

Chapter 5.8

Hands-On Experience in Building Institutional Grid Infrastructure

Xiaoyu Yang

University of Cambridge, UK

Gen-Tao Chiang

University of Cambridge, UK

ABSTRACT

It will become increasingly popular that scientists in research institutes will make use of Grid computing resources for running computer simulations and managing data. Although there are some production Grids available, it is often the case that many organizations and research projects need to build their own Grids. However, building Grid infrastructure is not a trivial job as it involves sharing and managing heterogeneous computing and data resources across different organizations, and involves installing many specific software packages and various middleware. This can be quite complicated and time-consuming. Building a Grid infrastructure also requires good knowledge and understanding of distributed computing, parallel computing and Grid technologies. Apart from building physical Grid, how to build a user infrastructure that can facilitate the use of and easy access to these physical infrastructures is also a challenging task. In this chapter, the authors summarize some hands-on experience in building an institutional Grid infrastructure. They describe knowledge and experience obtained in the installations of Condor pools, PBS clusters, Globus Toolkit, and SRB (Storage Resource Broker) for computing Grid and data Grid. The authors also propose to use a User-Centered Design (UCD) approach to develop a Grid user infrastructure which can facilitate the use of the Grid to improve the usability.

DOI: 10.4018/978-1-4666-0879-5.ch5.8

INTRODUCTION

“Grid computing is an infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources” (Foster,2001). However, this description still does not give a concrete definition to Grid and people use the term “Grid” with different meanings.

For instance, the most widely used “Grid” refers to computing Grids. This type of Grid refers to the sharing of computer resources, such as High Performance Computing (HPC), High Throughput Computing (HTC), or Condor-like desktop Grids, over the Internet. It aims ultimately to turn the global network of computers into one vast resource of computing power. LHC Computing Grid (LCG), now called Enabling Grid for e-Science (EGEE) (<http://www.eu-egee.org>), is a typical computing Grid project. It integrates more than 60,000 processors all over the world for processing petabytes of data generated by the Large Hadron Collider (LHC) at CERN (Lamanna, 2004).

The second interpretation of “Grid” is a data Grid. This type of Grid aims to allow data discovery and data sharing easily between collaborators. One example is the NERC Data Grid (NDG) (<http://ndg.badc.rl.ac.uk>). It creates a data Grid through which data from several NERC projects is accessible. Another example is the Biomedical Informatics Research Network (BIRN) (<http://www.nbirn.net/>). It is a biomedical science collaboration project within the USA. It consists of several projects on neuron-imaging of human neurological disorders and associated animal models.

The third interpretation of “Grid” is a collaborative Grid. This type of Grid tries to improve the communication between collaborators by using multiple participants video conferencing system and related applications. The Access Grid (AG) is an example. It provides such video conferencing along with other tools for sharing applications and data within a Virtual Organization (VO). A VO is a group of individuals or institutions who share

the computing and other resources of a “Grid” for a common goal (Clery, 2006).

However, none of these single interpretations reflects the whole picture of the “Grid”. An ideal Grid infrastructure should provide the functions of the most important parts from three types of Grids: Computing Grids, Data Grids, and Collaborative Grids. Since there are several collaborative tools and they are relatively easy in terms of deployment, this chapter will focus on building up computing and data Grid.

It will become increasingly popular that scientists in research institutes will make use of Grid computing resources for running computer simulations and managing data, this involves building a Grid infrastructure. Although there are some existing production Grids, such as National Grid Services (NGS) (<http://www.grid-support.org>) in UK, EGEE in EU, or Open Science Grid (OSG) (<http://www.opensciencegrid.org>) in the USA, it is still often the case that many organizations and research projects need to build their own Grids due to various reasons. For example, there are some security and data policy issues which do not allow either computing or data resources to be shared with others, that a test Grid environment is required for testing configurations and running applications before migrating to production sites, or building research project-specific Grid, etc. Apart from the physical Grid, a user-level infrastructure to facilitate the use of physical Grid also needs to be built.

In this chapter, we summarize some hands-on experience in building small-scale Grid infrastructure. As Grid infrastructure usually involves computing Grid and data Grid, we describe knowledge and experience obtained in installing Condor pools, PBS clusters, Globus Toolkit and SRB. We also introduce some job submission tools, and Grid monitoring and management tools. Furthermore, tools and technologies which can be employed in User-Centered Design (UCD) approach to develop a user-level infrastructure to facilitate the use of Grid are reviewed, and a case study is discussed.

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/hands-experience-building-institutional-grid/64534

Related Content

Automatic Dynamic Data Structures Recognition to Support the Migration of Applications to the Cloud

Beniamino Di Martino and Antonio Esposito (2015). *International Journal of Grid and High Performance Computing* (pp. 1-22).

www.irma-international.org/article/automatic-dynamic-data-structures-recognition-to-support-the-migration-of-applications-to-the-cloud/141297

Scalability of Mobile Ad Hoc Networks

Dan Grigoras, Daniel C. Doolan and Sabin Tabirca (2010). *Handbook of Research on Scalable Computing Technologies* (pp. 705-717).

www.irma-international.org/chapter/scalability-mobile-hoc-networks/36430

Adaptive Mobile Architecture with Utility Computing

Robin Singh Bhadoria and Chandrakant Patil (2016). *Emerging Research Surrounding Power Consumption and Performance Issues in Utility Computing* (pp. 386-404).

www.irma-international.org/chapter/adaptive-mobile-architecture-with-utility-computing/139854

Construction of a Research Model for Overseas Dissemination of Chinese Literature Based on Deep Learning Models

Hongjuan Zhang (2025). *International Journal of Distributed Systems and Technologies* (pp. 1-17).

www.irma-international.org/article/construction-of-a-research-model-for-overseas-dissemination-of-chinese-literature-based-on-deep-learning-models/375389

Parallel Distributed Patterns Mining Using Hadoop MapReduce Framework

Ishak H. A. Meddah and Khaled Belkadi (2017). *International Journal of Grid and High Performance Computing* (pp. 70-85).

www.irma-international.org/article/parallel-distributed-patterns-mining-using-hadoop-mapreduce-framework/182342