STANDARD PERFORMANCE EVALUATION CORPORATION

www.spec.org

March 18, 2009

To: Rebecca Duff - ICF International
Cc: Andrew Fanara - United States Environmental Protection Agency
Re: SPEC's Response to Draft 4 of ENERGY STAR Program Requirements for Computer Servers

Dear members of the EPA and ICF Consultants,

Twenty years ago, the computer industry formed the nonprofit Standard Performance Evaluation Corporation (SPEC) to establish industry standards for measuring compute performance. SPEC has since become the largest and most influential benchmark consortium. Its mission is to ensure that the marketplace has a fair and useful set of metrics to analyze the newest generation of IT Equipment.

SPEC welcomes this opportunity to review draft 4 of the ENERGY STAR Program Requirements for Computer Servers and is proudly looking forward to continuing our long-standing association with the EPA ENERGY STAR Product Development Team. The development of ENERGY STAR standards is an essential component in the ongoing effort to reduce worldwide energy consumption. We applaud the EPA for its goal to drive toward greater energy efficiency in IT Equipment, and SPEC considers the EPA ENERGY STAR Program as an industry partner in this effort.

The SPEC community has developed more than 30 industry-standard benchmarks for system performance evaluation in a variety of application areas and provided thousands of benchmark licenses to companies, resource centers, and educational institutions globally. Organizations using these benchmarks have published more than 10,000 peer-reviewed performance reports.

SPEC membership is open to any interested company or entity. Members and Associates are entitled to licensed copies of all released OSG benchmarks and unlimited publication of results on SPEC's public website. The initiation fee is 1000 USD and the annual dues are 6000 USD for members. Nonprofit organizations and educational institutions annual dues are 500 USD. Further membership information can be found on http://www.spec.org/osg/joining.html.

SPEC's current members, associates and supporting organization include: Acer Inc., Action S.A., Advanced Micro Devices, Apple Inc., Argonne National Laboratory, ASUSTeK Computer Inc., Austrian Energy Agency, Autodesk, Avere Systems, BEA Systems, BlueArc, Bull S.A., Center for Scientific Computing (CSC), Cisco Systems, Inc., Citrix Online, CommuniGate Systems, Dell, Department of Veteran's Affairs - Corporate Data Center Operations, Dresden University of Technology, E4 Computer Engineering SPA, EMC, EP Network Storage Performance Lab, FORMAT Sp. z o.o., Forschungszentrum Karlsruhe GmbH?, Fujitsu Limited, Fujitsu Siemens, Hitachi Data Systems, Hitachi Ltd., HP, Huawei Technologies, IBM, Indiana University, Intel, ION Computer Systems, Itautec S/A, JAIST, Leibniz Rechenzentrum - Germany, Lenovo, Microsoft, National University of Singapore, NEC - Japan, NetApp, Novell, NTT System, NVIDIA, Ohio State University, Oracle, Parallels, Penn State, Purdue University, Platform Computing Inc., Principled Technologies, QLogic Corporation, Rackable Systems, Red Hat, Results ByIQ, LLC, RWTH Aachen University, SAP AG, SGI, Sun Microsystems, Super Micro Computer, Inc., Symantec Corporation, Technical University of Darmstadt, The Portland Group, Tsinghua University, Unisys, University of Aizu - Japan, University of California - Berkeley, University of Cologne, University of Illinois - NCSA, University of Karlsruhe, University of Maryland, University of Miami, University of

Pavia, University of Texas at Austin, University of Tsukuba, Via Technologies, Virginia Polytechnic Institute and State University, VMware

The increasing demand for energy-efficient IT Equipment has resulted in the need for power and performance benchmarks and in response the SPEC community established SPECpower, an initiative to augment existing industry standard benchmarks with a power/energy measurement. The industry committed leading engineers and scientists in the fields of benchmark development and energy efficiency to tackle this task. The development of the first industry-standard benchmark that measures the power and performance characteristics of server-class compute equipment started on January 26th 2006. SPECpower_ssj2008 exercises the CPUs, caches, memory hierarchy and the scalability of shared memory processors on multiple load-levels. The benchmark runs on a wide variety of operating systems and hardware architectures. SPECpower_ssj2008 1.10 adds multi-node support (e.g. blade-support) and it will be released on April 8th 2009.

During a series of review discussions among the SPEC membership, we identified several shortcomings in draft 4 and compiled a list of these deficiencies and possible solutions. SPEC understands that Tier 1 will have a limited scope and several flaws but believes that with the following improvements it will be a step forward to educating it's target audience.

- SPEC continues to believe that performance is a necessary part of an energy-efficiency metric. Therefore the system idle-power-only criteria for enterprise servers will not effectively inform buyers to identify the most energy efficient Computer Servers in the marketplace. Also we need to point out that number of sockets, number of processor chips, and number of cores are all considered poor indicators of performance.
- It is also our opinion that comparing servers powered by AC against servers powered by DC is not fair, since the AC-DC conversion losses are not included in DC-powered server. We would recommend creating a separate category for DC-powered servers.
- Regarding the latest addition to the reporting requirements; the thermal flow rates of servers (Total Power Dissipation; Delta T at Exhaust at Peak Temperature; Airflow at Maximum Fan Speed at Peak Temperature; Airflow at Minimum Fan Speed at Peak Temperature). We believe that a unified reporting of the thermal flow rate across the industry could help facility managers to plan datacenters better with the additional thermal load of the servers in mind. Unfortunately, the volumetric flow depends on the configuration of the server and the EPA's concept of Product Family could not be used. Also we believe that the guideline described is not strict enough to ensure unified measurements and reporting across the industry and the additional cost for each partner needs to be considered as well. In order to qualify a server, a volumetric airflow bench (not a simple wind tunnel nor an anemometer) is necessary (Cost ~$5000) as well as a temperature chamber (Cost ~$5000). We estimate the time spent to produce these results is around 8 hours for each configuration. SPEC recommends to delete this requirement in order to enable smaller companies to become an ENERGY STAR partner as well.
- We would like to remind the EPA that different benchmark and stress tools will yield a different "Full Load (100%) Power". Therefore we would recommend to create a list of 2-5 acceptable benchmarks / stress tools; one might be the free SPECjvm2008 (http://www.spec.org/jvm2008/) benchmark.

The following comments are in order of to the line numbers of the draft 4:

- 163 Performance for Special Distinction
  - SPEC agrees and recommends additional recognition for voluntary measures.
  - We would recommend describing how this additional recognition would look like.
- 265 Editorial
  - From "designed to for technicians to easily add or replace multiple Computer Server boards in the field."
  - To "designed for technicians to easily add or replace multiple server blades in the field."
- 312 The expression "Capability" needs to be clarified.

- 313 The "management controller" needs to be defined.
- 335 Editorial
  - From “designed and built as a single enclosure and are not designed to for technicians to easily add or”
  - To “ designed and built as a single enclosure and are not designed for technicians to easily add or”
- 380 Does it need to be "removable" or "hard-wired"?
- 417 Editorial
  - From “InfiniBan” to “InfiniBand”
- 429 Product Family
  - A good idea, nevertheless a bad choice of words since EPA's definition of a Product Family is even more restrictive than product model.
- 466 It is a misconception that the maximum configuration yields the highest performance
- 476 The definition of "Typical" is not clear enough and let the specification open for manipulation
- 505 Multi-Node System, definition is missing in draft 4. Or is that suppose to be Dual-Node Servers?
- 604 Editorial
  - From “B. Active Power Requirements”
  - To “B. Idle Power Requirements”
  - In several lines (e.g. 612, 614, 622) the document switching between processor and socket
- 632 Editorial
  - In category C -> From “Dual” to “Single”
- 632 Table 3: Base System Idle Power Requirements
  - The Category B limits seem inconsistent and rather low when comparing A->B 10W and B->D 85W.
- 633 Table 4: Additional Power Allowances for Extra Components
  - The 8 Watt per drives allowance leaves the specification open for manipulation, a vendor can off-set high power consumption of system by using several additional SSD drives.
  - Also a definition of port and device is missing, since in general a device can have several ports.
  - A definition of active is necessary. Is a connection without traffic considered active?
  - Different Fiber Channel and InfiniBand components on the market, if not better defined the allowance can be misused in order to offset high server power consumption.
  - It might be a better approach for now to remove the additional allowance for IO cards altogether and state that an Energy Star qualified server will not change their status when additional IO card are installed.
- 797 Missing definition for BMC
- 874 Clarification needed:
  - “or other out-of-band technology shipped with the Computer Server”
  - Do the Temperature Sensor and Power Analyzer need to be part of the system or are add on devices allowed.
- 891 Processor Utilization
  - How does the EPA plan to verify accuracy?
  - How does the EPA define processor utilization for under-clocked processors and processors in different power states?
- 895 “Sampling Requirements: Hardware polling rates of the embedded sensors must meet a minimum of one sample per second. Data must be averaged on a rolling basis over a specific time period.”
  - Is it the really in the EPA's interest to requite the internal design of a server component?
- 931 1a.Tier2
  - SPEC is currently working on several additional power and performance benchmarks (various application areas) with release dates around mid and end 2009. We are confident that enough benchmarks will be available for Tier 2.
- 969 4.Tier2
  - IEEE 802.3az is not certified as a standard yet; therefore it seems unrealistic to assume that an implementation will be available on October 1, 2010.
- 1005 accredited, independent laboratory

    - It seems that is a huge burden (cost) on each partner, especially smaller companies. We recommend not to implement this requirement.
- 1096 Editorial
    - From “management features enabled by the operating system by default. Manufactures must also clearly”
    - To “management features enabled by the operating system by default. Manufacturers must also clearly”
- 1223 Ability to measure true power for all AC sources;
    - It probably should read "... true rms power .."
- 1225 Frequency response
    - What exactly is meant with "frequency response", it is generally not a part of a power analyzer specification. Do you mean the Current and Voltage Bandwidth?
- 1230 - It's quite unclear what is meant:
    - "… with a frequency of 1 second …" or less or more?
    - Is the "integrating time displayed" meant?
- 1224 An available current crest factor of 3 or more at its rated range value;
    - SPEC recommends the following as replacement
    - “Crest Factor – The analyzer must provide a current crest factor of a minimum value of 3. For Analyzers that do not specify the crest factor, the analyzer must be capable of measuring an amperage spike of at least 3 times the maximum amperage measured during any 1-second sample.”
- 1250 Accuracy
    - Wording suggestion for 1251-1256 since power analyzer vendors describe uncertainty in terms of "% of reading" or "% of range".
    - Measurements must be reported by the analyzer with an overall uncertainty of 1% or better for the ranges measured. Overall uncertainty means the sum of all specified analyzer uncertainties for the measurements made.
    - Examples:
        - An analyzer with a vendor-specified uncertainty of +/- 0.5% of reading +/- 4 digits, used in a test with a maximum wattage value of 200W, would have "overall" uncertainty of (((0.5%*200W)+0.4W)=1.4W/200W) or 0.7% at 200W.
        - An analyzer with a power range 20-400W, with a vendor-specified uncertainty of +/- 0.25% of range +/- 4 digits, used in a test with a maximum wattage value of 200W, would have "overall" uncertainty of (((0.25%*400W)+0.4W)=1.4W/200W) or 0.7% at 200W.
- 1266 Supply Voltage
    - In order to use a voltage within a 1% difference, an extra voltage source is needed. This will increase the cost for the partner, especially smaller companies.
    - SPEC recommends the adoptions of the following AC Line Voltage Source Standard. It covers data centers as well as households world wide and has been proven over the last year.
    - The AC Line Voltage Source needs to meet the following characteristics:
        - Frequency: (50Hz or 60Hz) ± 1%
        - Voltage: (100V, 110V, 120V, 208V, 220V, or 230V) ± 5%
- 1291 "Power consumption of the UUT shall be measured and tested from an external AC or DC source to the UUT."
    - For external DC sources the losses of the AC to DC conversion is not included; therefore, it shouldn't be compared with AC nor put in the same category.
- 1301 We do not understand the need to measure a half populated blade enclosure or even 1 blade server? Fully populated should be the only required configuration.
    - It might be easier for now to treat blades servers in the same way as 4 socket servers.
- 1326 Step 2 and Step 3 should be switched
- 1332 It seems that only a single power analyzer is allowed, it our belief that some blade configuration may require multiple power analyzers and therefore should be allowed.
- 1339-1340 Should be deleted

- 1341 How can a partner record the measurements? Manual only or is an automatic form of recording allowed as well?
- "1353 b. Only power management features that are shipped enabled may be enabled during testing, and all power management features used must be reported."
    - How does a partner handle the case where the OS has no default setting and only a selection of options?
- 1355 Clarification: Can the system be monitored via BMC as well?
- 1359 In cases where the spin-down drive is default, must the default be changes?
    - Also is booting from external storage allowed?
- 1375 "4. Switch on the Computer Server and begin recording elapsed time, starting either when the Computer Server is initially switched on, or immediately after completing any log in activity necessary to fully boot the system. Dual-Node Servers should be booted and logged on concurrently. Once logged in with the operating system fully loaded and ready, close any open windows so that the standard operational desktop screen or equivalent ready screen is displayed. Between 5 and 15 minutes after the initial boot or log in, set the analyzer to begin accumulating power values at an interval of greater than or equal to 1 reading per second. Accumulate power values for 5 additional minutes and record the average (arithmetic mean) value observed during 1383 that 5 minute period."
    - We believe this is too ambitious and should be simplified.
    - Why not just measure the energy for over 5 min and divide by 300 seconds?
    - Also: Is it allowed that a server go into a sleep state during the measurement?

A successful ENERGY STAR for Servers program has the potential to harmonize energy efficiency programs worldwide. Therefore SPEC is looking forward to the next revision and is confident that the EPA will include our comments in order to create a good, well-written, comprehensive ENERGY STAR Program specification that can be easily implemented by partners without huge investments.

By Order of the Board of Directors,

Klaus-Dieter Lange, Director, SPEC