IBM Comments to "ENERGY STAR® Program Requirements for Computer Servers Final Draft

IBM has reviewed the Final Draft document. Overall, EPA has addressed the majority of the concerns that we had with the Draft 4 requirements. IBM is supportive of the following changes made from Draft 4 to the Final Draft:

a. Classification of a product family based on the processor socket wattage and the available range of frequencies for that processor socket wattage. This better characterizes the systems;
b. Better definition of components under the "Product Family" definition, with expansion of the definition to acknowledge the fact that more than one manufacturer may supply a component under a specific manufacturer component part number;
c. Exclusion of standby power from the "additional output" 20 W limit;
d. Inclusion of SAN and RAID controllers in the I/O adder definition;
e. Revision to the definition of the maximum configuration to set the criteria as the configuration which fully utilized the available power to the server;
f. Exclusion of blade servers from this Final Draft. IBM has additional comments on a proposed approach to qualify blade servers in the specific comments.

These changes have served to make the requirements more workable. IBM appreciates EPA's efforts to engage and work with industry stakeholders to address the unique characteristics of server systems which have made the creation of the ENERGY STAR requirements a challenge.

There continue to be specific items on which we have commented previously which would benefit from modifications prior to implementation of the requirements. IBM offers the following specific comments.

Line 191: Under Definition of a Computer Server

IBM does not object to the requirement that a computer server include a hard drive. However, EPA should recognize that servers are being sold without hard drives with the capability to boot up and access storage through the Local Area Network. IBM recommends that EPA include a note or bullet in Definition V. Product Family (Line 341) which states that where a manufacturer sells a configuration that does not have a hard drive within a Product Family, that "hard driveless" configuration is qualified under ENERGY STAR® if the Product Family qualifies. A configuration without a harddrive should provide a more efficient option, as it is utilizing a shared storage resource which is likely to be inherently more efficient than a configuration having an operating hard drive on every server for some applications and workloads. The absence of a hard drive will result in a lower idle power.

Line 193: Under Definition of a Computer Server:

The statement: "All processors have access to shared system memory and are independently visible to a single OS or hypervisor" requires clarification. It is not always true; systems are available in which separate operating systems or hypervisors can access

a subset of the processors or cores on a system. It is recommended that you replace "single" with "one or more".

Line 441: Power Factor adjustment at 10% Loading on power supplies whose maximum power is >1000 W.

We reiterate our request to reduce the factor at 10% load for power supplies >1000 Watts to 0.75. A power factor of 0.8 at 10% load is technically difficult to achieve and to reliably measure. Looking at the PF limits for the 0 to 500 W and 500 W to 1000 W power supplies, both of which are set at .65, the jump to a PF of .8 for >1000 W systems is not justified. In addition, a server that utilizes a >1000 W power supply will not operate in the 5 to 15% power range for any meaningful period of time, making this criteria immaterial to the operation of the power supply. This is consistent with our previous comments in Drafts 2 and 3 and we continue to encourage EPA to either eliminate the 10% requirement or reduce the PF requirement to 0.75 for >1000 W power supplies loaded at 10% of capacity.

If appropriate, we would like to discuss this requirement with the appropriate EPA technical support person to understand and discuss the basis for this requirement and seek a workable compromise to this requirement. We believe this requirement is overly conservative and will preclude servers which would otherwise qualify for ENERGY STAR® from qualifying under the requirements.

Line 553: Standard Information Reporting Requirements: IBM will provide comments on the data sheet the week of May 11, 2009. This will give us time to test the data sheet using data from one of the IBM products.

Line 628: Power Supply Measurement Accuracy: IBM appreciates the changes that EPA made to recognize the variability in the accuracy of power supply measurements below the 100 watt (10% loading level). While the proposed change is effective for power supplies of 1000 W or less, it does not address the range of variability that occurs in large power supplies (>1000 W).

We propose the following language to address the fact that the accuracy of the power measurement on larger power supplies exceeds the 10 W limit at low loadings (<10% of the maximum load).

The proposal would provide specific accuracy requirements for power supplies with a maximum wattage of 1000 W or less and more than 1000 W;

Input power measurements for power supplies less then or equal to 1000 watts
-Measurements greater then 100 watts shall be +/-10%
-Measurements less then or equal to 100 watts shall be +/-10W
Input power measurements for power supplies greater then 1000 watts
-Measurements greater then 10% of max load shall be +/-10%

-Measurements less then or equal to 10% of full power supply rating shall be +/-1% of full range power

As an example, a 2000 watt power supply at 200 watts and below would have an accuracy measurement of +/-20 watts, and a 1000 watt power supply would be +/-10 watts.

A simpler option would be the following language:

Input Power Measurements:

- Measurements over the operating range of the server shall be +/-10%, where the operating range is defined as the power readings between the idle power and the maximum power of the server.
- Where the operating range includes input power measurements less than 100 watts, the accuracy will be +/-10 W below100 Watts.

The +/- 10W tolerance is just too tight to hold on power supplies with high wattage ranges. Planned high wattage power supplies that are currently in designs will not meet the accuracy requirements as detailed in the Final Draft. For large power supplies it stands, a 10% or 10 W limit below 10% of the maximum power supply output becomes less then +/-.5% of the full power supply range, which is not practical. The scaling is important, as the larger systems will not be running at 100watts or 10% of power supply range.

Line 610-611: We recommend that "non-proprietary" be removed from the sentence on data access so it reads: "…readable by third-party management systems." Our concern is that the data be readable by both proprietary and non-proprietary systems. Companies can choose to use either type of system and the data should be available to both.

Proposed, alternate language for utilization reporting:

Line 637: *Processor Utilization Measurements: The computer server will provide an estimation of the processor or system utilization that is visible to the operator or user of the computer server through the operating environment (operating system or hypervisor). The number is intended to provide the data center operator a qualitative indication of the amount of load on the system to provide guidance regarding opportunities for virtualization and consolidation of workloads to deliver more work for each unit of energy used.*

IBM proposes that the utilization requirement be qualitative in nature for the following reasons:

1. Each processor, system and operating system or hypervisor use slightly different technical techniques to estimate or quantify utilization, making the requirement of an

absolute or comparative number problematic given that the requirements are scheduled to go into effect on May 15, 2009.
2. With the introduction of power management functions and hardware multi-threaded systems, the current processor utilization measurement algorithms or functions cannot fully compensate for the presence of more than one thread or reductions in processor frequency. This in turn introduces significant inaccuracies into the measurement which make it unreasonable to establish a quantitative requirement with prescribed levels of accuracy. Each hardware and system software supplier has their own approach to resolving this for their customers. No single formula can represent all implementations.
3. Any option to provide a defined algorithm, as was done in the current draft to improve reporting accuracy will require a coding fix to implement. Given the variability in processor manufacturers, operating environments, and operating systems (OS versus hypervisor and proprietary source versus open source) there will be different timings for generating solutions across the range of systems which may disadvantage specific manufacturers and delay their ability to conform to the ENERGY STAR® computer server requirements. As this requirement is not one of the primary means by which ENERGY STAR® is encouraging the purchase of computer servers with more efficient power supplies and lower idle power demand, it should not be the one factor that would preclude a manufacturer from qualifying a computer server to the ENERGY STAR® requirements.
4. For the purpose of the Tier 1 ENERGY STAR® computer server requirements, the measurement only needs to be sufficiently accurate for the purposes of enabling decisions around the consolidation or reprovisioning of workloads or the migration of virtual machines. Existing CPU utilization measurement algorithms are already sufficiently accurate for this purpose. Many data centers today already rely on CPU utilization numbers as reported by currently shipping operating systems on currently shipping servers. They use this information successfully today to dynamically manage data center power consumption by reprovisioning workloads to minimize the underutilization of machines. Requiring any particular accuracy criterion for CPU utilization will not provide any particular incremental value to customers and will only increase the cost of the system due to the expensive additional micro-instrumentation required.

As an example of how the current utilization estimate can be used, a data center operations team at a specific facility tracked system utilization. In 2006, over 50% of the 1 and 2 processor servers in the data center were utilized 5% or less of the time and an additional 17% were utilized 5-10% of the time. By 2008, the percentage of 1 and 2 processor machines operating in these utilization ranges were reduced to 13% and 6% respectively. Much of the workload was moved to larger, more heavily utilized servers. Providing a qualitative measure of server utilization provides a data center operator with the information needed to identify these opportunities.

This language offers a sensible way to require the reporting of computer server utilization recognizing that there is no viable methodology to create a repeatable, quantitative method to measure utilization.

IBM Comments to "ENERGY STAR® Program Requirements for Computer Servers Final Draft

Line 669: Tier 2 Requirements

IBM recommends that the Target Effective Date for Tier 2 be extended until 12/31/2010 and that EPA maintain flexibility in setting the date for the final Tier 2 Requirements publication date. IBM is working internally and through The Green Grid to develop a proposal for a performance/power "synthetic workload" which will stress the key attributes of a server system. While IBM intends to work with the EPA and Green Grid to accelerate development, testing and data collection of the chosen system, the work will be difficult and subject to technical challenges. Under the proposed October 1, 2010 deadline, there may not be sufficient time to develop statistically valid data by the time the Tier 2 draft will need to be developed to settle on the appropriate 25% line for these benchmarks.

IBM is concerned that EPA appears to be dictating the outcome of the Tier 2 specification in the absence of an agreement on a criteria based on performance and power metrics. Given the range of discussion involved in the Tier 1 specification and the many changes that were made between the first and fourth drafts, changes that on the whole contributed to a more workable, robust set of criteria, it is inappropriate for EPA to attempt to specify an outcome to the discussion which will take place over the next 18 to 21 months. While we are not going to provide a detailed response to the items 1(b) through 4 of the Tier 2 proposal, we will note that in our comments to Draft 2 and 3 we have objected to the implementation of idle criteria for all servers, particularly more complex models, and to the Net Power Loss approach for power supply requirements. In both cases, we proposed workable alternatives which we believe provide a workable methodology to further EPA's and the industry's goal of providing more efficient IT equipment enabling improved energy utilization and efficiency in the data center.

Line 922: Accuracy requirements for the Power Meter used to measure server power specify a meter accuracy of .01 W at 10 W or less. This level of accuracy requires the use of a very expensive meter and we do not anticipate doing any idle measurements at less than 10 W. We ask that the requirement for a meter accuracy of .01 W below 10 W be removed from the requirements.

If EPA believes it is important to specify that level of accuracy at 10 W or less, we ask that you provide clarification that if no measurement is taken below 10 W then the measuring meter does not have to meet the requirement for .01 W accuracy below 10 W.

Line 410: Exclusion of Blades: IBM continues to be concerned that EPA intends to attempt to set idle power limits at various levels of blade population in a chassis. We provided extensive comments in draft 4, pages 4 to 7, on the difficulty of this approach. I reference those comments here, but will not repeat them. The variation in blade form factors, chassis are available with 7, 10, 14, and 24 or 25 blade slots, and the different disposition of fans, I/O, power supplies and other peripherals across the chassis and between the blade and the chassis, make it almost impossible to attain any meaningful

idle power qualification metric for blade systems. We propose the following potential solution to a Tier 1 blade system qualification criteria:

1. Require that processor level power management be enabled on the blade.

2. Require reporting of the following idle power and maximum power use information for the blade system in the product data sheet:
a. The idle power consumed by a blade chassis which is at least 70% populated with blades. This level of population of blades is most representative of conditions in a data center;
b. The idle and maximum power consumed by a single blade when tested in a test mount;
c. The idle and maximum power consumed by the chassis with any fans running at the low speed set point, a specified number of I/O devices engaged, etc. (this will need some additional clarification);

The power use information supplied as described in item (2) will allow the customer to ascertain the power use for a typical installation and adjust that power use to represent their intended configuration. They can also use the power configurator available from most manufacturers to perform the power use determinations. Requiring enablement of processor level power management and the reporting of the above described test data will drive increased energy utilization and efficiency in the data center. I am working with the technical team to get you a specific proposal for testing blade centers. We will try to do that over the next two weeks.

While IBM agrees that blade systems can compete directly with one and two processor socket systems, they also compete with 4 processor systems and they compete with 1 and 2 processor socket systems in the same way that a 4 processor socket systems competes with 1 and 2 processor socket systems. A data center operator has to analyze the benefits and difficulties in virtualizing the operations in blade or 4 socket systems against the operational, space and energy savings that can be achieved through the shared resources and expandability that can be achieved with a blade system. Rack servers are the better energy solution up to 4 servers vs. 4 blades, but blades become more economical and energy efficient above 4 servers.

Data Sheet Comments:

As a general comment, the use of a spreadsheet as a data reporting and management vehicle does not represent most efficient and error free approach to gathering and disseminating information on ENERGY STAR® products. It would be to EPA's benefit to develop, or have developed, a web-based tool for reporting and managing the data. This would benefit not only the Computer Server requirements, but the whole ENERYG STAR® products program.

1. Include a line reporting the make and model of the power meter used in the test.
2. Labels in Data Sheet are not consistent between the "Power Data" table and "Power and Performance Data for Benchmark #1" table. For instance, the idle power

calculated in step 8 of the procedure would (probably) be entered in "Measured Idle Power" for the "Power Data" table but under "Idle Power" under the other table. Same for "Power at Full Load." The terms in the data sheet should be made consistent from table to table.

3. In the Power and Performance table, it is recommended the "Power and Performance Score" item be relabeled as "Benchmark Score at Full Load Power" to be clear about the expectations.
4. Lines 16 and 17: Replace '"standard loading" with "specified loadings (10%, 20%, 50% & 100%)"
5. The information in lines 33, 34, 35, 36 is redundant with data requirements for lines 42, 43, 44, and 47. It is recommended that you remove lines 34, 35, 36, and 43.
6. The table for "Power and Performance Benchmark #2" should be labeled optional.
7. "Power Savings Features" Table: It would be appropriate to add three or four lines to the table for additional power saving features.
8. Line 79: Utilization Accuracy should be removed from the data sheet per our earlier comment.
9. Thermal Information Table:
   - The requirement for reporting Delta T across the server should be removed. It is dependent on cooling air flow and does not provide any meaningful information to the customer with regards to the efficiency or the set-up of the server in the data center.
   - The correct reference for the ASHRAE reports are: "ASHRAE Extended Environmental Envelope Final August 1, 2008" and "Thermal Guidelines for Data Processing Environments", ASHRAE, 2004, ISBN 1-931862-43-5".
   - Power dissipation should not be reported in the Thermal Information table, as it is already reported in the "maximum power data" on the "Power Data" Table. It should be the same information.
   - Airflow should be reported at the nominal fan speed and for maximum fan speed at 35 C as required by ASHRAE.
   - Currently, there are inconsistencies in the way that this data is reported across the industry. It is recommended that EPA work with ASHRAE and the Green Grid to establish standard, accepted guidelines for reporting this data.

This is a sample of an ASHRAE template for the IBM p520 server.

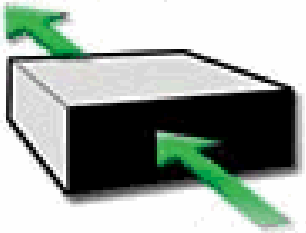
IBM Server Model 520 - Rack Mounted Drawer

Configuration	Condition								
	Typical Heat Release (Voltage 110 V)	Airflow				Weight		Overall System Dimensions[b] (W x D x H)	
		Nominal[a]		Maximum at 35°C					
	watts	cfm	(m^3/h)	cfm	(m^3/h)	lbs	kg	in.	mm
Minimum	420	28	44	40	68	117	53	25 x 37 x 23	630 x 933 x 584
Full	600	39	61	45	76	117	53	25 x 37 x 23	630 x 933 x 584
Typical	450	28	44	40	68	117	53	25 x 37 x 23	630 x 933 x 584

ASHRAE Class

3

Airflow Diagram Rack-mount Cooling scheme F-R



Front to Rear F-R

Configuration

Description	Model
Minimum	1-way, 1.5 GHz processor, 16 GB memory
Full	2-way, 1.65 GHz processor, maximum memory
Typical	1-way, 1.65 GHz processor, 16 GB memory

a. The airflow values are for an air density of 1.2 kg/m^3 (0.075 lbm/ft^3). This corresponds to air at 20°C (68°F), 101.3 kPA (14.7 psia), & 50% relative humidity.
b. Footprint does not include service clearance or cable management, which is zero on the sides, 40 in. (1168 mm) in front, & 40 in. (1016 mm) in the rear.

© 2004, ASHRAE Thermal Guidelines For Data Processing Environments adapted by DLB Associates, Consulting Engineers, P.C.

Appendix A: Test Procedure

The test procedure should include a "data reporting" section which specifically details where the data should be reported on the ENERGY STAR® Power and Performance Data Sheet and the Qualified Product Information Form.

Thank you for your consideration of our comments.