Chapter XIII
Service Oriented Storage System Grid

Yuhui Deng
Cranfield University, UK

Frank Zhigang Wang
Cranfield University, UK

Na Helian
Metropolitan University, UK

ABSTRACT

Storage Grid is a new model for deploying and managing the heterogeneous, dynamic, large-scale, and geographically distributed storage resources. This chapter discusses the challenges and solutions involved in building a Service Oriented Storage (SOS) Grid. By wrapping the diverse storage resources into atomic Grid services and federating multiple atomic Grid services into composite services, the SOS Grid can tackle the heterogeneity and interoperability. Peer-to-peer philosophy and techniques are employed in the SOS Grid to eliminate the system bottleneck and single point of failure of the traditional centralized or hierarchical Grid architecture, while providing dynamicity and scalability. Because Grid service is not designed for critical and real-time applications, the SOS Grid adopts Grid service to glue the distributed and heterogeneous storage resources, while using binary code to transfer data. The proposed methods strike a good balance among the heterogeneity, interoperability, scalability and performance of the SOS Grid.

INTRODUCTION

According to a new report from IDC (IDC white paper, 2007), 161 exabytes of digital information were created and copied in 2006. The growth will continue to increase exponentially. The amount of information in 2010 will surge more than six fold to 988 exabytes which amounts to a compound annual growth rate of 57%. About 70% of the digital information will be generated...
by individuals over the next three years. The data will be stored in a large number of data centers which are distributed across the Internet. The data centers may have completely heterogeneous operating systems, computer architectures, and IT infrastructures.

The explosive growth of data has been identified as the key driver to escalate storage requirements. There are two major technologies which impact the evolution of storage systems. The first one is parallel processing such as redundant arrays of inexpensive disks (RAID) (Gibson, et al, 1988). The second one is the influence of network technology on storage system architecture. Network based storage systems such as network attached storage (NAS) and storage area network (SAN) (Gibson and Meter, 2000; Morris and Truskowski, 2003) offer a robust and easy method to control and access large amounts of storage resources. However, the ever increasing amounts of data generated worldwide incur a significant impact on the storage systems we have today (Min, et al, 2005). It requires more sophisticated techniques and more flexible and reliable storage systems to store and manage the data (e.g. providing petabytes and even exabyte storage capacity, and aggregate bandwidth over 100 GB/s). Undoubtedly, NAS and SAN cannot meet the requirements. It is a big challenge to design an autonomous, dynamic, large-scale and scalable storage system which consolidates distributed and heterogeneous storage resources to satisfy both the bandwidth and storage capacity requirements.

A Grid is a flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources (Foster, et al, 2001). The objective is to virtualize resources including computers, networks, instruments and so on and allow users and applications to access the resources in a transparent manner. A Grid environment may consist of hundreds or even thousands of geographically distributed and heterogeneous resources to match the requirements imposed by all kinds of Grid applications. Grid computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large scale resource sharing and high performance orientation. Table 1 illustrates the characteristics of Grid and data storage.

Drawing inspiration from the Grid computing community, a storage Grid system is supposed to be able to satisfy the ever-increasing data storage requirements. The main goal of the storage Grid is to shield the heterogeneity of geographically distributed and diverse storage systems involved in the data storage and provide storage resources on demand for every authorized Grid user. Based on the existing Internet infrastructure, the storage Grid is a virtual organization which combines geographically distributed and heterogeneous storage systems into a logical community with only minimal administrative requirements.

**BACKGROUND**

In recent years, several research projects have investigated techniques in designing storage systems in a Grid environment. Storage Resource Broker (SRB) (Storage resource broker, 2007) supports

---

**Table 1. Characteristics comparison of Grid and data storage**

<table>
<thead>
<tr>
<th>Grid Characteristics</th>
<th>Data Storage Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large scale or global distribution</td>
<td>Worldwide</td>
</tr>
<tr>
<td>Dynamic coordination</td>
<td>Ever-increasing</td>
</tr>
<tr>
<td>Collaborative virtual organizations</td>
<td>Require Interoperability</td>
</tr>
<tr>
<td>User transparent</td>
<td>Involve large quantity of heterogeneous storage resources</td>
</tr>
<tr>
<td>Secured communication</td>
<td>High security, privacy, and reliability</td>
</tr>
</tbody>
</table>