A Case for the Global Access to Large Distributed Data Sets using Data Webs Employing Photonic Data Services

Robert L. Grossman
University of Illinois at Chicago
grossman@uic.edu

Yunhong Gu
University of Illinois at Chicago
gu@lac.uic.edu

Dave Hanley,
University of Illinois at Chicago
dave@lac.uic.edu

Xinwei Hong
University of Illinois at Chicago
xwhong@lac.uic.edu

Jorge Levera
University of Illinois at Chicago
jorge@lac.uic.edu

Marco Mazzucco
University of Illinois at Chicago
marco@lac.uic.edu

Dave Lillethun
Northwestern University
davel@alumni.northwestern.edu

Joe Mambretti
Northwestern University
j-mambretti@northwestern.edu

Jeremy Weinberger
Northwestern University
jeremy@northwestern.edu

October 18, 2003

This is a draft of a paper which later appeared in Proceedings of the 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2003), IEEE Computer Society, Los Alamitos, California, pages 62-66.

Abstract

We argue that data webs employing specialized path services, network protocols, and data protocols can be an effective platform to analyze and access millions of distributed Gigabyte (and larger) size data sets. We have built a prototype of such a data web today and demonstrated that it can effectively access, analyze and mine distributed Gigabyte size data sets even over thousands of miles by using specialized network and data protocols. The prototype uses a server which employs the DataSpace Transfer Protocol or DSTP. Our assumption is that WSDL/SOAP/UDDI-based discovery and description services will enable this same infrastructure to scale to millions of such DSTP-Servers.

1 Introduction

In this paper, we argue that data webs employing specialized path services, network protocols, and data protocols can be an effective platform to access millions of distributed Gigabyte (and larger) size data sets. We have built a prototype of such a data web today and demonstrated that it can effectively access, analyze and mine distributed Gigabyte size data sets even over thousands of miles by using specialized network protocols.

There are three broad approaches emerging for providing services to remote and distributed data. Currently, the most common is to provide web front ends to remote file systems, databases, archival systems, and hierarchical storage systems. For a general survey of work in this and related areas, see [9].

Data grids and data webs are two emerging technologies which are complementary to this approach. Data grids are infrastructures which provide authentication, authorization, and access controls (AAA) so that individuals can employ remote and distributed high end computing resources. Globus is the most common grid middleware [2]. Data grids are grids in which certain additional services are run, including services for parallel TCP striping (GridFTP), and data replication services (Globus Replica Catalog and Globus Replica Management) [1].

Data webs are web based infrastructures for data employing web services to provide access to remote and distributed data [5]. Data webs for working with large data sets also employ specialized protocols and services for precisely this purpose. We describe these specialized protocols and services below.

We note that recently there have been efforts to incorporate some web services into the grid framework [3].

In this paper, we describe a data web we have built called DataSpace [4]. Version 1.0 of DataSpace was developed before standards for web services had been defined and used XML for working with metadata and a protocol called the DataSpace Transport Protocol (DSTP) for packaging data and metadata for transport. The current version of DataSpace is Version 2.2. For these experiments, we developed an experimental version of DataSpace derived from Version 2.2, which includes support for standard web services, including SOAP/XML for packaging metadata and the Web Service Description Language (WSDL) for the description and advertising of metadata and data services. These services should scale and easily provide web based access to the *metadata* for millions of distributed data sets.

On the other hand, as we will show below, SOAP/XML and standard network protocols such as TCP, FTP, and HTTP are not adequate for dealing with the data itself. We have introduced specialized path services, network protocols and data services which to handle the transport of the data itself, which we describe below. Finally, we describe experiments showing the speedup gained with this approach applied to some typical data mining algorithms, such as computing simple correlations for streaming geo-science data.

2 Data Webs

A fundamental challenge is to develop technology which scales as the number of data providers and data consumers grows. Data webs are web based infrastructures for data which are designed to scale to very large numbers of data producers and data consumers in the same way that the web today scales to very large numbers of document producers and document consumers. Data webs are designed to facilitate the access, remote analysis and distributed mining of remote and distributed data.

Today, there a variety of technologies to access, manage, and mine distributed data. One way to sort out these different technologies is to differentiate the action (viewing, data mining, or computing) and the object of the action (files, attributes/columns, or higher order concepts), as in the table below. From this viewpoint, data webs are designed for remote data analysis and distributed data mining. Data grids are designed to support high end computation of distributed data. Persistent archives are designed to support the viewing of archived files.

To make these distinctions a little clearer, consider a petabyte of data. To those interested in persistent archives, a petabyte is often thought of 1015 bytes managed by hierarchical storage system providing archival services, caching, and migration. To those interested in data grids, a petabyte is often thought of as one thousand one TB data sets. The goal is to develop middleware services supporting authorization, authentication, and access controls (AAA) so that virtual high performance computing resources can be used to analyze the data. To those interested in data webs, a petabyte is often thought of as either one million 1 GB data sets or one hundred million 10 MB data sets. The goal is to support the discovery, correlation, normalization, transformation, and analysis of millions of remote and distributed data sets.

3 Photonic Data Services

In this section, we describe the layered model we use for building data webs. The model is one of the standard one models used and combines the standard TCP layered model with one of the models being used for web services.

Standard web services do not scale to large remote and distributed sets as is becoming well known and as we point out in the Section on Experimental Studies below. To overcome this limitation, we have recently integrated the following three services:

1. path services to set up, tear down and check the status of photonic paths on a per application basis;
2. high performance network transport protocols specifically designed to work with large remote data sets;
3. and data services specifically designed for remote data analysis and distributed data mining, including mechanisms for separating data, metadata, and keys, for merging data streams by keys, for sampling, etc.

We call the integration of these three services *Photonic Data Services.* For the work described in this paper, we have recently developed a data web which uses standard web services for working with metadata and small data sets, while using Photonic Data Services for working with the data per se. This combination should scale to millions of Gigabyte size data sets and larger.

1. Physical Links. We assume that the physical links are provided by multichannel wavelength-division multiplexed (WDM) communications, as well as by Ethernet, and other technologies.

2. Path Services Layer. We assume that there are services allowing us to set up paths between devices, tear down paths, check the status of paths, set up routing, etc. For our applications, we use the ODIN and related services for this layer [7].

3. Internet Layer. This layer provides a common network addressing and routing across multiple networks. For our applications, we use the Internet Protocol (IP) in this layer.

4. Network Protocol Services Layer. We assume that there are transport services including TCP, UDP, and other more specialized protocols providing high performance over the paths. DataSpace includes two specialized services for high performance data transport: PSockets for striped TCP and SABUL for sending data using UDP, while employing TCP for error control, rate control and congestion control [6].

5. Data Services Layer. This layer uses the DataSpace Transport Protocol or DSTP for these services [4]. In addition, beginning with DataSpace Version 2.3, any SOAP-based web service may also be used in this layer. DSTP provides direct support for metadata, keys, vector valued keys, data range queries, merging, sampling and support for missing values.

6. Description Services Layer. For the purposes here, the Web Services Description Language or WSDL has emerged as the standard for web services so that other services and applications can use them [11]. Here we assume that the metadata and data sources and services are described in WSDL.

7. Discovery Services Layer. For the purposes here involving the discovery data, metadata, and data services, we assume that UDDI will be the standard [11].

8. Application Layer. We assume that the remote data analysis and distributed data mining applications can request standard and specialized network services depending upon the applications requirements. The applications use the discovery services and description services to locate data sets and services of interest.

4 Status

In this section, we briefly describe the status of Photonic Data Services.

These experiments used SABUL as the network protocol. SABUL was developed by Project DataSpace and is maintained as a Source Forge Project. The current release is Version 2.1 [10].

These experiments used DSTP as the data service. Version 2.2 of the DSTP-Server developed by Project DataSpace was obtained from Source Forge [10] for these experiments.

Currently, the ODIN-based path services are designed for a single administrative domain. For these experiments, test applications were developed which integrated ODIN and SABUL within the OMNInet adminstrative domain. It is an open research question to extend path services across adminstrative domains.

These experiments used WSDL/SOAP for the description services layer. An experimental version of the DSTP-server developed by Project DataSpace was used for these experiments. WSDL/SOAP support is expected to be included in Version 2.3 of the Project DataSpace DSTP-Server.

5 Experimental Studies

For our experiments we used the Chicago area OMNInet [8] and the global Terra Wide Data Mining Testbed [12]. OMNInet is an optical networking testbed deployed in the Chicago metropolitan area. OMNInet currently provides 1 GE and 10 GE services between Northwestern, the University of Illinois at Chicago, and the StarLight facility in Chicago. The Terra Wide Data Mining Testbed (TWDM) is a testbed built on top of DataSpace for the remote data analysis and distributed data mining. Currently, the TWDM Testbed consists of five geographically distributed workstation clusters linked by optical networks through StarLight in Chicago.

The experimental setup was as follows. Data servers were located at the SARA research facility in Amsterdam and at the University of Illinois at Chicago and connected via an OC-12 network. Data was streamed from both sources and merged at StarLight using a common key. StarLight and the University of Illinois at Chicago are located several miles a part. The machine performing the distributed merge was connected by OC-12 paths to both remote data sources. Merging data in this fashion is a basic primitive operation in distributed data mining.

The machine in Amsterdam was a dual P4, 1700 Mhz, with 512M RAM. The machines in Chicago were dual PIIIs, 1000Mhz, with 512M RAM. The machines were all running Linux, with the 2.4.x kernels. The network traffic was over SurfNet and OMNInet with routing providing 622 Mb/s of maximum bandwidth.

The first experiments used TCP as the network protocol and DSTP as the data service protocol. Each data stream was 300 MB in size. As can be seen from the Table 2, the average speed varied between 4-5 Mb/s, despite the fact that each link had a maximum available bandwidth of 622 Mb/s.

The second set of experiments used a path service set up by ODIN, SABUL as the network protocol, and DSTP as the data service protocol. This time each data stream was 1.8 GB in size. As can be seen from Table 3, the average speed was above 500 Mb/s.

The third set of experiments compared GridFTP and SABUL as alternate network protcols. The results for GridFTP depend critcally upon the parameter values. For these sets of experiments between StarLight and SARA we found that a a TCP buffer of 1024KB and 32 parallel sockets provided optimal results. The fourth table shows the data rates for data sets of sizes 100 MB, 500 MB, 1 GB and 2 GB. GridFTP's data rates varied between 95 Mb/s and 324 Mb/s. SABUL's data rate varied between 476 Mb/s and 527 Mb/s. It is important to note that SABUL did not require any tuning, while sveral hours were spent searching for the optimal parameter values for GridFTP.

1. The first table summarizes some of the differences between data webs, data grids and persistent archives.

2. The second table shows the result of merging two remote 300 MB data sets using standard network protocols and DSTP-based data services. As the data become more disordered, the algorithm lowers the accuracy so as to process the data at line speed. The tests were done between StarLight and SARA using the TWDM testbed.

3. The third table shows the result of merging two remote 1.8 GB data sets with SABUL-based network services and DSTP-based data ser-

Table 1: Some of the differences between data webs, data grids and persistent archives.

Knowledge	Digital Libraries	Knowledge Mining	Semantic Web
Attributes	Web accessible databases	Data Webs	Data Grids
Files	Persistent Archives	Distributed Data Mining	Grids
	View	**Mine/Discover**	**Compute**

Table 2: Merging two remote 300 MB data sets using standard network protocols and data services using DSTP.

Rand %	**Match %**	**Time (sec)**	**Data Rate (Mb/s)**
2	96.6	513	4.68
10	89.9	540	4.44
20	81.5	531	4.52
33	73.1	563	4.26

Table 3: Merging two remote 1.8 GB data sets with a path set up using ODIN, network services using ODIN, and data services using DSTP. Standard web services can be used to discover these two data sets and set up access to them.

Rand %	**Match %**	**Time (sec)**	**Data Rate (Mb/s)**
2	99	53.3	550
10	91	52.4	550
20	83	56.2	512
33	78	54.6	527

Table 4: The data rate for GridFTP and SABUL over the TWDM testbed between StarLight and SARA for different data set sizes. GridFTP used a TCP buffer of 1024KB and 32 parallel sockets.

Data Set (MBs)	**GridFTP (Mb/s)**	**SABUL (Mb/s)**
100	94.9	527
500	246	476
1000	324	506
2000	315	506

vices. The tests were done between StarLight and SARA using the TWDM testbed.

4. The fourth table compares data rate for GridFTP and SABUL over the TWDM testbed between StarLight and SARA for different data set sizes. The results for GridFTP used a TCP buffer of 1024KB and 32 parallel sockets, which was close to the optimal for the test conditions.

References

[1] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke. The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. *Journal of Network and Computer Applications*, 23:187–200, 2001.

[2] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco, California, 1999.

[3] Globus. Retrieved from http://www.globus.org, September 2, 2002.

[4] R. Grossman, M. Hornick, and G. Meyer. Data mining standards initiatives. *Communications of the ACM*, 45(8):59–61, 2002.

[5] R. Grossman and M. Mazzucco. Dataspace - a web infrastructure for the exploratory analysis and mining of data. *IEEE Computing in Science and Engineering*, 2002.

[6] R. L. Grossman, M. Mazzucco, H. Sivakumar, and Y. Pan. Simple available bandwidth utilization library for high-speed wide area networks. *submitted for publication*, 2001.

[7] D. Lillethun, J. Mambretti, and J. Weinberger. Odin: Path services for optical networks, in preparation. *www.icair.org*, 2002.

[8] J. Mambretti. Omninet. *www.icair.org/omninet*, 2002.

[9] R. W. Moore, C. Baru, R. Marciano, A. Rajasekar, and M. Wan. Data-intensive computing. In Foster:Grid [2], pages 105–129.

[10] Source forge repository for project dataspace. Retrieved from www.sourceforge.net/projects/dataspace, September 2, 2002.

[11] W3c semantic web. Retrieved from www.w3.org/2001/sw/, September 2, 2002.

[12] Terra wide data mining testbed. Retrieved from www.ncdm.uic.edu/testbeds.htm, September 2, 2002.