ByteIO Performance Tests

Mark Morgan, University of Virginia
mark-morgan@Virginia.EDU

Abstract

This document describes the experimental setup for and the results of a performance test that we carried out between various University and Teragrid RP sites. With these tests we showed that the ByteIO specification for reading and writing remote grid data files can be used in a global file system to achieve reasonable bandwidth during bulk data transfers.

Description

The purpose of this performance test was to evaluate how ByteIO [1][2] performed in both the LAN and WAN environments. To do this, we decided to acquire performance numbers for peak performance of data transfers between various machines at both the University and Teragrid RP level. For all of these tests we used a Genesis II [3] grid client to download a 1 Gigabyte file from a Genesis II ByteIO service.

Experimental Setup

The ByteIO Performance Tests were designed to provide peak transfer rates for a broad selection of sites. To perform these tests a set of seven machines were selected from two universities and three Teragrid RPs. Specifically, Cicero, Sulla, and Mark iMac were used from the University of Virginia, Hess was selected from Virginia Tech, and Abe1196, BigBen, and Ranger were chosen from NCSA, PSC, and TACC respectively[1].

Rather than run the performance tests between all possible permutations of two machines (mathematically 7P2 or 54 total permutations), we instead select a representative subset of permutations that illustrates typical bandwidths possible between sites. Specifically, we measure the performance between the following pairs (the first computer in a pair being the client or reader in the indicated test):

- Mark iMac -> Abe1196

[1] In some cases, the machines listed indicate large parallel machines and the tests were in fact run from front-end or login nodes considered equivalent to the parallel machines.

- Hess -> BigBen
- Ranger -> BigBen
- Ranger -> Abe1196
- Sulla -> Cicero[2]

In the following sections I describe the specific characteristics of each of the involved machines, the interconnects between the machines, and the software used to perform the tests.

Machines

Machine	Architecture	Operating System	CPU Speed	Cores	Memory
Mark iMac	Intel Core 2 Duo	Mac OS X 10.5	2.66 GHz	2	4 GB
Sulla	Intel(R) Xeon(R)	Ubuntu 8.04.3	2.33 GHz	8	48 GB
Cicero	Intel(R) Xeon(R)	Ubuntu 8.04.3	2.33 GHz	8	24 GB
Abe1196	Intel 64	Redhat Enterprise Linux 4	2.33 GHz	8	8 GB
Ranger	Opteron	CentOS	2.3 GHz	8	16 GB
BigBen	AMD Opteron	SuSE Linux	2.6 GHz	2	2 GB
Hess	Intel(R) Xeon(R)	Linux	2.66 GHz	4	4 GB

Table 1: Hardware Characteristics of Machines Involved in Performance Tests

Interconnects

The performance tests were carried out over a number of different interconnect speeds ranging from a relatively slow 10 Mb/s to a comparatively speedy 10 Gb/s. The following table sums up the minimum interconnect speed involved in testing each pair of hosts.

Machine Pair	Minimum Interconnect Speed
Mark iMac -> Abe1196	100 Mb/s
Hess -> BigBen	1,000 Mb/s
Ranger -> BigBen	10,000 Mb/s
Ranger -> Abe1196	10,000 Mb/s
Sulla -> Cicero	1,000 Mb/s

Table 2: Interconnect Speeds Involved in Performance Tests

[2] This particular pairing of machines represents two machines in a LAN environment connected by a GigE connection.

Software

All of the performance tests were run using Genesis II as the software implementation for both the client and the server. For each pair of machines, a 1 Gigabyte file was transferred using 7 client threads, 8 Megabyte block sizes, and the ByteIO MTOM transfer mechanism.

The 1 Gigabyte file size was used because it represents a large enough data transfer to give us a good estimate on peak transfer performance. Previous performance tests have shown that 7 client threads, 8 Megabyte block sizes, and MTOM as a transfer mechanism give highly performant results. However, these values are not considered to be optimal for all machine pairings and one would expect that each of these transfer parameters could be tuned to achieve even better performance results. One should particularly consider alternative transfer mechanisms for bulk data transfer (such as GridFTP) as allowed by the ByteIO specification.

To collect the results, a client was written that would start a timer right before starting to download the file and would stop right after finishing the complete transfer. This elapsed time does not include client start-up costs or global namespace lookup. Each test was run ten times for each pair of machines. From these ten results, the highest and lowest values were thrown out and the remaining eight were averaged.

Results

Machine Pair	Average Bandwidth (Mb/s)	Standard Deviation (Mb/s)
Mark iMac -> Abe1196	65.52	1.84
Hess -> BigBen	156.12	5.10
Ranger -> BigBen	284.37	7.13
Ranger -> Abe1196	193.54	20.14
Sulla -> Cicero	600.69	24.65

Summary

It's difficult to summarize the results of these tests as more than absolute numbers without an indication of a target performance. On the surface we can see that performance between machines sitting next to each other on a rack in a machine room and connected by a GigE interconnect can achieve more than 60% of the predicted maximum bandwidth. However, as the machines move further apart and the network load between the machines increases we see the expected decrease in performance. Two interesting data sets exist however against which we can make rough comparisons. Specifically, these are performance measurements made by PSC for Lustre performance and for GridFTP performance. For the former, the PSC Lustre WANPage (http://quipu.psc.teragrid.org/wanpage/www/wanpage.php) shows that the Lustre performance between DC-WAN at Indiana and PSC tends to be around 50 MB/s or 400 Mb/s. Similarly, the PSC Speedpage website (http://speedpage.psc.edu) for GridFTP performance indicates numbers on the order of 80 to 150 MB/s for GridFTP transfers between Abe, Ranger, and BigBen (or, 640 Mb/s to 1200 Mb/s. By these estimates the ByteIO performance numbers are within an order of magnitude. Unfortunately, it is difficult to make direct comparisons as these numbers were taken with different machines and in some cases (such as the GridFTP numbers) the files are striped across machines.

Works Cited

1. Morgan, M. ed. "ByteIO Specification 1.0." GFD-R-P.087. 31 October 2007. http://www.ogf.org/documents/GFD.87.pdf.
2. Morgan, M. ed. "ByteIO OGSA WSRF Basic Profile Rendering 1.0." GFD-R-P.088. 31 October 2007. http://www.ogf.org/documents/GFD.88.pdf.
3. http://vcgr.cs.virginia.edu