi-institutional Virtual Organizations (VOs), where these new communities overlay classical organization structures, and these virtual organizations may be large or small, static or dynamic. The LHC Computing Grid Project at CERN [9] is a classic example of where VOs are being used in anger.

1.3 GRID-RELATED STANDARDS BODIES

For Grid-related technologies, tools and utilities to be taken up widely by the community at large, it is vital that developers design their software to conform to the relevant standards. For the Grid community, the most important standards organizations are the Global Grid Forum (GGF) [10], which is the primary standards setting organization for the Grid, and OASIS [11], a not-for-profit consortium that drives the development, convergence and adoption of e-business standards, which is having an increasing influence on Grid standards. Other bodies that are involved with related standards efforts are the Distributed Management Task Force (DMTF) [12], here there are overlaps and on-going collaborative efforts with the management standards, the Common Information Model (CIM) [13] and the Web-Based Enterprise Management (WBEM) [14]. In addition, the World Wide Web Consortium (W3C) [15] is also active in setting Web services standards, particularly those that relate to XML.

The GGF produces four document types related to standards that are defined as:

- *Informational*: These are used to inform the community about a useful idea or set of ideas, for example GFD.7 (A Grid Monitoring Architecture), GFD.8 (A Simple Case Study of a Grid Performance System) and GFD.11 (Grid Scheduling Dictionary of Terms and Keywords). There are currently eighteen Informational documents from a range of working groups.
- *Experimental*: These are used to inform the community about a useful experiment, testbed or implementation of an idea or set of ideas, for example GFD.5 (Advanced Reservation API), GFD.21 (GridFTP Protocol Improvements) and GFD.24 (GSS-API Extensions). There are currently three Experimental documents.
- *Community practice*: These are to inform the community of common practice or process, with the objective to influence the community, for example GFD.1 (GGF Document Series), GFD.3 (GGF Management) and GFD.16 (GGF Certificate Policy Model). There are currently four Common Practice documents.
- *Recommendations*: These are used to document a specification, analogous to an Internet Standards track document, for example GFD.15 (Open Grid Services Infrastructure), GFD.20 (GridFTP:

Protocol Extensions to FTP for the Grid) and GFD.23 (A Hierarchy of Network Performance Characteristics for Grid Applications and Services). There are currently four Recommendation documents.

1.4 THE ARCHITECTURE OF THE GRID

Perhaps the most important standard that has emerged recently is the Open Grid Services Architecture (OGSA), which was developed by the GGF. OGSA is an Informational specification that aims to define a common, standard and open architecture for Grid-based applications. The goal of OGSA is to standardize almost all the services that a grid application may use, for example job and resource management services, communications and security. OGSA specifies a Service-Oriented Architecture (SOA) for the Grid that realizes a model of a computing system as a set of distributed computing patterns realized using Web services as the underlying technology. Basically, the OGSA standard defines service interfaces and identifies the protocols for invoking these services.

OGSA was first announced at GGF4 in February 2002. In March 2004, at GGF10, it was declared as the GGF's flagship architecture. The OGSA document, first released at GGF11 in June 2004, explains the OGSA Working Group's current thinking on the required capabilities and was released in order to stimulate further discussion. Instantiations of OGSA depend on emerging specifications (e.g. WS-RF and WS-Notification). Currently the OGSA document does not contain sufficient information to develop an actual implementation of an OSGA-based system. A comprehensive analysis of OGSA was undertaken by Gannon *et al.*, and is well worth reading [16].

There are many standards involved in building a service-oriented Grid architecture, which form the basic building blocks that allow applications execute service requests. The Web services-based standards and specifications include:

- Program-to-program interaction (SOAP, WSDL and UDDI);
- Data sharing (eXtensible Markup Language – XML);
- Messaging (SOAP and WS-Addressing);
- Reliable messaging (WS-ReliableMessaging);

- Managing workload (WS-Management);
- Transaction-handling (WS-Coordination and WS-AtomicTransaction);
- Managing resources (WS-RF or Web Services Resource Framework);
- Establishing security (WS-Security, WS-SecureConversation, WS-Trust and WS-Federation);
- Handling metadata (WSDL, UDDI and WS-Policy);
- Building and integrating Web Services architecture over a Grid (see OGSA);
- Overlaying business process flow (Business Process Execution Language for Web Services – BPEL4WS);
- Triggering process flow events (WS-Notification).

As the aforementioned list indicates, developing a solid and concrete instantiation of OGSA is currently difficult as there is a moving target – as the choice of which standard or specification will emerge and/or become popular is unknown. This is causing the Grid community a dilemma as to exactly what route to use to develop their middleware. For example, WS-GAF [17] and WS-I [18] are being mooted as possible alternative routes to WS-RF [19].

Later in this book (Chapters 2 and 3), we describe in depth what is briefly outlined here in Sections 1.2–1.4.

1.5 REFERENCES

- [1] Ian Foster and Carl Kesselman (eds), *The Grid: Blueprint for a New Computing Infrastructure*, 1st edition, Morgan Kaufmann Publishers, San Francisco, USA (1 November 1998), ISBN: 1558604758.
- [2] Smarr, L. and Catlett, C., Metacomputing, *Communication of the ACM*, 35, 1992, pp. 44–52, ISSN: 0001-0782.
- [3] De Roure, D., Baker, M.A., Jennings, N. and Shadbolt, N., The Evolution of the Grid, in *Grid Computing: Making the Global Infrastructure a Reality*, Fran Berman, Anthony J.G. Hey and Geoffrey Fox (eds), pp. 65–100, John Wiley & Sons, Chichester, England (8 April 2003), ISBN: 0470853190.
- [4] CCA, <http://www.extreme.indiana.edu/ccat/glossary.html>.
- [5] IPG, <http://www.ipg.nasa.gov/ipgflat/aboutipg/glossary.html>.
- [6] Foster, I., Kesselman, C. and Tuecke, S., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of Supercomputer Applications*, 15(3), 2001.
- [7] Grid Checklist, <http://www.gridtoday.com/02/0722/100136.html>.

-
- [8] IBM Grid Computing, http://www-1.ibm.com/grid/grid_literature.shtml.
 - [9] LCG, <http://lcg.web.cern.ch/LCG/>.
 - [10] GGF, <http://www.ggf.org>.
 - [11] OASIS, <http://www.oasis-open.org>.
 - [12] DMTF, <http://www.dmtf.org>.
 - [13] CIM, <http://www.dmtf.org/standards/cim>.
 - [14] WBEM, <http://www.dmtf.org/standards/wbem>.
 - [15] W3C, <http://www.w3.org>.
 - [16] Gannon, D., Chiu, K., Govindaraju, M. and Slominski, A., A Revised Analysis of the Open Grid Services Infrastructure, *Journal of Computing and Informatics*, 21, 2002, 321–332, http://www.extreme.indiana.edu/~aslom/papers/ogsa_analysis4.pdf.
 - [17] WS-GAF, <http://www.neresc.ac.uk/ws-gaf>.
 - [18] WS-I, <http://www.ws-i.org>.
 - [19] WS-RF, <http://www.globus.org/wsrp>.

