Chapter 8.10
Leveraging the Power of the Grid with Opal

Sriram Krishnan
University of California San Diego, USA

Luca Clementi
University of California San Diego, USA

Zhaohui Ding
Jilin University, China

Wilfred Li
University of California San Diego, USA

ABSTRACT

Grid systems provide mechanisms for single sign-on, and uniform APIs for job submission and data transfer, in order to allow the coupling of distributed resources in a seamless manner. However, new users face a daunting barrier of entry due to the high cost of deployment and maintenance. They are often required to learn complex concepts relative to Grid infrastructures (credential management, scheduling systems, data staging, etc). To most scientific users, running their applications with minimal changes and yet getting results faster is highly desirable, without having to know much about how the resources are used. Hence, a higher level of abstraction must be provided for the underlying infrastructure to be used effectively. For this purpose, we have developed the Opal toolkit for exposing applications on Grid resources as simple Web services. Opal provides a basic set of Application Programming Interfaces (APIs) that allows users to execute their deployed applications, query job status, and retrieve results. Opal also provides a mechanism to define command-line arguments and automatically generates user interfaces for the Web services dynamically. In addition, Opal services can be hooked up to a Metascheduler such as CSF4 to leverage a distributed set of resources, and accessed via a multitude of interfaces such as Web browsers, rich desktop environments, workflow tools, and command-line clients.

DOI: 10.4018/978-1-4666-0879-5.ch8.10

INTRODUCTION

In 1992, the term "Metacomputing" was coined by Smarr and Catlett to describe a connected network computing environment (Smarr, Catlett, 1992), which would eventually enable everyone to obtain information on demand "all from their own desktop workstations." Metacomputing has now been popularized as Grid computing with the emergence of the Globus toolkit (Foster, Kesselman, 1997) and other related software, with the goal of providing dynamic accessibility to computing resources like an electric power grid. With the release of the Atkins report in 2003 (Atkins et al, 2003), the term "cyberinfrastructure" is generally used to refer to the national and international network of computers, storage, software and human resources dedicated to support the advancement of science, engineering and medicine. Various countries have made major investments in the development of the Grid middleware, and software stacks that bridge the network and computing resources to the domain specific applications and users (Foster, Kesselman, 1999).

Since 1997, US funding agencies such as the National Science Foundation (NSF) and the Department of Energy (DOE) have funded the development of cyberinfrastructures through a series of initiatives including, but not limited to:

- The National Partnership for Advanced Computational Infrastructure (NPACI: http://www.npaci.edu/), National Middleware Initiatives (NMI), Software Development for Cyberinfrastructure (SDCI), and the Science Discovery through Advanced Computing (SciDAC) for software;
- The OptIputer (A Powerful Distributed Cyberinfrastructure to Support Data-Intensive Scientific Research and Collaboration: http://www.optiputer.net) and the Global Ring Network for Advanced Application Development (GLORIAD: http://www.gloriad.org) for networking;
- The TeraGrid (http://www.teragrid.org), the Open Science Grid (OSG: http://www.opensciencegrid.org), and the Petascale Computing Environment for high throughput and high performance computing.

A number of international Grid activities have also sprung up, such as UK e-Science Programme (Hey, Trefethen, 2005), The Enabling Grids for E-science (EGEE: http://public.eu-egee.org) from the European Union, and the Pacific Rim Grid Applications and Middleware Assembly (PRAGMA) (Arzberger, Papadopoulos, 2006) supported by NSF and member institutes in the Asian Pacific region. In the public health and biomedicine sector, the National Center of Research Resources (NCRR) at the National Institutes of Health (NIH) has funded the development of the Biomedical Informatics Research Network (BIRN) (Grethe et al, 2005) for data sharing, and the National Cancer Institute (NCI) has supported the development of the Cancer Biomedical Informatics Grid (caBIG) for Grid services for biomedical data models and integration (Oster, 2007).

The impact of cyberinfrastructures on e-science is most felt through the collaborative research it enables on a scale never imagined before. One of the key components is the ability to form "virtual organizations," where researchers and professionals alike are able to share data, applications, and visualization environment transparently and seamlessly with the necessary authentication and authorization (Siebenlist, Magaratnam, Welch, Neuman, 2004). For example, the BIRN project enables the biomedical researchers to share neuroscience imaging datasets, analysis tools, as well as access to TeraGrid resources through its portal environment. The OptIportals allow researchers to visualize large datasets over high speed optical networks. Similarly, the OSG supports high energy physics research, as well as computational biology virtual organizations (Stevens, 2006).

Related Content