An overview of the SPEC CPU2006 benchmark on HP ProLiant servers and server blades

Executive summary

This paper will explain the purpose of SPEC CPU2006, how the benchmark is conducted, and how to interpret the results. In general, SPEC CPU2006 represents rich compute intensive environments in mathematics and the sciences, with a small focus on business application workloads.

What SPEC CPU2006 measures

The SPEC CPU2006 benchmark is intended to stress the computer processor (CPU), the memory architecture, the compilers, and the chipset/front side bus. It is important to note that though most servers do extensive I/O, either through a NIC, a disk controller, or some other device, SPEC CPU2006 does not measure the performance of these pathways.

SPEC CPU2006 uses workloads that represent real-world problems of R&D environments or a highly specialized field such as weather forecasting. Some of the more common applications are ray tracing, data compression, video compression, PERL language interpretation and XML processing. SPEC CPU2006 results are useful in the following situations:

- To compare disparate systems (different processor platforms for example)
- To see the performance impact of a specific hardware component (i.e. performance delta between two different processor speeds in the same server)
- When the target environment has a CPU intensive, likely scientific workload
- When there are no other benchmark results available on the systems that you want to compare

The importance of SPEC CPU2006

There are several reasons why SPEC CPU2006 can be an important benchmark. Many other benchmarks stress I/O, operating system, and other system components other than just the CPU. The benchmarks that stress these other components are much more complicated to run and therefore have a more limited set of published results with which to compare.

MHz/GHz does not tell the whole performance story

For different platforms, processor speed has a different impact (i.e. AMD based systems have lower clock rate than comparable Intel-based processors). Other factors affecting performance include processor cache size, front-side bus/hypertransport speeds, and memory speed technology. SPEC CPU2006 results utilize all of these components to arrive at a set of performance metrics that represent overall system performance.

SPEC CPU2006 results are ubiquitous

SPEC CPU2006 results are available on a large number of systems and system configurations. There are three reasons why so many results are available:

The SPEC CPU2006 benchmark functions across different server platforms.

The SPEC CPU2006 workloads are supplied by the SPEC organization as C, C++ and FORTRAN source code. The suite has been ported to and runs on numerous platforms and operating systems.

The SPEC CPU2006 benchmark does not take many hardware resources to run.

SPEC CPU2006 is a self contained benchmark – it does not take a stack of "driver" computers to supply the workload (like TPC-C or SPECweb2005)

SPEC CPU2006 does very little I/O so benchmark sponsors do not have to supply options like disk drives, RAID controllers, NICs, network switches or JBODs.

The European Commission Directive 93/36/EEC.

The European Commission Directive 93/36/EEC disallows references to a specific brand of microprocessors in a public tender and Article 28 of the EC Treaty has been interpreted to disallow the specification of a clock rate as a measure of performance for a microprocessor. Due to these limitations, SPEC CPU2006 metrics have become a common way to specify system performance requirements for competitive IT procurement in many EU public tenders.

The SPEC CPU2006 workload suites and key calculations

INT and FP: the two benchmark suites in SPEC CPU2006

The following table lists the applications used in the two different suites of SPEC CPU2006, the INT suite and the FP suite. INT stands for integer-math based workloads, and FP stands for floating point-math based workloads. The INT suite consists of a number of workloads involving integer math supplied as C and C++ code. The FP suite consists of highly scientific workloads involving floating point math supplied as C, C++ and FORTRAN code.

Table 1. INT and FP comparison

<table>
<tr><th colspan="2">SPEC CPU2006 INT benchmark suite</th><th colspan="2">SPEC CPU2006 FP benchmark suite</th></tr>
<tr><td rowspan="3">Business applications</td><td>C, Perl programming language workloads</td><td rowspan="3">Chemistry and biology</td><td>Quantum chemistry</td></tr>
<tr><td>"bzip" file compression workload</td><td>Quantum chromodynamics</td></tr>
<tr><td>XML processing workload</td><td>Molecular dynamics</td></tr>
<tr><td rowspan="2">Scientific applications</td><td>Search gene sequencing</td><td rowspan="5">Physics</td><td>Fluid dynamics</td></tr>
<tr><td>Quantum computing</td><td>General relativity</td></tr>
<tr><td rowspan="7">Problem solving applications</td><td>Video compression</td><td>Structural mechanics</td></tr>
<tr><td>Combinatorial optimization</td><td>Electromagnetics</td></tr>
<tr><td>Discrete event simulation</td><td>Weather prediction</td></tr>
<tr><td rowspan="4">Artificial intelligence applications</td><td rowspan="4">Mathematics</td><td>Finite element analysis</td></tr>
<tr><td>Linear programming</td></tr>
<tr><td>Ray-tracing</td></tr>
<tr><td>Speech recognition</td></tr>
</table>

Geometric mean

To understand how SPEC CPU2006 works, it is helpful to have an understanding of geometric mean. SPEC CPU2006 results are a geometric mean of the run time of the workloads run in the benchmark, normalized to the performance of an older common server platform.

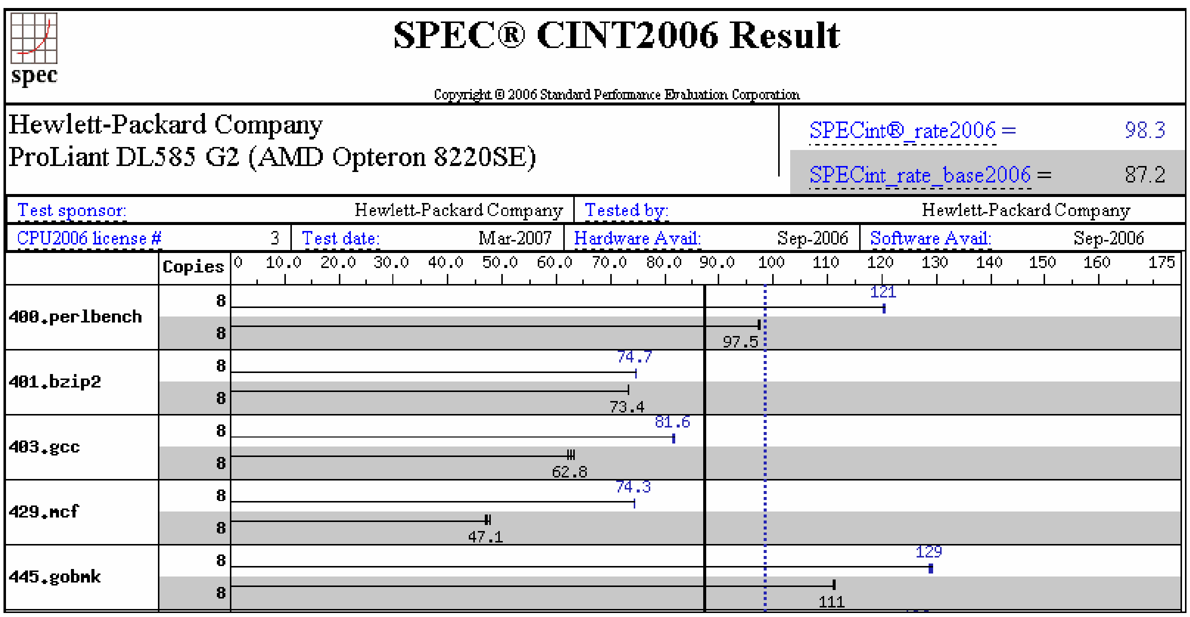
An **arithmetic** mean is relevant when several quantities **add** together to produce a total. In the same way, the **geometric** mean is relevant when several quantities **multiply** together to produce a product. The geometric mean is a complicated calculation devised to answer the question, "If all the quantities had the same value, what must that value be in order to achieve the same product?"

An example of this would be in calculating the average growth of compound interest. Suppose you have a value which grows 10% during the first period, 20% during the second period, and 30% during the third period. What is its average rate of return?

To calculate this answer (which happens to be 18.17%), you would use the geometric mean rather than the arithmetic mean since the value was multiplied by 1.10, 1.20 and 1.30 each year respectively rather than being added to.

By using the geometric mean, SPEC CPU2006 accurately accounts for the amount of time each individual workload takes to complete in the cumulative INT of FP suites.

Figure 1. Portion of sample HP ProLiant DL585 G2 SPEC CINT2006 result



Interpreting the results

The following is a list of results categories that can be obtained from running the SPEC CPU2006 benchmark (specific results viewable at www.spec.org):

- SPECint_2006
- SPECint_base2006
- SPECfp_2006
- SPECfp_base2006
- SPECint_rate2006
- SPECint_rate_base2006
- SPECfp_rate2006
- SPECfp_rate_base2006

FP vs. INT

FP stands for floating point-math based workloads, and INT stands for integer-math based workloads.

Base vs. peak

The ***base*** metrics (e.g. SPECint_base2006) are measured when the workload applications are compiled with "standard" compiler options. The base metrics are required for all reported SPEC CPU2006 results and represent a user who compiles applications by using a simple build process.

The ***peak*** metrics (e.g. SPECint2006) are optional and have less strict compilation requirements. These runs represent the user with the patience to tune and optimize the build process.

Speed vs. rate

The speed run measures how long it takes to run the benchmark workloads end to end. The rate run measures the number of tasks a system can run in a certain amount of time.

The SPEC speed metrics (e.g., SPECint2006) are used for comparing the ability of a computer to complete single tasks. This metric will measure the performance the system using a single core of the system processor. This metric is greatly affected by the clock speed of the processor and the processor cache size.

The SPEC rate metrics (e.g., SPECint_rate2006) measure the throughput or rate of a machine carrying out a number of simultaneous tasks. This metric provides a good overall measure of the performance of today's multi-core servers. SPEC rate metrics are typically most affected by the number of processor cores on a system and then by the clock speed and processor cache size.

Comparing SPEC CPU2006 results

SPEC CPU2006 results are available on the http://www.spec.org website. When comparing SPEC CPU2006 results, it is important to only compare like metrics. For example, it is useful to compare a SPECint_2006 score on one system with another SPECint_2006 score from another system; however it is not useful to compare a SPECint_2006 score with a SPECfp_rate_base2006 score. It is likewise not fair to compare a "peak" result from one system to a "base" result on another system. When comparing the overall performance of a multi-core system, use the SPEC rate metrics. When weighing single-threaded performance, consider the SPEC speed metrics.

HP ProLiant servers, server blades and SPEC CPU2006

Market leadership

HP ProLiant servers and server blades are a vital part of the HP success story. HP is the #1 vendor in worldwide server shipments. For the 21st consecutive quarter, more than 5 years, HP is the #1 vendor in worldwide server shipments. HP ships more than one out of every three servers in the world.[1]

Proven performance

Proven performance is part of the reason that HP is #1 in server shipments. HP has posted hundreds of SPEC CPU2006 benchmark results on ProLiant servers and server blades, helping customers to identify reasons to be confident in HP.

Appendix

CINT2006 Workloads	**CFP2006 Workloads**
CINT2006: C & C++ applications	**CFP2006: C, C++ & FORTRAN**
400.perlbench C PERL Programming Language	410.bwaves Fortran Fluid Dynamics
401.bzip2 C Compression	416.gamess Fortran Quantum Chemistry
403.gcc C C Compiler	433.milc C Physics: Quantum Chromodynamics
429.mcf C Combinatorial Optimization	434.zeusmp Fortran Physics/CFD
445.gobmk C Artificial Intelligence: go	435.gromacs C/Fortran Biochemistry/Molecular Dynamics
456.hmmer C Search Gene Sequence	436.cactusADM C/Fortran Physics/General Relativity
458.sjeng C Artificial Intelligence: chess	437.leslie3d Fortran Fluid Dynamics
462.libquantum C Physics: Quantum Computing	444.namd C++ Biology/Molecular Dynamics
464.h264ref C Video Compression	447.dealII C++ Finite Element Analysis
471.omnetpp C++ Discrete Event Simulation	450.soplex C++ Linear Programming, Optimization
473.astar C++ Path-finding Algorithms	453.povray C++ Image Ray-tracing
483.xalancbmk C++ XML Processing	454.calculix C/Fortran Structural Mechanics
	459.GemsFDTD Fortran Computational Electromagnetics
	465.tonto Fortran Quantum Chemistry
	470.lbm C Fluid Dynamics
	481.wrf C/Fortran Weather Prediction
	482.sphinx3 C Speech recognition

For more information

www.hp.com/servers/benchmarks

www.spec.org

[1] IDC, Worldwide Technical Server QView Q1 2007, August 2007.

October 2007