私立東海大學

資訊工程與科學研究所

碩士論文

指導教授: 楊 朝 棟 博士

使用平行虛擬檔案系統建置網格大型儲存系統

On Construction of a Large File System Using PVFS for Grid

研 究 生: 潘 介 棟

中華民國九十四年六月

摘要

網格 (Grid) 是繼 Internet 之後又一項重大的網路進步。這是由於網格系統 (Grid System) 具有廣泛使用性、高效能等特質。然而，網格系統中資料儲存空間的一致性和共用性亦受到挑戰。因此，計算網格 (Computational Grid)與資料網格 (Data Grid)協同應用上的問題更加顯的重要。運用叢集系統及其平行技術建置一套具有可用度高、擴充性強、高運算能力又兼具大型的儲存空間在網格下，將可解決上述問題。在這篇論文中，我們提出一個可以在網格系統裏，同時擁有高速運算的計算網格及高存取率的資料網格解決方法。在我們所架設環境中，計算網格部份，使用了叢集計算系統 (Cluster Computing)增加系統協同運算效能。資料網格部份，運用較高存取效能的平行檔案系統 (PVFS)，已充分運用網格系統中有效的儲存空間。整個系統設計提高了網格系統的多用性能，提高存取率及運算效能。不論對大量資料存取或是巨量運算，都提供相當不錯的平台給網格使用者運用。

關鍵字: 叢集計算、網格計算、資料網格、平行虛擬檔案系統

Abstract

Grid is a great progress of network after Internet because the Grid System provides a specialty that it can be used popularly and effectively. However, it is a challenge to the consistency and community of use on the data storages space of Grid System. Therefore, the problem of application for the Computational Grid and Data Grid is more important. It can set up a usability, expandability, high operation capability, and large memory space in Grid with the Cluster system and parallel technique in order to solve the problem. In this paper, we provided a Grid with high operation capability and higher storage to solve the problem. As to the Grid setting, we take use of the Cluster computing to increase the operation effect for computing, and a PVFS2 with more storages effect for data. It can supply a quite correct platform for Grid user whether for large data access or huge operation.

Keywords: Cluster computing, Grid Computing, Data Grid, Parallel Virtual File Systems (PVFS), GridFTP

Acknowledgements

I would like to thank all of those people who have helped me through the completion of this thesis. In particular, I would like to thank my advisor, Dr. Chao-Tung Yang, who introduced me to this topic and gave me a broad support and guidance. I would also like to thank Professor Wuu Yang, Professor Ching-Hsien Hsu, Professor Hsien-Cheng Yu and Professor Yi-Min Wang for their valuable comments and advice on my oral defense for this thesis.

There are many people whom I would like to thanks, especially to my family. Without their encouragement and support, I couldn't pursue further education in the graduate school of Tunghai. Many of my classmates have encouraged and supported me through this research. For this reason, I could complete my work without fear.

Lastly, I would like to thank all my coworkers and friends whose unconditional support making this thesis possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

High performance, distributed computing and computational sciences require large data sets, fast and efficient ways of getting to that data, and a security model that will protect the integrity of the stored data. In order to create enough usable space without spending large amounts of money for storage, multiple storage servers need to be used in groups.

The Grid Computing discipline involves the actual networking services and connections of a potentially unlimited number of ubiquitous computing devices within a “Grid”. This new innovative approach to computing can be most simply thought of as a massively large power “utility” grid, such as what provides power to our homes and businesses each and every day. This delivery of utility-based power has become second nature to many of us, worldwide. We know that by simply walking into a room and turning on the lights, the power will be directed to the proper devices of our choice for that moment in time. In this same utility fashion, Grid Computing openly seeks and is capable of adding an infinite number of computing devices into any grid environment, adding to the computing capability and problem resolution tasks within the operational grid environment.

Data grids are used to provide secure access to remote data resources: flat-file data, relational data, and streaming data. For example, two collaborators at sites A and B need to share the results of a computation performed at site A, or perhaps design data for a new part needs to be accessible by multiple team members working on a new product at different sites—and in different companies.

Grid is a great progress of network after Internet because the Grid System provides a specialty that it can be used popularly and effectively. However, it is a challenge to the consistency and community of use on the data storages space of Grid System. Therefore, the problem of application for the Computational Grid and Data Grid is more important. It can set up a usability, expandability, high operation capability, and large memory space in Grid with the Cluster system and parallel technique in order to solve the problem. In this paper, we provided a Grid with high operation capability and higher memories to solve the problem. As to the Grid setting, we take use of the Cluster computing to increase the operation effect for computing, and a PVFS2 with more storages effect for data. It can supply a quite correct platform for Grid user whether for large data access or huge operation.

The remainder of this article is organized as follows. Background and related studies are presented in Chapter 2 and our system architecture is outlined in Chapter 3. Experimental results and a performance evaluation of our scheme are presented in Chapter 4 and Chapter 5. Chapter 6 concludes this research article.

Chapter 2

Background

2.1. Cluster Computing

Scalable computing clusters, ranging from a cluster of (homogeneous or heterogeneous) PCs or workstations to SMP (Symmetric Multiprocessors)[1], are rapidly becoming the standard platforms for high-performance and large-scale computing[2][3][4][5]. A cluster is a group of independent computer systems and thus forms a loosely coupled multiprocessor system as shown in Figure 2.1. A network is used to provide inter-processor communications. Applications that are distributed across the processors of the cluster use either message passing or network shared memory for communication. A cluster computing system is a compromise between a massively parallel processing system and a distributed system. An MPP (Massively Parallel Processors) system node typically cannot serve as a standalone computer; a cluster node usually contains its own disk and is equipped with complete operating systems, and therefore, it also can handle interactive jobs. In a distributed system, each node can function only as an individual resource while a cluster system presents itself as a single system to the user.

The concept of Beowulf clusters is originated at the Center of Excellence in Space Data and Information Sciences (CESDIS), located at the NASA Goddard Space Flight Center in Maryland. The goal of building a Beowulf cluster is to create a cost-effective parallel computing system from commodity components to satisfy specific computational requirements for the earth and space sciences community. The first Beowulf cluster was built from 16 Intel® DX4TM processors connected by a channel-bonded 10 Mbps Ethernet, and it ran the Linux operating system. It was an instant success, demonstrating the concept of using a commodity cluster as an alternative choice for high-performance computing (HPC). After the success of the first Beowulf cluster, several more were built by CESDIS using several generations and families of processors and network [6].

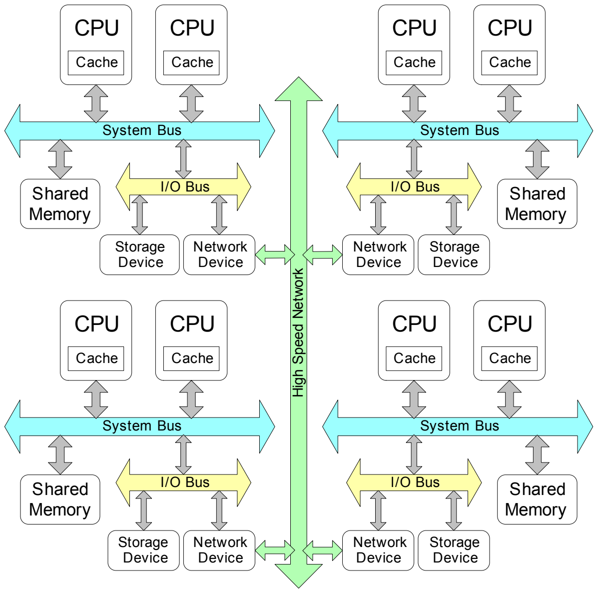

Figure 2.1: A cluster system by connecting four SMPs.

Beowulf is a concept of clustering commodity computers to form a parallel, virtual supercomputer. It is easy to build a unique Beowulf cluster from components that you consider most appropriate for your applications. Such a system can provide a cost-effective way to gain features and benefits (fast and reliable services) that have historically been found only on more expensive proprietary shared memory systems.

2.2. Grid Computing

Grid computing is a form of distributed computing that involves coordinating and sharing computing, application, data, storage, or network resources across dynamic and geographically dispersed organizations [7][8][9][10][11][12][13]. Grid technologies promise to change the way organizations tackle complex computational problems. However, the vision of large scale resource sharing is not yet a reality in many areas -

Grid computing is an evolving area of computing, where standards and technology are still being developed to enable this new paradigm. The layered Grid architecture and its relationship to the Internet protocol architecture is shown in Figure 2.2.

Grid computing is applying the resources of many computers in a network to a single problem at the same time - usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data. A well-known example of grid computing in the public domain is the ongoing SETI@Home (Search for Extraterrestrial Intelligence) project[14] in which thousands of people are sharing the unused processor cycles of their PCs in the vast search for signs of "rational" signals from outer space. According to John Patrick, IBM's vice-president for Internet strategies, "the next big thing will be grid computing."

Grid computing requires the use of software that can divide and farm out pieces of a program to as many as several thousand computers. Grid computing can be thought of as distributed and large-scale cluster computing and as a form of network-distributed parallel processing. It can be confined to the network of computer workstations within a corporation or it can be a public collaboration (in which case it is also sometimes known as a form of peer-to-peer computing).

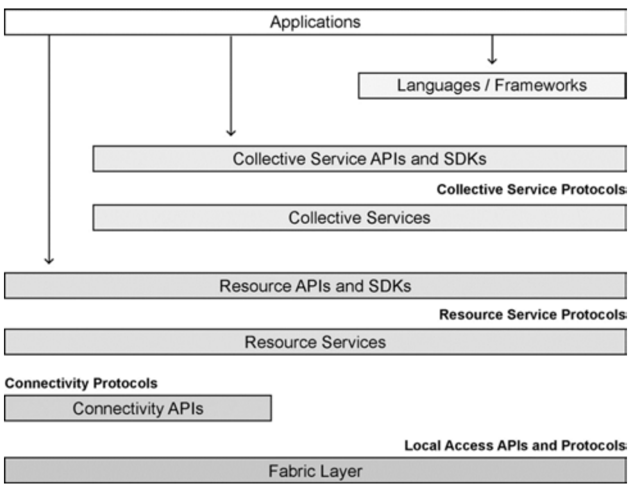

Figure 2.2: The Architecture model of Grid computing.

2.3. Data Grids

Some newly developed scientific studies such as high energy physics and computational genomics demand access of large amounts of data, a requirement that leads to such considerations such as data file management, replicated files management and transfer, and distributed data access management. The data grid infrastructure is meant to integrate the data storage devices and data management service into the grid environment. Data Grid consists of scattered computing and storage resources, which, though located in different countries, remain accessible to users.

In this research, we have adopted a grid middleware called Globus Toolkit [15][16] as the infrastructure for the data grid. Globus Toolkit provides solutions to such considerations as security, resource management, data management, and information service.

The storage system is a basic data grid component. It defines all storage technology capability such as adding, deleting, reading, writing and operating file instances. In this research we will apply PVFS2 [17] as the Data Grid storage system.

Data grids are used to provide secure access to remote data resources: flat-file data, relational data, and streaming data. For example, two collaborators at sites A and B need to share the results of a computation performed at site A, or perhaps design data for a new part needs to be accessible by multiple team members working on a new product at different sites—and in different companies[18][19].

The Data Grid project will bring together researchers from Biological Science, Earth Observation, and High-Energy Physics where large-scale, data-intensive computing is essential. The needs of these fields over coming years will provide the data - the Data Grid project will provide the computational means to handle them.

A grid server can also be configured to perform monitoring services on other grid components. Monitoring typically involves determining the response time of other components to ping messages. As the Avaki Data Grid software [20] evolves to incorporate access to relational data, the grid server is expected to perform database tasks such as opening a connection, issuing a query, or executing a stored procedure and reporting results into the data grid.

2.4. PVFS

Parallel I/O continues to be a topic of active development. Recent years have seen the creation of many new options. Even with these new choices, certain factors remain constant. Parallel applications need a fast I/O subsystem. Clusters need a parallel file system that can scale as the number of nodes increases to the thousands and tens of thousands. PVFS is our answer.

Many institutions and researchers have used the first generation of the Parallel Virtual File System (PVFS) with much success. The system architecture is shown in Figure 2.3. The time has come for the second generation. PVFS2 continues to serve as both a platform for parallel I/O research as well as a production files system for the cluster computing community.