A Comparison of the IBM @server 325 with the HP Integrity rx2600

Douglas M. Pase, Ph.D.
IBM eServer xSeries Performance Development & Analysis
Research Triangle Park, NC USA

Abstract

The IBM® eServer™ 325, based on the AMD Opteron™ processor, is a powerful server for high performance technical computing. Two of its rivals are the Intel® Itanium™ 2 processor-based HP Integrity rx2600 and its smaller relative, the rx1600. The rx2600 is often presented as having greater performance than the e325, but when compared dollar for dollar the e325 is several times faster, even when high-performance networks like InfiniBand are used. Both systems are well-suited for scientific and technical workloads, and several technical benchmarks from SPEC are used for this comparison. The e325 also has much better memory performance as demonstrated in measurements with the STREAM benchmark, which measures sustainable memory bandwidth in MB/s.

Introduction

The IBM eServer 325 [1], or e325, is a rack-optimized 1U system based on the AMD Opteron processor. It is optimized for scientific and technical workloads requiring efficient floating-point processing and high bandwidth to memory. Ideally suited for use in high-performance clusters, the e325 supports one or two Opteron processors and up to 12GB of DDR333 memory. Disk subsystems may be either IDE or hot-swap SCSI. The e325 uses a Non-Uniform Memory Access (NUMA) architecture that gives each processor its own path to local memory and enables the memory bandwidth to scale with the number of processors used. The processor uses a 1MB Level 2 cache to reduce the latency to data.

One of the e325 server's strongest competitors is the rack-mounted 2U Hewlett-Packard Integrity rx2600 server [2], or rx2600. Based on the Intel® Itanium™ 2 processor, the rx2600 is also optimized for scientific and technical workloads and designed to be used in clusters. Unlike the Opteron processor, Itanium 2 uses a shared Front-Side Bus (FSB) design in which both processors share the available bandwidth to memory. This design provides maximum bandwidth per processor when there is only one processor in the system, but it effectively divides the bandwidth per processor in half when there are two processors. To reduce the impact of the loss of memory performance, the Itanium 2 processor uses the very large 6MB Level 3 cache. The smaller rx1600 uses a similar design, with the same chip set and Front-Side Bus configuration, but uses slower processors and is packaged in a 1U chassis.

Large caches help in some workloads, but not in all, and memory performance is important. In this paper, we compare the memory performance of the two systems and discuss the differences.

HP benchmarks the rx2600 system aggressively to assert that its performance is superior, especially for scientific and technical applications. Comparisons are typically done on a system-by-system or processor-by-processor basis. This is a common practice in the industry, and such comparisons have their uses. But while performance comparisons are common, comparisons involving a price component are rarely mentioned. One reason may be the instability of system prices. A more likely reason is that the rx2600 is often many times more expensive than the system against which it is compared. When systems are similarly priced, it may be reasonable to do a processor-by-processor comparison. But in the case of the e325 and rx2600, they have similar performance but they are not similarly priced.

In this paper we bring the price component into the comparison and show how that changes the conclusions.

The focus of this paper is scientific and technical workloads, so we examine performance on standard benchmarks for the technical community. Benchmark suites used for comparison are SPEC CPU2000 [3], SPEC HPC2002 [4] and STREAM [5].

Nodes in a scientific cluster are connected together using a communication fabric such as Gigabit Ethernet, Myrinet or InfiniBand. Gigabit Ethernet, or GbE, is a mature standard. It is by far the least expensive of the three and has significantly less performance. Myrinet is a commonly deployed proprietary product from Myricom that has much greater performance than GbE. InfiniBand is the most expensive and has the highest performance. InfiniBand is an emerging standard that is gaining some attention. Since applications tend to scale less efficiently as the number of nodes increases, it is often cost-effective to use the faster and more expensive networks in large clusters.

Memory Throughput

Memory throughput reflects how effectively the memory can feed the processors with data. The fastest processor can operate only as quickly as it receives data. Scientific and technical applications are especially sensitive to memory performance.

The e325 is a cache-coherent Non-Uniform Memory Access (ccNUMA) system that provides excellent memory throughput relative to Intel Xeon™ and Itanium 2 processor-based systems. NUMA means that as a processor attempts to store or retrieve data from memory, some parts of memory appear to be closer than others. The advantage of NUMA is that as processors are added to a system, data paths are added to memory ─ allowing the memory bandwidth to increase.

For purposes of this paper the term *memory bandwidth* will be used to refer to the peak transfer rate the hardware can support as measured by a hardware device such as a logic analyzer. The term *memory throughput* will be used to refer to the peak payload transfer rate as measured by a software benchmark such as STREAM.

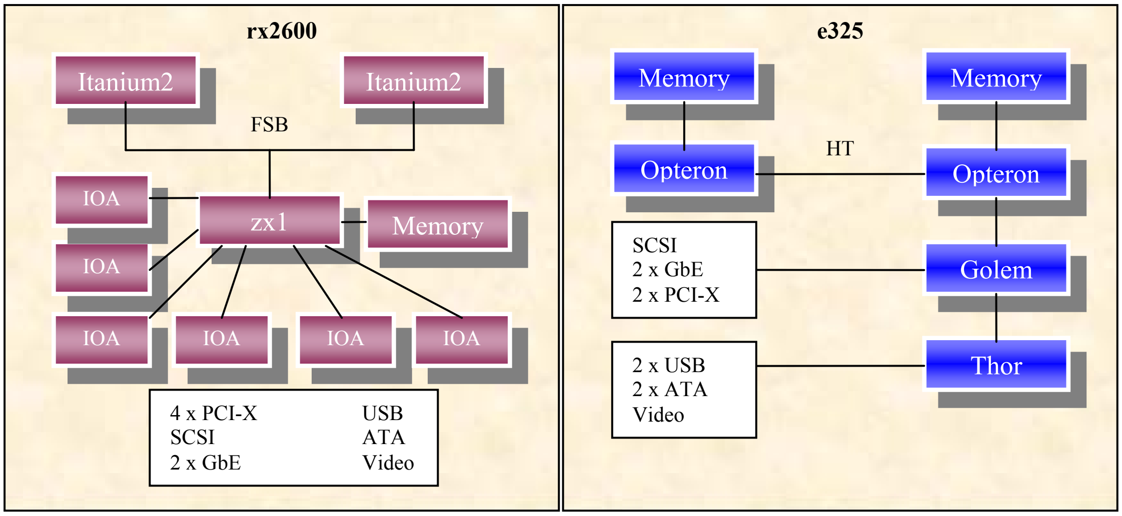

Figure 1. Block Diagrams of Memory for the rx2600 (a) and e325 (b)

To understand the significance of NUMA, consider the block diagrams in Figure 1. Figure 1a is the block diagram for the HP rx2600 and is similar to all Intel Itanium 2 processor-based systems. Notice that both Itanium 2 processors share a common Front-Side Bus (FSB) to access both I/O and memory. All memory accesses follow a common path regardless of the data being accessed or the processor originating the access request. Every byte of memory takes the same time to access from any processor. However, when both processors attempt to access memory at the same time, one will have to wait.

The memory bandwidth is limited by the speed of the Front-Side Bus, the processor and the memory controller to which it connects. The rx2600 uses the zx1 memory controller. The Itanium 2 Front-Side Bus is 16 bytes wide. Current Itanium 2 systems operate with a 400MHz clock. This limits the Front-Side Bus bandwidth to 6.4GB/s.

The bus connecting the memory controller to the memory is much faster. It supports four channels of DDR266 memory for a total bandwidth of 8.5GB/s. This allows the processors to fetch data while I/O activity also transfers data to and from memory.

Figure 1b is the block diagram for the e325. Each processor has a block of memory attached. A processor may access its local memory directly, or it may access memory attached to the other processor by using the HyperTransport Link (HT). Accessing local memory takes less time than accessing remote memory. The second processor provides a second path to memory. When both processors access their local memory at the same time, there is no interference between the processors. Access can potentially be up to double the rate of single-processor access.

Opteron processor-based systems do not have a shared Front-Side Bus, and the memory controller is integrated into the processor. They currently use 333MHz memory (DDR333). Each processor has two separate channels to memory, and each channel is 8 bytes wide. Each processor offers 5.3GB/s of bandwidth; therefore, a two-processor system supports twice the single-processor bandwidth for a total of 10.6GB/s.

Memory bandwidth is important, but memory throughput that can be used by an application is more important. We use the STREAM benchmark to measure system memory throughput. For consistency in measurements, we use only the untuned results. This benchmark performs a series of very simple arithmetic loops similar to those commonly found in scientific applications. Each loop iteration references data found in memory and not in the processor cache. The floating-point operations performed within the loop are much faster than the memory operations.

As such, the benchmark gives a good estimate of the memory throughput normally available to applications.

STREAM performance results for various systems are presented in Figure 2. The HP rx2600 dual-processor system uses the HP zx1 chipset, which is highly optimized for the Itanium 2 processor. It uses four channels of DDR266 memory. However, in spite of the high bandwidth between the chipset and memory, STREAM performance is limited by the speed of the processor's Front-Side Bus. For that reason, the HP rx2600 results should be viewed as the best results that can be obtained with an Itanium 2 processor. In Figure 2, the 1 GHz result is measured and reported publicly [6]; the 1.5 GHz result is estimated. Results for the 1 GHz rx1600 are expected to be the same as for the 1 GHz rx2600. All e325 results are measured.

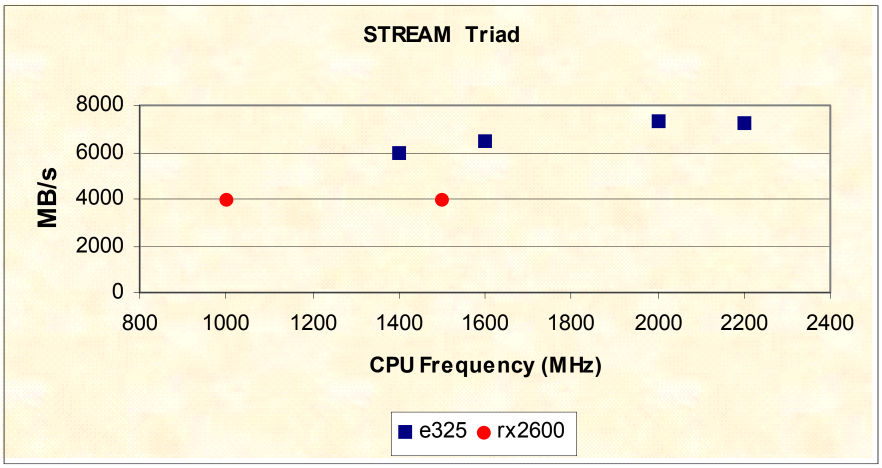

Figure 2. STREAM Performance

Itanium 2 systems are limited by the performance of the Front-Side Bus, and STREAM performance is not affected by the speed of the processor. Since the Opteron memory controller is integrated with the processor, memory performance improves with improvements in processor speed. That is why the e325 server's STREAM performance increases when the processor speed is increased. At 2.2 GHz the STREAM results are slightly lower than at 2.0 GHz because of the way in which the memory frequency must be an exact multiple of the processor frequency. Since 333 MHz is not an exact multiple of 2.2 GHz, the memory speed is slowed slightly until it is a multiple.

Figure 2 is significant because it shows the enormous difference in memory throughput realized by the different systems. Note that, depending on its frequency, memory performance on the e325 ranges from 33% to 80% better than on the rx2600.

Price/Performance

Most would agree that high performance is an important characteristic of many systems, especially larger ones. However, even the largest customers have finite budgets. Some customers have a specified budget and want the greatest performance they can obtain for that budget, while others require a minimum level of performance and seek a favorable price. Whichever it is, we're talking about price/performance.

Consider the SPEC CFP2000 Rate benchmark, which is a collection of applications intended to measure system floating-point performance under realistic workloads. The applications are run in batches with a copy of one application running simultaneously on each processor. (All pricing and performance data for the e325 and rx2600 in Figures 3 and 4 were gathered on January 8,

2004. Rx1600 price and performance data were gathered on February 27, 2004. Systems were priced as configured during measurements [3, 7, 11 and 15].)

As Figure 3a shows, when compared on a processor-by-processor basis, the rx2600 server's performance is superior to the performance of the e325. Although this comparison seems natural, it is misleading because the rx2600 is many times more expensive than the e325. When one compares system performance on a dollar-for-dollar basis (see Figure 3b) the conclusions are very different. In Figure 3b, note that lower cost is to the left on the horizontal axis and higher performance is above on the vertical axis. So the best systems (measured by price/performance) would be above and to the left of other systems.

Note that although the rx2600 is about 50% faster, its cost is more than four times as much for that amount of performance. Put differently, spending $30,000 (the cost of an rx2600) for e325 servers would produce a system capable of about 2.5 times greater performance than the rx2600 server can deliver.

The situation is a little better for the rx1600. In this case, a dual-processor rx1600 has performance that is very similar to an e325's performance, but it costs about 1.5 times as much. So the rx1600's price/performance is better than that of the rx2600, but the e325 is still much better than both.

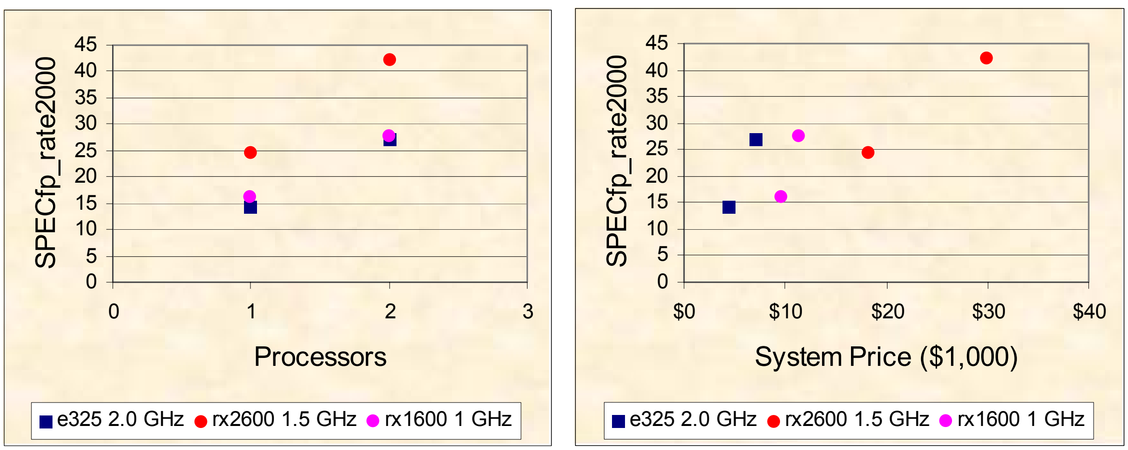

Figure 3. SPEC CFP2000 Rate Performance by Processors (a) and by System Price (b)

Because the Itanium 2 processor is not especially well-optimized for integer processing, the effect is even more dramatic when the same comparison is made using the SPEC CINT2000 Rate benchmark (see Figure 4).

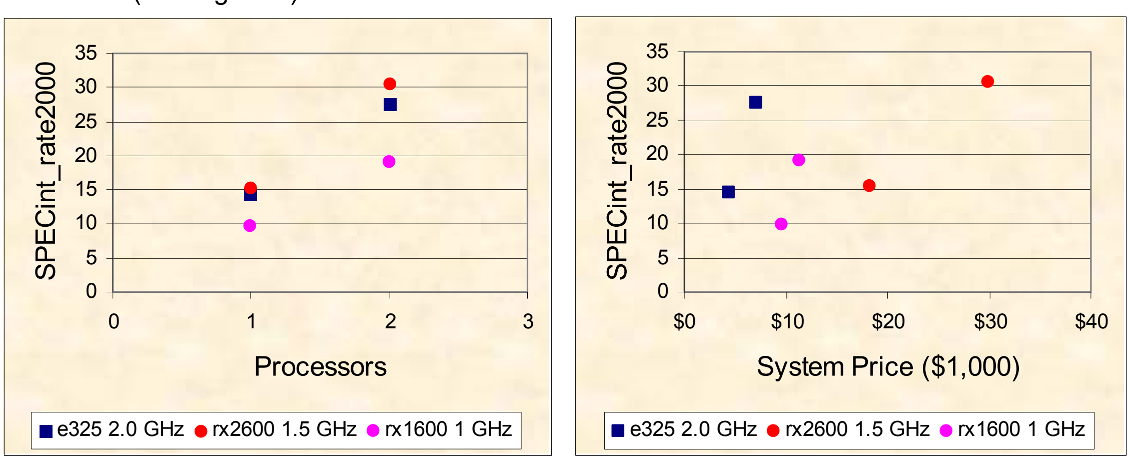

Figure 4. SPEC CINT2000 Rate Performance by Processors (a) and by System Price (b)

Both the e325 and the rx2600 are designed to operate in a clustered environment. So while results from single-node benchmarks like SPEC CFP2000 Rate and SPEC CINT2000 Rate are illuminating, it is more realistic to use cluster benchmarks for the comparison. For this reason, two cluster benchmarks, SPEC CHEM2002 and SPEC ENV2002, were chosen. Each benchmark has a Small and Medium data set. The Medium data set was used for this comparison.

SPEC CHEM2002 is a popular quantum chemistry code called GAMESS, used in the life sciences. Figure 5a shows GAMESS performance by processor count. From the figure, two things are clear. First, as the processor count increases above about 60, the benchmark becomes increasingly communication-bound. Adding more processors does not make the application run faster. Second, the rx2600 cluster appears to be slightly faster than the e325 cluster.

Once again, results from a processor-by-processor comparison do not take price into account. Figure 5b shows the same performance using a dollar-for-dollar comparison. The rx2600 cluster still shows slightly better performance on the high end, but notice that the cost of the smaller rx2600 cluster is much higher than even the largest e325 cluster used. (Data was not available for the rx1600.)

To make this comparison, each cluster was priced according to the components as listed in the published benchmark disclosures and the prices posted on the HP and IBM Web sites. All major components such as compute nodes, file servers and networks were included in the estimates. Incidental components such as racks and cabling were not included. All performance and pricing data for the e325 and for the rx2600 using Myrinet were collected on January 8 through January 22, 2004. Data for the rx2600 using InfiniBand was updated March 4, 2004 [4, 7-14].

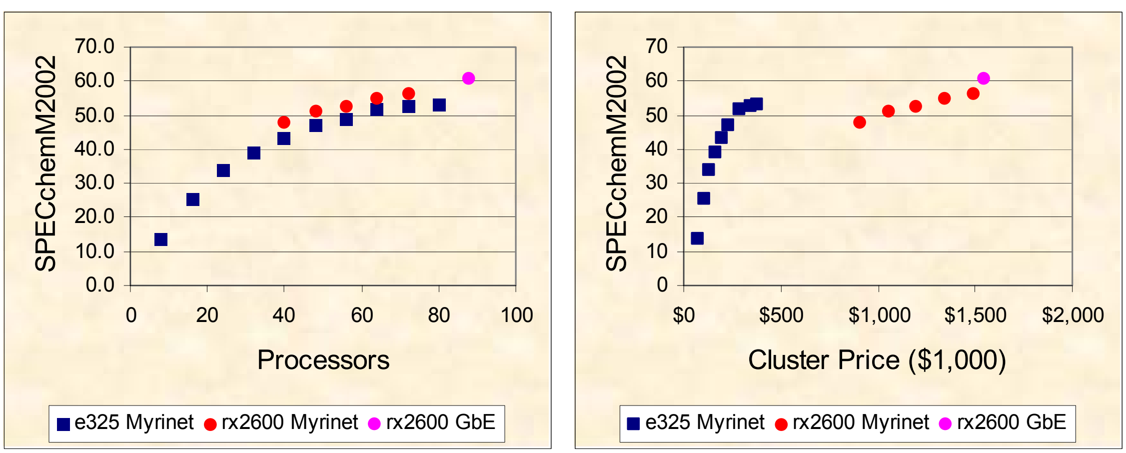

Figure 5. SPEC CHEM2002 Performance by Number of Processors (a) and by Cluster Price (b) (All rx2600 results were published. All e325 results were measured, some were published and the remaining results are regarded as measured estimates.)

SPEC ENV2002 is the environmental code WRF. It computes a weather forecasting model that is cache-friendly. As a result, the rx2600 with its larger cache performs somewhat better than the e325, as Figure 6 shows. Once again, when factoring in the cluster cost, the rx2600 is at a serious disadvantage even though the benchmark is well-suited to the Itanium 2 processor. Using a processor-by-processor comparison of both systems using Myrinet, the rx2600 can be as much as 37% faster than the e325. But using a dollar-for-dollar comparison the rx2600 is more than 3.3 times **slower**.

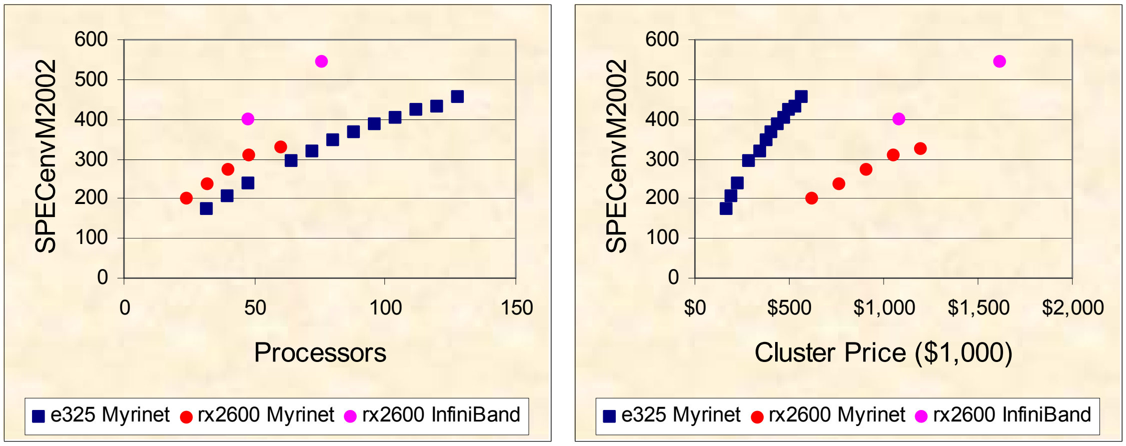

Figure 6. SPEC ENV2002 Performance by Number of Processors (a) and by Cluster Price (b) (All e325 and rx2600 results were published.)

Conclusions

The AMD Opteron processor-based, dual-processor IBM eServer 325 is a powerful 1U system. Its competition is the Intel Itanium 2 processor-based, dual-processor HP Integrity rx2600. Both systems are well-suited for scientific and technical applications. Both systems are designed to be used in clusters.

The e325 supports high memory bandwidth and throughput because of its integrated memory controller and NUMA architecture. The second processor in the system adds more bandwidth to the system, preventing memory bottlenecks. The rx2600's bandwidth and throughput are limited because of its shared Front-Side Bus. Adding a second processor to an rx2600 increases the demand and creates contention for the bus, reducing system performance. As a result, the e325 memory performance is up to 80% faster than that of the rx2600.

The e325's performance is also superior when measured on a dollar-for-dollar basis against the rx2600 in many benchmarks. When taking price into account, the e325's price/performance ratio is more than 2.5 times better when measured using the SPEC CFP2000 Rate benchmark. It is 3.8 times better when measured using SPEC CINT2000 Rate. In a cluster environment, the e325 is 3.3 and 3.5 times better using SPEC ENV2002 and SPEC CHEM2002, respectively. A similar comparison shows the e325 is superior to the rx1600, but by a smaller margin.

A single rx2600 may be priced at 3 to 5 times the price of an e325 with only marginal improvement in performance. The rx1600 is priced from 1.5 to 2 times the price with little or no performance advantage. This allows the cost-conscious customer to purchase an e325 or e325 cluster with the same level of performance at a small fraction of the cost of a similar rx2600 or rx1600 solution. At the same time, it allows the performance-conscious customer to purchase much greater performance for the same dollars than would be possible if purchasing an rx2600 or rx1600 solution.

References

1. IBM eServer 325, ftp.software.ibm.com/pc/pccbbs/pc_servers_pdf/e325spec.pdf.
2. HP Integrity rx2600 Server – summary, www.hp.com/products1/servers/integrity/entry_level/rx2600/index.html.
3. SPEC CPU2000, www.spec.org/cpu2000/.
4. SPEC HPC2002, www.spec.org/hpc2002/.
5. Memory Bandwidth: STREAM Benchmark Performance Results, www.cs.virginia.edu/stream/.
6. HP rx2600 Stream Results, www.cs.virginia.edu/stream/stream_mail/2002/0025.html.
7. Servers from IBM: eServer 325, www-132.ibm.com/webapp/wcs/stores/servlet/CategoryDisplay?catalogId=-840&storeId=1&categoryId=2583808&langId=-1&dualCurrId=73.
8. Servers from IBM: xSeries 345, www-132.ibm.com/webapp/wcs/stores/servlet/CategoryDisplay?catalogId=-840&storeId=1&langId=-1&dualCurrId=73&categoryId=2559454.
9. Cisco Catalyst 4003, shop.store.yahoo.com/shopmiracle/ws-c4003-s1.html.
10. Myrinet Product List, www.myricom.com/myrinet/product_list.html.
11. HP's Small & Medium Business Online Store (Servers), h71016.www7.hp.com/dstore/SubFamilies.asp?oi=E9CED&BEID=19701&SBLID=&ProductLineId=431&FamilyId=1474.
12. HEWLETT-PACKARD HP 9000 L Class HP server rp5470, pcau.com/l3/HPU6144.php.
13. Price Compare, Find Lowest Priced Network Cards, www.activecompare.com/price_compare.asp?OpID=CatTypeModels&BOT=598948913040121&CatType=TRN20021231160552264348704&KW=A6825A.
14. HP's Small & Medium Business Online Store, h71016.www7.hp.com/dstore/optionCategories.asp?supercategoryid=168&CategoryId=2225&oi=E9CED&BEID=19701#C2225.
15. Business Wire, "HP Strengthens Industry-standard Offerings with New Servers and Storage; New Products Enable Customers to Reduce Costs While Simplifying Change," http://home.businesswire.com/portal/site/home/index.jsp%3Fepi-content%3DGENERIC&newsId%3D20040209005516&newsLang%3Den&beanID%3D202776713&viewID%3Dnews_view

IBM Systems and Technology Group

Department 5MX

Research Triangle Park NC 27709

Produced in the USA.

3-04