



## Contents

Introduction: Choosing a Benchmark .....	1
Important System Benchmark	
Quick Reference Chart by Application .....	3
Important System Benchmarks .....	4
TPC-C™ .....	4
TPC-H™ .....	5
TPC-App™ .....	6
MMB3 .....	7
SPEC CPU® .....	8
SPECweb® .....	9
SPECjbb® .....	10
SPEC SFS® 3.0 .....	11
SPECjAppServer® .....	12
Other Benchmarks .....	13
Whetstone and Dhrystone .....	13
NetBench and WebBench .....	14
LINPACK .....	15
STREAM .....	16

## Introduction: Choosing a Benchmark

Users need to select the most appropriate benchmark to get the most accurate picture of performance for their application(s). Benchmarks vary widely in what they measure: they can measure anything from CPU/compiler integration to very specific application performance. So there are several considerations for users before they attempt to assess published results:

### ***1. Is the user's application unique to his/her organization?***

Assessment of a unique application may not be available through application-specific benchmarks. In that case the SPEC CPU benchmarks are appropriate, depending on how the application is threaded:

#### A. Single Threaded Applications

SPECint2000

SPECfp2000

#### B. Multi-Threaded Applications

SPECint\_rate2000

SPECfp\_rate2000

### ***2. Does the application have a benchmark already associated with it?***

If the application already has a legitimate benchmark test associated with it, these results may provide the closest picture to what the user can expect:

#### A. Database Applications (i.e., Oracle, Microsoft SQL Server)

TPC-C (OLTP)

TPC-H (Decision Support)

*(Continued on page 2)*



# Dell Guide to Server Benchmarks



(Continued from page 1)

- B. Messaging (i.e., e-mail, groupware)
  - MMB (Microsoft Exchange Server only)
  - SPECmail
- C. Web Serving (i.e., web server, FTP server)
  - SPECweb
  - WebBench
- D. File Server
  - SPECsfs
  - NetBench
- E. Application Server
  - TPC-App
  - SPECjAppServer

In addition to these questions, when evaluating a benchmark users need to ask:

### **3. Does the benchmark accurately represent current workloads?**

Benchmarks like Whetstone and Dhrystone were developed to assess workloads from the 1960's, and are not applicable to today's workloads.

### **4. Can the benchmark scale?**

Some benchmarks may not scale to match a specific customer environment.

### **5. Does the benchmark measure system or subsystem performance?**

Some benchmarks only measure a subsystem rather than overall system performance. The table below lists some subsystems and common benchmarks that measure those subsystems:

<b>Subsystem</b>	<b>Common Benchmarks</b>
CPU	SPEC CPU2000
Disk I/O	IOmeter, IObench, IOzone
Network I/O	Netperf, Iperf
Memory bandwidth	STREAM

### **6. What exactly does the benchmark measure? Are the results replicable? Reviewed?**

Dell recommends that customers leverage benchmarks that are owned by standards bodies to help ensure that a good result has been reviewed by a vendor's peers. Also, by using the benchmarks owned by standards bodies, there is no bias or conflict of interest to favor one vendor's products over another's.

# Dell Guide to Server Benchmarks



Important System Benchmark  
Quick Reference Chart by Application

**Key:**

- ☆ Best choice
- ⊕ Limited Applicability
- NA Not Applicable

Benchmark	Database - OLTP	Database - Decision Support	Messaging	Application Serving	Web Serving	Java VM Applications	Network File Serving	Proprietary (Internal)
TPC-C	☆	⊕	NA	NA	NA	NA	NA	NA
TPC-H	⊕	☆	NA	NA	NA	NA	NA	NA
TPC-App	⊕	⊕	NA	☆	⊕	⊕	NA	⊕
MMB	NA	NA	☆	NA	NA	NA	NA	NA
SPEC CPU	NA	NA	NA	NA	NA	⊕	NA	⊕
SPECweb	NA	NA	NA	⊕	☆	NA	NA	⊕
SPECjbb	NA	NA	NA	⊕	NA	☆	NA	NA
SPEC SFS	NA	NA	NA	NA	NA	NA	☆	NA
SPECjAppServer	⊕	NA	NA	☆	⊕	☆	NA	⊕

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### TPC-C™

<http://www.tpc.org/tpcc/>

#### *Application(s) Measured*

Database

#### *Benchmark Organization*

The Transaction Processing Council (TPC) is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable performance data based on those benchmarks.

#### *Description*

TPC Benchmark C (TPC-C) simulates a computing environment where a population of users executes transactions against a database. The benchmark is centered on the principal activities (transactions) of an order entry environment. These transactions include entering and delivering orders, recording payments, checking the status of orders, and monitoring the level of stock at the warehouses. TPC-C can represent just about any industry that may use tools like the following:

- Point-of-sale system
- Reservation system
- Order entry system
- Warehouse distribution center

#### *Why it's Important*

A long standing, highly regarded, and visible benchmark that is referred to by customers, analysts, and press

#### *Interpreting results*

TPC-C performance is measured in new-order transactions per minute. The primary metrics are the transaction rate (tpmC) and the price per transaction (\$/tpmC). The higher the tpmC, the better the performance. The lower the \$/tpmC, the better the price/performance.

#### *Factors to Consider when Evaluating Results*

- Cost of configuration
- Availability of configuration
- Performance deltas
- Price/performance

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### TPC-H™

<http://www.tpc.org/tpch/>

#### *Application(s) Measured*

Database

#### *Benchmark Organization*

The Transaction Processing Council (TPC) is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry.

#### *Description*

The TPC-H Benchmark (TPC-H) is a decision support benchmark. It consists of a suite of business-oriented ad-hoc queries and concurrent data modifications. This benchmark illustrates decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and give answers to critical business questions. TPC-H can represent just about any industry that may use tools like the following:

- Data warehouse
- Data mining
- Financial/business analysis and reporting

#### *Why it's Important*

This benchmark goes further than the TPC-C benchmark in the complexity of what it measures. It is also highly regarded and visible in the industry.

#### *Interpreting results*

TPC-H results are presented as the Composite Query-per-Hour Performance Metric (QphH), and the related price/performance (\$/QphH). TPC-H reflects multiple aspects of the capability of the system to process queries. These aspects include: the database size; the query processing power when queries are submitted by a single stream; and the query throughput when queries are submitted by multiple concurrent users. The TPC-H results are presented based on the size of the database: 100GB, 300GB, 1,000GB, 3,000GB, and 10,000GB.

#### *Factors to Consider when Evaluating Results*

- Cost of configuration
- Availability of configuration
- Actual performance differences
- Price/performance

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### TPC-App™

[http://www.tpc.org/tpc\\_app/](http://www.tpc.org/tpc_app/)

#### *Application(s) Measured*

Application Servers: Web service applications

#### *Benchmark Organization*

The Transaction Processing Council (TPC) is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry.

#### *Description*

The TPC-App benchmark is an application server and web services benchmark. The workload is performed in a managed environment that simulates the activities of a business-to-business transactional web application server operating in a 24x7 environment. TPC-App showcases the performance capabilities of application server systems. This benchmark was introduced by the TPC in December 2004. TPC-App can represent just about any industry that may leverage the following:

- Application layer of a web-based ordering system
- Online business-to-business e-commerce
- Software managed environments: J2EE.Net; BEA WebLogic; IBM Websphere

#### *Why it's Important*

This application server benchmark comes from the Transaction Processing Council, which is highly regarded for its benchmarks.

#### *Interpreting results*

TPC-App results are presented as Web Service Interactions per second (SIPS) per Application Server SYSTEM, the related price/performance (\$/SIP), and the Total SIPS, which is the total number of SIPS for the entire tested configuration or system under test (SUT).

#### *Factors to Consider when Evaluating Results*

None at press time.

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### MMB3

#### *Overview:*

<http://www.microsoft.com/exchange/evaluation/performance/mmb3.mspx>

#### *Most Current Results:*

<http://www.microsoft.com/exchange/evaluation/performance/default.mspx>

#### *Application(s) Measured*

Microsoft Exchange Server 2003

#### *Description*

MAPI Messaging Benchmark 3 (MMB3) measures the performance and scalability of computers running Microsoft Exchange Server 2003. The MMB3 benchmark is designed to mimic the workload requirements of the typical email/messaging user in an Exchange environment.

#### *Why it's Important*

Microsoft developed this benchmark and it is recognized as the defacto way of measuring Exchange Server 2003 performance, with results submitted by major computer vendors, including Dell, Fujitsu-Siemens, HP, and IBM.

#### *Interpreting Results*

An operating environment with higher MMB3 results than another is able to hold more users per server. MMB offers results based on single node servers and clustered servers. While single nodes may show higher performance results, the clustered systems offer high availability: in the case of a system failure, there is minimized (if any) downtime.

#### *Factors to Consider when Evaluating Results*

- Results should be interpreted as a benchmark for messaging throughput, not to be confused with deployment recommendations
- Factors such as backup/restore, topology and other issues should be considered when planning a deployment
- Every version of the MMB benchmark corresponds to a specific version of Exchange Server. As such, the results of one version of the benchmark are not comparable to the results of another version (i.e., MMB2 results are not comparable to MMB3 results)

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### SPEC CPU®

<http://www.spec.org/cpu2000/results/>

#### *Applications Measured*

Not application specific, SPEC CPU measures processor, chipset, and compiler speed (SPECint and SPECfp) and throughput (SPECint\_rate and SPECfp\_rate).

#### *Benchmark Organization*

The Standard Performance Evaluation Corporation (SPEC) is a non-profit group founded in 1988 to develop “realistic, standardized performance tests” for a variety of application types.

#### *Description*

The most current version of SPEC CPU is SPEC CPU2000. The benchmarks that make up SPEC CPU2000 measure the performance of the computer's processor, memory architecture and compiler. The SPECint and SPECint\_rate benchmarks measure compute-intensive integer performance, and SPECfp and SPECfp\_rate measure compute-intensive floating point performance.

#### *Why it's Important*

SPEC is a recognized standards body that provides constantly evolving benchmarks, which are highly regarded in the industry. The SPEC CPU benchmark isolates the CPU/chipset and compiler performance, so it is used heavily when positioning processors like Intel Xeon and AMD Opteron.

#### *Interpreting results*

The SPEC CPU results are as follows:

- SPECint measures how fast the computer completes a single task – this is a speed measurement
- SPECfp measures how many tasks a computer can accomplish in a certain amount of time; this is called a throughput, capacity or rate measurement.
- SPECint\_rate and SPECfp\_rate are typically used to demonstrate the performance of multi-processor systems. The results are given as base and peak, with the higher the result, the better the performance

#### *Factors to Consider when Evaluating Results*

- This benchmark is limited to the chip and chipset and associated compiler, so it does not cover I/O or other system components.
- SPECfp and SPECfp\_rate are not representative of higher-end server workloads: the tests were derived from applications such as crash simulations, ocean modeling, and quantum chromodynamics, and are more representative of HPC environments rather than database or web infrastructure

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### SPECweb®

<http://www.spec.org/web2005/results/>

#### *Applications Measured*

Web serving

#### *Benchmark Organization*

The Standard Performance Evaluation Corporation (SPEC) is a non-profit group founded in 1988 to develop “realistic, standardized performance tests” for a variety of application types.

#### *Description*

The most current version of SPECweb is SPECweb2005. It is designed to measure a system's ability to act as a web server servicing static and dynamic page requests.

#### *Why it's Important*

Like the TPC, SPEC is a recognized standards body that provides constantly evolving benchmarks, which are highly regarded in the industry. SPECweb was introduced close to 10 years ago to measure web serving performance.

#### *Interpreting results*

SPECweb2005 measures the maximum number of simultaneous connections, requesting the predefined benchmark workload that a secure web server is able to support while still meeting specific throughput and error rate requirements. The higher the result the better the performance.

#### *Factors to Consider when Evaluating Results*

- Amount of memory used in the tested configuration
- Cost of configuration

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### SPECjbb®

<http://www.spec.org/jbb2005/>

#### *Applications Measured*

Java Virtual Machine servers/applications

#### *Benchmark Organization*

The Standard Performance Evaluation Corporation (SPEC) is a non-profit group founded in 1988 to develop “realistic, standardized performance tests” for a variety of application types.

#### *Description*

The SPECjbb2005 is a Java program emulating a 3-tier system with emphasis on the middle tier, to analyze the performance of systems using Java as middleware between a database and customers. Originally developed over five years ago, SPECjbb2005 provides a new enhanced workload, implemented in a more object-oriented manner to reflect how real-world applications are designed.

#### *Why it's Important*

Like the TPC, SPEC is a recognized standards body that provides constantly evolving benchmarks, which are highly regarded in the industry. SPECjbb was introduced close to 10 years ago to measure web serving performance.

#### *Interpreting results*

The performance metrics for SPECjbb2005 are bops (business operations per second). The bops metric measures the overall throughput achieved by all the JVMs in a benchmark run. The measure bops/JVM reflects the contribution of a single JVM to the overall metric, and shows the performance and scaling of a single JVM. A higher result on bop/JVM translates to better performance of the system.

#### *Factors to Consider when Evaluating Results*

- Vendor and version of JVM used (efficiency of JVM will impact the results)
- CPU use is critical (pushes the CPU to the max)

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### SPEC SFS® 3.0

<http://www.spec.org/sfs97r1/results/>

#### *Applications Measured*

Network file serving

#### *Benchmark Organization*

The Standard Performance Evaluation Corporation (SPEC) is a non-profit group founded in 1988 to develop “realistic, standardized performance tests” for a variety of application types.

#### *Description*

SPEC SFS is a benchmark that measures network file server (NFS) throughput and response time. It evaluates the speed and request-handling capabilities of NFS systems.

#### *Why it's Important*

Like the TPC, SPEC is a recognized standards body that provides constantly evolving benchmarks, which are highly regarded in the industry. SPEC SFS was introduced 13 years ago to measure network file serving performance.

#### *Interpreting results*

SPEC SFS 3.0 is a system-level benchmark that heavily exercises CPU, mass storage, and network components. The greatest emphasis is on I/O, especially as it relates to operating and file system software. The performance metrics for SPEC SFS 3.0 include a throughput measure (in operations per second) and an overall response time measure (the average response time per operation).

#### *Factors to Consider when Evaluating Results*

- Additional hardware in the system (memory, disk controllers, disks, network controllers and buffer cache)
- Total cost of the tested configuration

# Dell Guide to Server Benchmarks



## Important System Benchmarks

---

### SPECjAppServer®

<http://www.spec.org/sfs97r1/results/>

#### *Applications Measured*

Network file serving

#### *Benchmark Organization*

The Standard Performance Evaluation Corporation (SPEC) is a non-profit group founded in 1988 to develop “realistic, standardized performance tests” for a variety of application types.

#### *Description*

SPECjAPPServer is a multi-tier benchmark for measuring the performance of Java 2 Enterprise Edition (J2EE) technology-based application servers. These application servers provide the foundation software in many modern e-commerce implementations. SPECjAppServer can represent just about any industry that may leverage the following:

- Application layer of a web-based ordering system
- Online business-to-business e-commerce
- Software managed environments: J2EE.Net; BEA WebLogic; IBM Websphere

#### *Why it's Important*

Like the TPC, SPEC is a recognized standards body that provides constantly evolving benchmarks, which are highly regarded in the industry. SPECjAPPServer was introduced in April 2004 to measure the performance of Java 2 Enterprise Edition (J2EE) application servers, and provides a level playing field on which to test and compare the latest J2EE hardware and software platforms.

#### *Interpreting results*

The primary SPECjAppServer 2004 metric is JOPS, which is the average number of successful "jAppServer Operations Per Second" completed during the measurement interval. JOPS is composed of the total number of Business Transactions completed in the Dealer Domain, added to the total number of work orders completed in the Manufacturing Domain, normalized per second.

#### *Factors to Consider when Evaluating Results*

None at press time.



## Other Benchmarks

---

### Whetstone and Dhrystone

#### *Description*

The Whetstone benchmark was the first intentionally written to measure computer performance and was designed to simulate floating point numerical applications. The Whetstone benchmark was introduced in 1972. The Fortran version, introduced in 1973, became the first general purpose benchmark that set industry standards of performance.

The benchmark produced speed ratings in terms of Thousands of Whetstone Instructions per Second (KWIPS). In 1978, self timing versions produced speed ratings in MOPS (Millions of Operations per Second) and MFLOPS (Floating Point) and overall rating in MWIPS

The Dhrystone benchmark was designed to test performance factors important in non numeric systems programming (operating systems, compilers, word processors, etc.). Developed in 1984, Dhrystone tests a system's integer performance. The program is CPU bound, performing no I/O functions or operating system calls.

Original versions of the benchmark gave performance ratings in terms of Dhrystones per second. This was later changed to VAX MIPS by dividing Dhrystones per second by 1757, the DEC VAX 11/780 result.

#### *Why they're not relevant today*

These are dated benchmarks that are not appropriate to today's workloads. Variants of these benchmarks are available through SiSoftware CPU benchmark, in its Sandra product, a suite of diagnostic tests. This is a private company, not an industry consortium, so the tests are proprietary and the composition/details on the benchmark are not available to the public.

**NOTE:** In the words of Dr. Reinhold Weicker, the original creator of Dhrystone, "Although the Dhrystone benchmark that I published in 1984 was useful at the time, it cannot claim to be useful for modern workloads and CPUs because it is so short, it fits in on-chip caches, and fails to stress the memory system. Also, because it is so short and does not read from an input file, special compiler optimizations can benefit Dhrystone performance more than normal program performance." <sup>1</sup>

1. Embedded Microprocessor Benchmark Consortium, EEMBC Journal - Autumn 2002, [http://www.eng.mu.edu/corlissg/173.04F/on\\_eembc\\_benchmarking.html](http://www.eng.mu.edu/corlissg/173.04F/on_eembc_benchmarking.html)

# Dell Guide to Server Benchmarks



## Other Benchmarks

---

### NetBench and WebBench

<http://www.veritest.com/benchmarks/netbench/default.asp>

#### *Description*

NetBench is a benchmark program that measures how well a file server handles file I/O requests from 32-bit Windows clients, which pelt the server with requests for network file operations. Clients represent physical machines, while engines represent processes that execute multiple, simultaneous instances of the test suite on a physical client.

WebBench helps measure Web server software performance by running different Web server packages on the same server hardware, or by running a given Web server package on different hardware platforms.

#### *When It's Useful*

NetBench is a tool that stresses a networked server performing basic file & print services (like SPEC SFS). It is limited to evaluating systems with just one network interface card (NIC) for I/O. WebBench is a tool that stresses a web server. It is limited to evaluating systems that have a file set up to 200MB.

#### *Interpreting results*

- NetBench: NetBench reports throughput (in megabits per second – Mbps – the higher the better) and client response time measurements (the lower the better).
- WebBench: WebBench reports two overall scores for the server: requests per second and throughput as measured in bytes per second -- for both scores, the higher the better.

#### *Factors to Consider when Evaluating Results*

- Although Veritest is no longer supporting these benchmarks, Dell has found that they still provide a relevant comparison of performance, especially between generations of servers.
- The limitations of these tests mean that they may not fully stress the latest features of new technologies (i.e., dual core, hyperthreading)

# Dell Guide to Server Benchmarks



## Other Benchmarks

---

### LINPACK

<http://www.netlib.org/benchmark/hpl/>

#### *Description*

The LINPACK Benchmark is something that grew out of the LINPACK software project. It was originally intended to give users of the package a feeling for how long it would take to solve certain matrix problems. The benchmark started as an appendix to the LINPACK Users' Guide and has grown since the LINPACK User's Guide was published in 1979. LINPACK uses a numerical analysis technique called LU decomposition to solve a series of simultaneous equations. LINPACK measures a computer's floating-point rate of execution.

#### *How It's Useful*

The high performance LINPACK distributes the matrix problems between computers, enabling the performance assessment of clusters, specifically HPCC clusters.

#### *Interpreting results*

LINPACK results are presented in megaflops (MFLOPS), with a higher result indicating a faster performance.

#### *Factors to Consider when Evaluating Results*

- While LINPACK is currently useful in characterizing HPCC cluster performance, it is not as effective as a measure in non-clustered environments.
- LINPACK results are sensitive to cluster interconnects (Ethernet, Infiniband) and the bandwidth constraints associated with them

# Dell Guide to Server Benchmarks



## Other Benchmarks

---

### STREAM

<http://www.cs.virginia.edu/stream/>

#### *Description*

STREAM is a benchmark program that measures sustainable memory bandwidth in megabytes per second and is designed to highlight memory performance. It uses an HPC-like workload derived originally from computational fluid dynamics. It measures the computation rate and corresponding memory bandwidth for simple vector kernels, which are the building blocks of many different HPC workloads.

Since processors are getting faster much more quickly than computer memory systems, STREAM can help measure performance more accurately for users of applications that are memory intensive. As processors continue to improve their performance with multiple cores and higher clock speeds, more and more programs will be limited in performance by the memory bandwidth of the system, rather than by the computational performance of the CPU.

#### *How It's Useful*

The STREAM benchmark is specifically designed to work with datasets much larger than the available cache on any given system, so the results should be indicative of the performance of very large, vector style applications.

#### *Interpreting results*

STREAM consists of four tests that add independent information to the results:

- "Copy" measures transfer rates in the absence of arithmetic.
- "Scale" adds a simple arithmetic operation.
- "Sum" adds a third operand to allow multiple load/store ports on vector machines to be tested.
- "Triad" allows chained/overlapped/fused multiply/add operations

The results are presented as bandwidth in megabytes per second (MB/s), with a higher result indicating better performance.

#### *Factors to Consider when Evaluating Results*

None at press time.

# Dell Guide to Server Benchmarks



THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

Dell and PowerEdge are trademarks of Dell Inc. SPEC CPU2000, SPECint, SPECfp, SPECjAppServer, SPECjbb, SPECsfs, and SPECweb are registered trademarks of the Standard Performance Evaluation Corporation (SPEC). TPC-C, TPC-H, and TPC-App are trademarks of the Transaction Processing Council (TPC). Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

©Copyright 2006 Dell Inc. All rights reserved. Reproduction in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell.