October 16, 2009

To: Evan Haines
ICF International

CC: Andrew Fanara
United States Environmental Protection Agency

Re: Hewlett-Packard Response to Preliminary Draft 1 of the Tier 2 *ENERGY STAR®* *Program Requirements for Computer Servers*

From: Hewlett-Packard Company, Enterprise Servers and Networking Business Unit

This document may be published on the Energy Star website.

Hewlett-Packard (HP) has a long-standing association with the *ENERGY STAR®* program and HP welcomes this opportunity to participate as a valued stakeholder in the process of creating of the Tier 2 *ENERGY STAR® Program Requirements for Computer Servers* (hereafter in this feedback document called "Energy Star for Computer Servers"). Comments and issues are focused on Eligibility Criteria.

1. Energy Star Product Eligibility Criteria

The following is a compiled list of HP comments, referenced to the section numbers or line numbers listed in the document. Comments are broken into two general lists. The *Substantive Feedback* section lists substantial changes to the approach documented in the Final Draft, while the *Editorial Feedback* section details issues that are important to clarify the intent of the document or to eliminate errors.

1.1. Substantive Feedback

1.1.1. Definition Issues

1.1.1.1. Lines 268-273: The definition of *Resilient Server* needs some improvement.

a. One of the key features of resilient servers is that they can handle multiple faults without requiring a reboot. Double-chip-sparing technology is a unique firmware function that can permanently detect/correct an error in a DRAM by recognizing failure and then "erasing" DRAM bits from the Error Correcting Code (ECC) calculations. It also allows the ECC logic to correct for a second DRAM failure in the same ECC codeword. This greatly reduces the likelihood of a memory-induced crash compared to single-chip-sparing and is more cost and power effective than memory mirroring

on x86 servers. HP recommends that double-chip sparing technology be added to the definition of resilient servers.

b. A second practical factor for resilient servers is that many of them offer redundant buses along with bus retries. System buses are the interconnects between processors, memory, I/O, and on large servers, different parts of the server. Bus retries are when there is an error or timeout and the requestor resends the request. Today, resilient servers often have bus retries, and HP recommends that bus retries be part of the resilient server definition. HP also recommends that the definition notes that many of these servers also have redundant buses that may consume additional power.

c. In the RISC/EPIC market, a significant portion of the resilient server market, approximately 75% of the units shipped are less than 4 socket servers (72% of RISC/EPIC units in 2008 according to IDC). This is often because while larger systems are used for production deployments, entry class systems are often used for development environments, initial test environments, and system management consoles. In particular, since test and development usually are done on servers with similar features, functionality, processors, and redundancy features of the larger production systems, even these entry class resilient servers have additional functionality than standard volume (x86) servers. Therefore, we feel that the resilient server category should not be limited to 4 socket and greater systems but also include 1 and 2 socket servers.

d. Resilient servers often offer expansion characteristics that are much greater than standard volume servers. For instance, many of the mid-range and high-end servers in this category have a modular design that allows additional processors, memory, system boards, and I/O expansion boards to be added to the server. This implies that the underlying infrastructure needs to support workloads that may start out as a fraction of the server, but grow either temporarily or permanently to significantly larger sizes. As a reflection of this, many vendors offer some sort of permanent or temporary instant capacity on demand offering where customers can pay for the additional

capacity only if and when they require it. HP recommends that the resilient server category definition mention the fact that many of these servers initially ship with additional room for quick and easy expansion, since that often increases the overall power requirements of the system

e. Resilient servers as a general rule offer hot-swap components. This means that some components, such as power supplies, processors, cell boards, memory, I/O cards and more can be changed while the server is up and running. While this type of infrastructure does increase the reliability of the server, it also means that there is some level of redundancy that consumes power. This is one of the key reasons that resilient servers should not be in the same category as volume servers.

f. Resilient servers can be of either rack mount or blade server form factors. In fact, the concepts of cell boards, processor books, and dynamic partitions could be seen as predecessor to today's blade servers. HP believes that the resilient server category should specifically mention that resilient servers can be of either rack mount or blade server form factors.

g. Resilient servers are often used for a limited set of workloads. They include the broad categories of consolidation/virtualization, business processing and decision support. For instance, approximately 60% of all UNIX servers are used for business processing or decision support, versus 25% for Windows and Linux server (IDC workload study). HP recommends that the resilient server definition reflect that resilient servers are often used for these purposes. Often these workloads are not CPU bound; they tend to have a balanced demand across multiple resources such as processor, memory, and I/O. These broad demands mean that a benchmark that stresses one part of the system, or a mixed "decathlon" type of benchmark would not be a representative of how end users deploy these systems.

h. Resilient server utilization levels are usually higher than in the volume server category. For instance, in the late 1990's, and HP internal

study showed that average month utilization over a large number of mission critical servers, partially managed by HP for our customers, had an average utilization of approximately 30%. This is significantly higher than the sub-20% utilization that was typical on volume servers at the time, and in fact, has increased with the increased usage of consolidation and virtualization technologies. HP recommends that the resilient server definition reflect that resilient servers ordinarily have higher utilization levels than volume servers.

i. Resilient servers usually have longer development and deployment cycles than volume servers. Resilient servers are often designed and sold for many years, although there may be periodic upgrades to the processors, memory, and/or I/O infrastructure. For instance, HP Superdome servers have been sold based on the same chassis since 2000. In addition, once customers purchase a resilient server, it is often deployed for five or more years. One study suggested that the average UNIX server age is approximately 7 years old, a much longer deployment life cycle than most volume servers. In addition, end users of these systems are often more concerned about reliability, and therefore don't adopt new technologies until they are proven in the market place. HP recommends that the resilient server category reflect the longer development, adoption and deployment lifecycles. HP also recommends that the EPA recognize that since this category of servers doesn't change as often as volume servers, they are often not in a position to take advantage of leading edge technology, such as Energy Efficient Ethernet, as soon as it is released to the market.

1.1.1.2. Lines 269: The term **"**RAS" is not defined. Also "RAS" should not be treated as one feature. Reliability, Availability and Serviceability should each be defined, and several types of servers should be recognized for their value-add in each RAS area.

- ✓ **Warranty length should be an item in your data collection spreadsheet.** A server with a 3-year warranty, by design and configuration, has more reliability (and possibly more power) than a server with a 1-year warranty. 3-year warranties disallow use of certain components, and component choices may affect a server's active and idle power.
- ✓ Servers designed for "Availability" have several possible features that keep a server running despite imperfect power quality, environmental conditions, and component failures. The data collection spreadsheet should ask about the expected minimum downtime per year and also the length of time guaranteed for power outage ride-through.

- One example of availability enhancement would be a feature (like a battery or large capacitor) that enables a server to stay "on" despite an interruption in power. A measurement that could be quoted in the data collection spreadsheet is the minimum time period of power interruption that a server will survive. Even simple power supplies have product differentiation in how long they can "ride-through" power interruptions.
- "Serviceability" features may increase availability with hot pluggable features like fans, power supplies and hard drives. Supporting hot plug requires circuitry that needs power. The data collection spreadsheet should ask about hot plug fans and hot plug drives.

1.1.1.3. Line 313: power supplies will not always look exactly like these descriptions and will become more difficult to measure and impossible to compare to traditional power supplies. Energy Star needs to remove requirements for power supply efficiency, since power supply efficiency is measured as part of active and idle power measurements.

1.1.1.4. Lines 346-347: "Server Utilization" is not adequately defined. Utilization is typically measured by operating systems. Virtualization disconnects the utilization of physical server hardware from that measured by the operating system(s). In a Virtualized data center a "Server" is often a server OS instance and not a physical server.

1.1.2. Tier 2 Qualifying Products

1.1.2.1. HP requests that blade servers and multi-node servers be added included as qualifying products.

- ✓ Multi-node servers are no more difficult to qualify than dual-node servers and should have been included in Tier 1. Unlike the note on line 414, multi-node are not similar to fully fault tolerant servers in any way except being excluded from Tier 1. The sales volume of multi-node servers is increasing and will be deployed in large data centers.
- ✓ Blade servers need to be qualified in specific chassis enclosures. It would be possible for a blade to be compliant in some enclosures and not in others, since different enclosures can have different power supplies and power management features.
- ✓ Acceptable option 1 for defining blade server compliance is to take an approach like 4S servers in Tier 1. Require blade enclosures to support power management and reporting capabilities, require minimum power supply efficiency levels and then ignore the power levels. A blade server solution is going to use far less power than a comparable rack-mount solution, plus it often includes network switches and non-server blades.

- ✓ Acceptable option 2 for defining blade server compliance is to leverage rack mount server compliance test requirements and pass/fail criteria, and to qualify a blade server by filling a particular chassis/enclosure with enough homogeneous server blades that when divided by the number of blade servers the rack mount pass fail criteria is met.
- ✓ Regardless of option 1 or option 2, Energy Star server blades should be able to be shipped without an enclosure. Server blades must be able to retain their Energy Star status if they are plugged into a chassis or enclosure type in which the blade has been qualified. The chassis or enclosure should not be required to have the same set of blades, switches and routers used during compliance testing, but should have the same or better power supplies and power management features.

1.1.2.2. Support for including >4S servers is going to be contingent upon the availability of a suitable benchmark that adequately illustrates the performance and value of larger servers.

1.1.2.3. HP does not support the inclusion of fully fault tolerant servers and server appliances. Sales volumes are inconsequential and the market does not demand Energy Star for those products.

1.1.2.4. Support for including resilient servers (and >4S servers) is going to be contingent upon suitable accommodations for the reliability, availability and serviceability features that are enhanced on those platforms. HP believes that the list of qualified products should be expanded to include the resilient server category. While this is a smaller part of the overall market (~5% of units), it is still a significant part of the overall server spend (UNIX servers, a significant part of this category, are consistently 31%-32% of the annual server revenue according to IDC – 1996 to 2008).

The category of resilient servers often have additional redundancy features that consume power, operate at higher average levels of utilization (much less time spent at idle), offer additional expansion characteristics, process different workloads, and may be in different form factors such as rack mount or blade form factor chassis. Therefore, HP believes that a direct comparison of these servers with volume servers would often exclude these servers and limit the effectiveness of the Energy Star program to provide customers with a guide to which of these servers are the power efficient.

HP recommends that these servers have slightly different qualification requirements than volume servers. HP proposes the following requirements for resilient servers:

A. Power Supply Efficiency Requirements
   i. HP recommends that this requirement should line up with Climate Savers Power Supply Efficiency requirements for

simplicity. HP recommends that resilient servers should have at least a Climate Savers Gold level. HP agrees with the Green Grid that Net Power Loss is not the best way to evaluate power supply efficiency.

B. Active Power Requirements

i. HP recommends that over a longer time frame, some mechanism of active power measurement and reporting be created. However, this measure should reflect the fact that resilient servers often run different types of workloads and should not be the same measurement as volume servers. However, since developing an additional measurement tool with different weightings for various workloads, HP recommends that this requirement be dropped for resilient servers in Tier 2. Alternatively, HP recommends requiring reporting (but not power requirements) of power efficiency with at least one benchmark that is often used with this category of servers, such as SPECjAppServer2004 or TPC-H. Since resilient servers are often used for different workloads, HP recommends that the EPA select a number of possible benchmarks to report against in this category, and require the vendor to select and report on at least one benchmark.

ii. HP recommends that since resilient servers often have higher utilization levels than volume servers, and often spend less time in idle modes, that idle power requirement be dropped, although potentially still reported. This would be consistent with the implementation of 3 and 4 socket servers in the Tier 1 definition. The workloads that customers tend to run on these servers are usually workloads that keep the servers busy, so idle power is less important in this category.

iii. HP agrees that all the power management features as set forth in Table 5: Power Management Features should be required for resilient servers.

iv. HP recommends that since resilient servers have longer development life cycles and slower customer adoption of new technologies, Energy Efficient Ethernet should not be included as a requirement for this category (nor in any other Tier 2 category).

v. HP supports the EPA's requirement for standardized reporting. HP recommends that the reporting for resilient servers should be as consistent with volume servers as is feasible.

vi. HP recommends that data formatting on the various forms should be consistent between the forms. If data can be copied and pasted between the forms, it simplifies submission while providing the same information for the end user.

C. Data Measurement and Output Requirements

i. HP broadly agrees that the data measurement and output requirements as stated in the proposed Energy Star preliminary draft version 1.0 Tier 2 should apply to resilient servers.

In addition, HP recommends that the resilient server category should have a blade form factor sub-category. The resilient server blade form factor sub-category should have similar Power Supply Efficiency, configuration, standard information reporting requirements, and data measurement and output requirements as volume blade servers. However, the active power requirements for this blade based resilient servers should be the same as for rack mount resilient servers.

1.1.2.5. Blade Discussion Questions:

a. What efficiency/power/performance information do purchasers routinely request when investigating a new blade system? Do these requests change at all if the purchase is intended to replace standalone servers rather than other blades?
Manufacturers provide data showing that power per server is lower for comparable blade servers relative to rack-mount servers. At this time, there are no standard benchmarks that are used by customers to compare rack-mount servers to blade servers.

b. What efficiency/power/performance information would be useful to blade purchasers that is *not* routinely requested that could influence the provisioning process?
Blade customers tend to know what to ask.

c. What assumptions must typically be reported when marketing comparative blade performance and efficiency?

Using a CPU burn application like Prime95, companies have been marketing the relative power consumption of blade server solutions versus rack mount server "equivalents". This particular test is not a fair comparison, however, since there is no performance aspect in the test. There is some hope that SPECpower and other benchmarks will soon provide a fair means to compare blades and rack mount servers.

d. Regarding infrastructure overhead (power distribution/supply, cooling), how can the most efficient implementations be identified? What assumptions are fair/relevant to ensure fair comparison?

1. It is preferable that an enclosure <u>not</u> have its own Energy Star compliance process. We have suggested a means for qualifying blades in certain enclosures.

2. An easy and low cost approach to measuring the power of the enclosure/chassis infrastructure would be to measure the power with one server in an enclosure and then measure power with two servers installed. Then subtract twice the power difference to negate the server blade power and see the infrastructure power. Run a benchmark on both servers during the power measurements, so that there is nothing hibernating. The value of this measurement is dubious, however.

e. Given the server focus of this specification, what are ways that blade storage and network equipment could be addressed to create stable testing conditions between competing implementations?

1. Compliance testing of blade servers should not require the presence of storage and networking blades.

2. The presence of storage and networking blades in an as-shipped solution must not affect a blade server's Energy Star compliance.

f. Is analysis at the chassis level a valid approach to determining requirements for blades?

Testing the aggregated power and performance of a chassis with blades in it is the appropriate way to test compliance. However, the compliance test set-up should not have to be the same as the as-shipped or as-deployed solution.

g. Are there any anticipated purchasing practices when a user moves to a blade architecture (e.g. customers typically purchase blade chassis fully populated, half populated, etc.)?
It depends. If a customer buys one enclosure, then it is probably partially-populated, but if they buy many enclosures then many (but not all) are probably fully-populated. Partially-populated enclosures are more likely over time to be heterogeneously populated with dissimilar blades.
An enclosure is considered an option that is purchased with a blade server. This doesn't preclude a blade from being bought without an enclosure option and plugged into an existing enclosure. This illustrates how the blade and enclosure are sold and in HP's case (at this time) that the minimum population of blades in an enclosure is one...when shipped.

1.1.3. Efficiency Requirements for Qualifying Products

1.1.3.1. Line 449: HP is **not** supportive of the net power loss approach. Putting that test methodology into practice requires a supply chain to dramatically alter its test procedures and it forces them to test and certify power supplies at many times the number of load points. While it is nice to think of each power supply having an "efficiency curve," in reality the power supplies are not tested and guaranteed except at four load conditions. The rest of the efficiency curve is theoretical.
There is no way to accurately measure the DC power output of a power supply when it is installed in system. Every SKU has different power ranges, so net power loss would require hundreds of test to be run on every power supply.
Requiring NPL testing is a significant change and cost adder for our supply chain. It doesn't matter that in a few cases it is marginally better in predicting power supply efficiency. With performance/watt testing power supply efficiency becomes irrelevant.

1.1.3.2. Line 449: Requiring NPL testing is an additional and unnecessary cost for every Energy Star server.

1.1.3.3. Future improvements for power supply efficiency are reaching the point of diminishing returns. On server types that have performance per watt benchmarking pass/fail criteria, it makes no sense to require power supply efficiency requirements. With the benchmark, the server can be viewed as an opaque box. Removing power supply testing, once performance per watt

benchmarks are in place, supports the ability for this specification to support future changes to power supply and power deliver architectures.

1.1.3.4. Given the small data sets possible, we believe that servers with >2 sockets should continue with the current Energy Star compliance criteria. 4 socket server SKUs that are compliant with Tier 1 are rare in the industry, due to compliance difficulties on those platforms.

1.1.3.5. Lines 796-798. Tier 2 Requirements – HP believes that servers originally qualified under Tier 1 should remain qualified (grandfathered) for some period of time after Tier 2 takes effect, for instance for one year or until the server model is discontinued, whichever is less. This will encourage manufactures to qualify servers under Energy Star as soon as they are available, without the additional burden of having to re-qualify all existing qualified servers as soon as Tier 2 takes effect. Alternatively, HP recommends that the EPA allow servers to qualify under the Tier 2 rules in the 9 month period between the Tier 2 specification being finalized and the effective date, and that these servers automatically also qualify for immediate Energy Star Tier 1 certification. This would allow the manufacturers to test and submit their servers once during this period, for consumers to get the energy efficiency information, and eliminate the requirements for double testing and submission.

1.1.4. Test Criteria

1.1.4.1. In classes of servers where there are Tier 2 measurements for power and performance, there is no need to measure power supply efficiency.

1.1.4.2. HP believes that the EPA's approach to an active mode efficiency tool for general purpose/volume servers is a good approach, in particular if the tool is stable for a long enough periods that the vendors can learn how to test with the tool and that results are comparable to a large number of systems (e.g. In place with no changes for multiple years). HP believes that this will provide a significant amount of usable data for general purpose servers that customers may utilize for decision making purposes.

However, HP notes that not all servers in the market are general purpose/volume servers. As an example, resilient servers are primarily used for business processing and/or decision support. These systems are designed for higher levels of availability, but also designed to run these types of workloads efficiently. In other words, these types of servers excel in one area, but might not have a high overall score on a blend of workloads. The industry has recognized this over the years as demonstrated by the multiple benchmarks that approximate different types of workloads.

HP supports the EPA with the goal of providing much needed power consumption information for general purpose servers. However, HP believes that while the first focus should be on the general purpose servers, that the EPA should also consider expanding the coverage to more specialized servers. This could be in the form of allowing certain classes of servers to add power information (perhaps using the SPECpower power monitoring procedures) to a handful of other existing industry standard benchmarks, and allowing top performers in each category to qualify for Energy Star. Since vendors who create these more specialized servers likely already run some of these benchmarks, adding power information to them would be a small step, and in the end, provide customers with power and performance data on servers that otherwise would likely not have this information published.

1.1.5. Effective Date

1.1.5.1. October 15, 2010 is an acceptable release date for the Tier 2 specification, but there is inadequate time to test for compliance. There must be a 6-9 month window available for manufacturers to be able to continue to ship Energy Star Tier 1 products, so that supply chains can be managed and customer orders are not interrupted. The changes from Tier 1 to Tier 2 are dramatic and commerce will be interrupted if there is no grandfathering period.

1.1.6. Future Specification Revisions

1.1.6.1. The EPA should look into the future and decide at what point they should stop changing the Energy Star specification. Market forces have been improving energy efficiency very effectively and there comes a point where government intervention needs to stop "improving" the Energy Star specification and declare victory.

1.2. Editorial Feedback

1.2.1. pp1 line 4. Some confusion exists on why this is "Version 1.0 Tier 2" and not "Draft 1 of Tier 2" or perhaps "Version 2.0 Preliminary Draft 1" of the Energy Star specification. Saying "Version 1.0" looks like this is part of the specification release on May 15th. It is doubtful that there will be fractional revisions (1.1, 1.2, etc.) of the preliminary drafts for the version 2.0 specification.