INFO SIZING

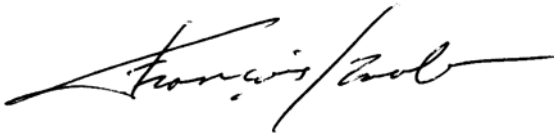
Ernie.Ostic
Ascential Software, Inc.
601 Carlson Pkwy.
Suite 330
Minnetonka, MN 55305

December 16, 2002

As per your request I verified the performance of the **Ascential DataStage XE Parallel Extender.**

Please find attached InfoSizing's detailed report and attestation of the results obtained.

Respectfully Yours,



François Raab
President

P.J. - Performance Benchmark Report
- InfoSizing's expert profile and background

Performance Benchmark Report

DataStage XE Parallel Extender

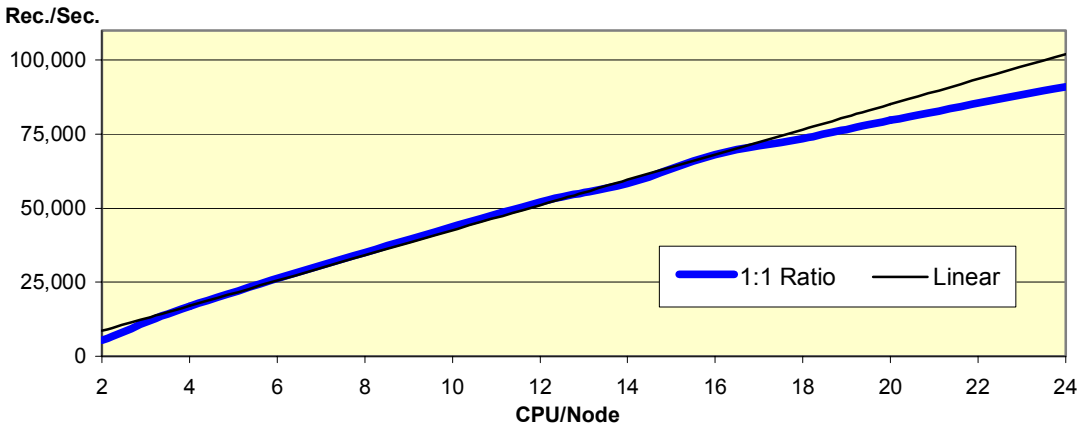
Highlights

On October 2002, at Ascential Software's request, InfoSizing verified the performance of the **Ascential *DataStage XE Parallel Extender*** by defining and auditing the execution of a benchmark on a test platform assembled at Sun Microsystem's lab in Palo Alto, Ca.

The **Ascential *DataStage XE Parallel Extender***, running on Sun's system and storage subsystem has demonstrated significant achievements:

- Near-linear scalability of the data transformation process.
- A high degree of predictability in capacity planning.
- Can process a Terabyte of data in under 7 hours on a 24 way Sun Server.

The following diagram highlights the results of the performance tests that were audited.



The graph shows how a 1 to 1 ratio of *Parallel Extender* nodes to Sun processors results in a scalability that almost perfectly matches the theoretically optimal "Linear" plot.

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

Detailed Report

The Ascential platform included a 24-way Sun multi-processor system, interconnected through fiber channels to a storage sub-system holding all input and output files. The following table presents an overview of the tested configuration.

Processor: Sun Fire 6800
• 24 US III at 900Mhz • 192 GB of Memory • 4 Fiber Channels
***Dataset:* Input Files**
• 768 GB of Raw Data • 144 Flat Files • 1.4 Billion Rows
***Storage:* Sun StorEdge 9900**
• 4 T3 Storage Arrays • 1 GB Read Cache per Array • 256 MB Write Cache per Array • 9 72GB Disks per Array • 1 RAID-0 Partition per Array • 7 Veritas Partitions Across all 4 Arrays

The following performance measurement tests were executed:

- **Node Scaling**, scaling the number of *Parallel Extender* nodes at a fixed number of CPUs.
- **CPU Scaling**
- **Sustained Throughput**, processing large amounts of records over long periods of time.

InfoSizing verified the benchmark environment and the execution of the tests. InfoSizing collected and analyzed the results of the tests. This report contains the details of our findings.

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

Data Set Population

The data set used in the test is made of flat files generated to contain random data. Each record is 534 bytes long and made of the following columns:

Number of Columns	*Datatype*	*COBOL*
30	Binary Integer	COMP
20	Packed Decimal	COMP-3
4	17 to 20 bytes	Packet

Each file contains 10,724,014 (ten million) records for a file size of 5.3 GB.

During a test, the number of files processed varies depending on the desired duration for the test and the number of *Parallel Extender* Nodes configured for the test. During the scaling tests each *Parallel Extender* Nodes processed 21GB of raw data. During the throughput test each of the 24 *Parallel Extender* Nodes processed 32GB of raw data for a total of 768 GB.

Data Transformation

The processing executed against the dataset is a data transformation. Using meta-data definitions, **Ascential *DataStage XE Parallel Extender*** converts each column of the dataset from its existing datatype to a chosen datatype. In addition, the resulting content of the columns is processed according to specified transformation rules.

The following set of transformations were executed during each test:

- 3 trim operations (removal of leading and trailing blanks)
- 1 date conversion
- 3 sub string extractions
- 3 string concatenations
- 1 column swapping (replacing the content of one by another)
- 2 column generation based on other column values
- 1 integer division
- 1 integer multiplication
- 1 addition

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

Performance Tests

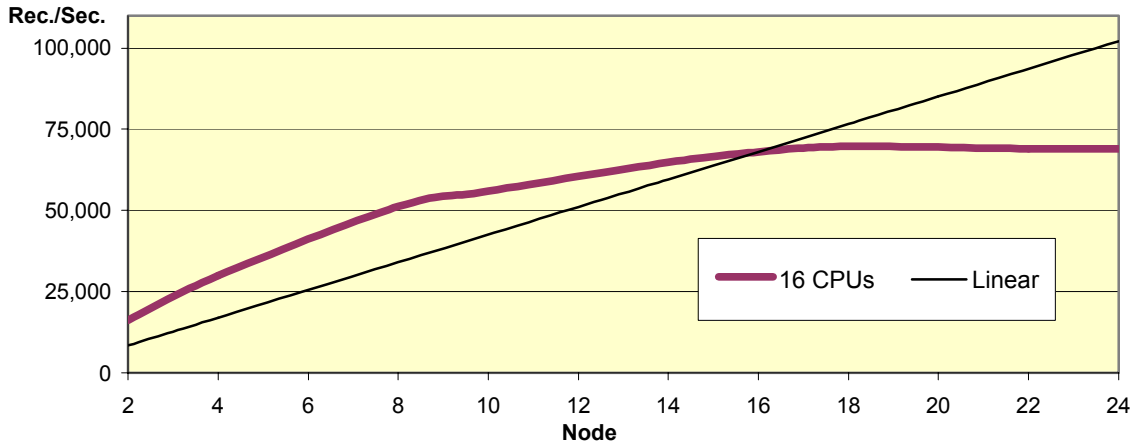
This section presents the results of each of the tests executed and audited during the benchmark.

Nodes Scaling Test

The first test evaluates the optimal number of Nodes for a fixed number of CPUs. The test system is configured with 16 UltraSparc III processors and the number of *Parallel Extender* Nodes is scaled from 2 to 24. The number of records processed per second is measured for each scaling level.

The I/O subsystem of the test system being sized to serve a full 24 processor configuration, the 16 CPU configuration is chosen to avoid reaching an I/O bottleneck and focus on the processing power. Indeed, I/O statistics collected during the test show that the average service time rarely exceeds 30 milliseconds.

The following diagram illustrates the results of the Node Scaling test.



The following observations are made:

- While the number of CPUs is greater than the number of Nodes there is excess processing power and the processing rate scales near linearly.
- As the number of Nodes reaches the number of CPUs, the processor utilization is maximized and the record-processing rate reaches its peak level.
- With a number of nodes greater than the number of CPUs the system is saturated and the record-processing rate no longer increases. In fact, it tends to diminish slightly due to excessive contention for the processing resources.

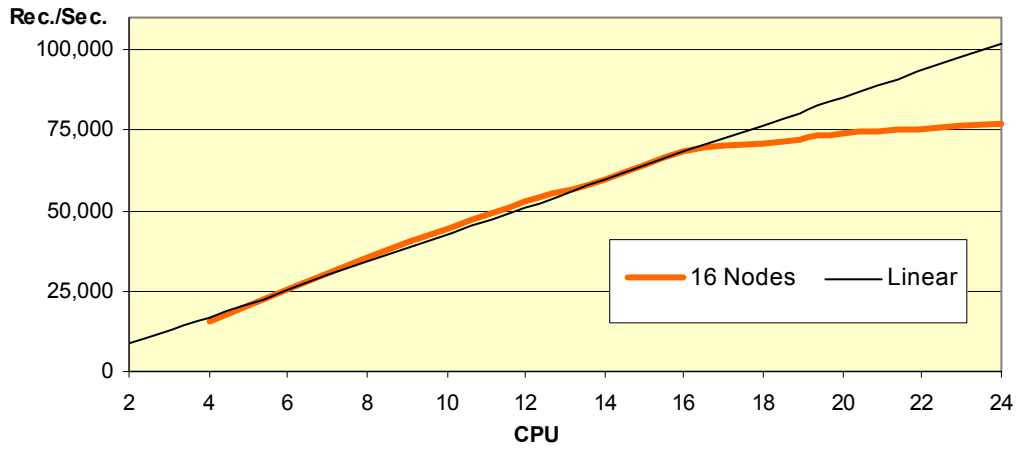
1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

CPU Scaling Test

The second test further explores the case where the number of *Parallel Extender* Nodes is greater than the number of CPUs. The test system is now configured with a fixed number of Nodes set at 16 and the number of UltraSparc III processors is scaled from 4 to 24. The number of records processed per second is measured for each scaling level.

The 16 Nodes configuration is chosen to avoid reaching an I/O bottleneck by keeping the demand on this sub-system well below its theoretical limit. The focus is on identifying the optimal number of processors needed to serve 16 Nodes. I/O statistics collected during the test show that the average service time rarely exceeds 40 milliseconds.

The following diagram illustrates the results of the CPU Scaling test.



The following observations are made:

- While the number of CPUs is lower than the number of Nodes the processing rate is limited by the number of available processors and scales linearly as this number is increased.
- As the number of CPUs reaches the number of Nodes, the demand of the 16 nodes receives near full service by the 16 processors. CPU utilization is still in the high 90% but some idle time is now consistently present.
- With a number of CPUs greater than the number of Nodes a slight increase in processing rate is observed, but it is well below linear scaling and significant idle time is now shown by the CPU statistics. This confirms that the system is best utilized when loaded with a number of *Parallel Extender* Nodes equal to or slightly greater than the number of CPUs.

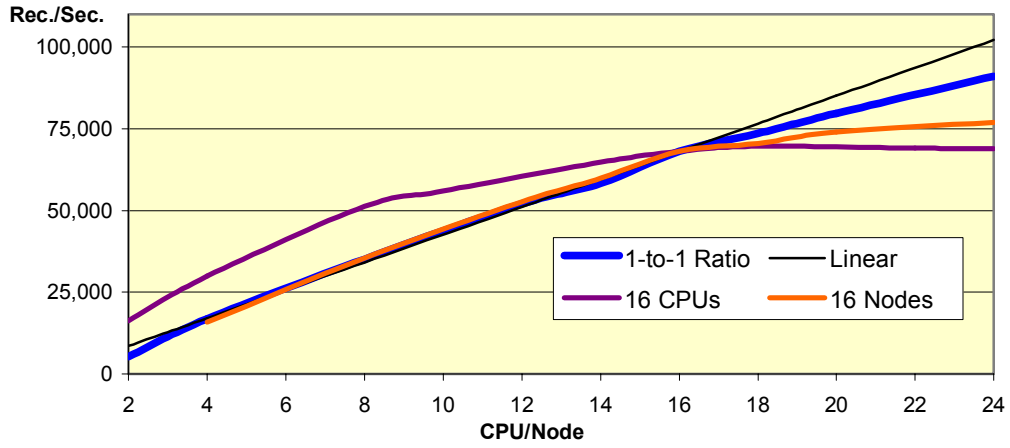
1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

Linear Scaling Test

This third test further examines the validity of the 1-to-1 configuration guideline between the number of *Parallel Extender* Nodes and the number of configured CPUs, as established in the previous two tests.

This time, the CPU configuration matches the number of Nodes and is scaled between 2 and 24. The focus of this test is to evaluate how well the 1-to-1 configuration guideline applies to the entire scaling spectrum of the test system. Once again, I/O statistics collected during the test show little I/O bottleneck with an average service time that rarely exceeds 45 milliseconds.

The following diagram illustrates the results of the Linear Scaling test. The results of the previous two tests are included for clarity.



The following observations are made:

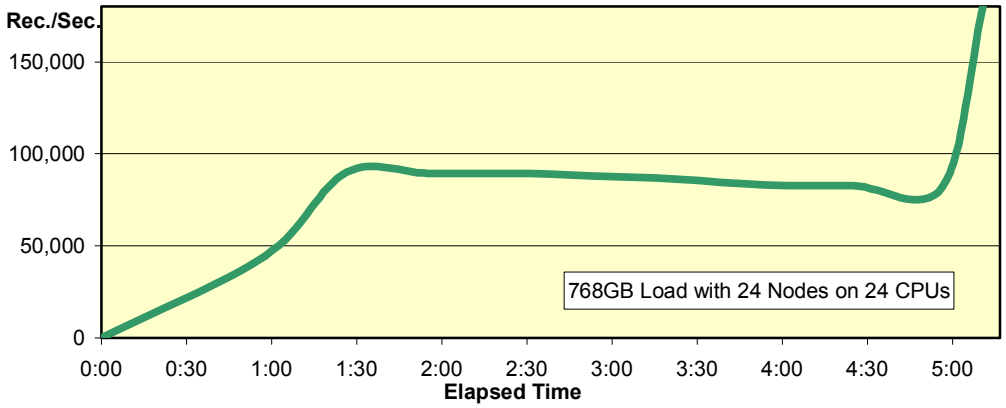
- The 1-to-1 configuration guideline holds near perfectly while scaling from 2 to 16. The scaling follows the ideal linear plot.
- Past the 16 processors level the scaling become slightly sub-linear. However, at the full capacity of 24 CPUs and 24 Nodes the scaling is only 12% shy of the ideal linear plot.
- The scaling behavior of the *Parallel Extender* is very predictable and near perfect.

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

Sustained Throughput Test

This test consists in running the Parallel Extender at full scaling capacity, with 24 nodes on 24 CPUs, for an extended period of time. The purpose of the test is to identify general processing behaviors, sustained throughput levels and derive reliable guidelines over the Terabyte level.

The system is configured with a full 192 GB of main memory. In addition, each T3s make full use of their 1 GB read ahead cache and their 256 MB write back cache.



The following observations are made:

- The processing rate is subject to a significant ramp-up during which all caches and in-memory buffers are being filled with data. During that period, the number of read I/Os is pushed to the maximum capacity of the I/O sub-system.
- Once all the caches and in-memory buffers are full, the processing rate rapidly stabilizes to a steady state level. The number of read I/Os diminishes to match the number of write I/Os as records are only read once room has been made by records processed and written back to disk.
- During the period at which records are processed at a steady state level, a gradual decrease in throughput of approximately 3% per hour is observed. This may be explained by a progressive loss of efficiency in the use of caches and in-memory buffers.
- At the tail end of the data set, all remaining records have been read into caches and in-memory buffers and the system no longer has to do any read I/Os. This elimination of part of the load results in a sharp increase in processing rate until the end of the data set is reached.
- Given the above, it is safe to extrapolate that a 24 Node configuration running on 24 UltraSparc III processors could process a Terabyte of data in under 7 hours.

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554

INFO SIZING

About InfoSizing:

Founded in 1991, InfoSizing, Inc. (www.sizing.com) is a consulting firm specializing in accurate sizing of information systems through the design, management and analysis of performance benchmarks.

InfoSizing provides consulting services in the areas of TPC Standard Benchmarks, Custom Benchmarks and Vendor Supported Benchmarks.

Benchmark design and auditing services are delivered to most leading enterprise system vendors, to the Transaction Procession Performance Council, and to innovative users of information technology.

InfoSizng's activities include:

- Consulting on the definition of performance engineering strategies;
- Designing custom benchmarks for performance tuning and technology assessment;
- Acting as independent consultant to the TPC and its membership;
- Acting as technical editor for the design of industry standards benchmarks;
- Acting as expert witness in the area of database and OLTP performance;
- Auditing benchmark implementations and results prior to public release;
- Providing independent consulting on public and private procurement.

Clients include government agencies concerned with the impartiality of their procurement projects, IT organizations looking to increase their competitive advantage by acquiring cutting edge technologies and IT vendors with a focus on engineering and marketing products where performance is a vital component and a strategic asset. Example of past and present clients include: Boeing, Bull, Compaq, Dell, Fujitsu, HP, IBM, Informix, Intel, MCI-Worldcom, Microsoft, Oracle, SGI, Siemens, Sun, Sybase, Teradata, Ariba-Tradex and UAL.

About Francois Raab:

Mr. Raab is President and founder of InfoSizing, Inc.

Recognized, award winning expert in the field of performance engineering, benchmark design and system testing, Francois Raab is the original author of the TPC-C Benchmark, the most successful industry standard measure of commercial computing performance, and is co-author of "The Benchmark Handbook" (pub. Morgan Kaufmann).

Accredited as a Certified Benchmark Auditor by the Transaction Processing Performance Council (www.tpc.org), Mr. Raab's consulting services are retained by all major system vendors as well as Fortune-500 IT organizations. With 20 years of experience in the field of databases and commercial processing, Mr. Raab is a key figure in the performance measurement, system sizing and technology evaluation community.

1373 North Franklin Street • Colorado Springs, CO 80903-2527 • Office: 719/473-7555 • Fax: 719/473-7554