ENERGY STAR® Data Center Storage
Supplemental Data Set Development
February 2011

Objectives

As part of the ongoing Version 1.0 ENERGY STAR Data Center Storage specification development process[1], the EPA is interested in augmenting its current data set with the results of a second round of energy performance tests on data center storage systems using the protocol specified in this document.

EPA initially built its storage data set in early 2010 with substantial help from stakeholders. A total of 17 unique storage product platforms were tested, some several times, for a total of 42 data points. The number of data points for each segment of the SNIA Storage Taxonomy is illustrated in Table 1:

Table 1: Details of Current ENERGY STAR Data Set

# Systems (# Configs)	**Online**	**Near-online**	**Removable Media Library**	**Virtual Media Library**
Group 2	--	--	**1 (2)**	2 (2)
Group 3	**5 (7)**	--	**2 (10)**	--
Group 4	**5 (19)**	**1 (1)**	**1 (1)**	--

In order to most effectively supplement this existing data set, EPA wishes to consider data for systems in the taxonomy segments highlighted in yellow in Table 1 (Online 2/3/4 and Removable Media Library 2/3/4). EPA is also interested in suggested refinements to the test method.

The goal of this effort is to further EPA's understanding of the relationship between hardware/software configuration and energy performance in both active and idle states. EPA is specifically interested in the effects of single-variable configuration changes on power consumption. Items such as Hard Disk Drive (HDD) selection (e.g., capacity vs. performance), Reliability-Availability-Serviceability (RAS) features (e.g., single vs. redundant controllers, RAID level), and use of Small Form Factor (SFF) and Solid State Disk (SSD) drive technologies are all of interest. EPA has provided additional guidance on specific areas of focus for this supplemental data set construction process in the "Priorities for Supplemental Data" section of this document.

[1] A chronological history of ENERGY STAR Data Center Storage specification development is available at: http://www.energystar.gov/index.cfm?c=new_specs.enterprise_storage

EPA is appreciative of all stakeholder contributions to date, understands the continued limitations on time, personnel, and hardware resources for this effort, and appreciates that all responses are subject to the goodwill and best efforts of program participants.

Additional data will be anonymized and posted to the ENERGY STAR website in preparation for a subsequent draft of the Version 1.0 ENERGY STAR Data Center Storage specification.

Preliminary Conclusions from Current Data

Following an assessment of the existing data set[2] and subsequent conversations with individual vendors and industry organizations, EPA has arrived at several preliminary hypotheses about storage system energy performance:

- HDD selection (type and quantity) for Online systems is the most significant factor in active energy performance.
- Energy performance results are highly variable within and across taxonomy categories. Various systems and system configurations exhibit individual strengths during active and idle assessments. There is no clear energy performance leader, either within a taxonomy category or across several categories.
- There is no obvious difference in energy performance between Online Group 3 and Online Group 4. Group 3 systems were expected to exhibit better active energy performance than comparable Group 4 systems, given the hardware infrastructure required to support additional scalability in Group 4. The same also appears true for Removable Media Library Groups 2, 3, and 4, though this observation is based on a smaller sample size.

These and future observations will inform the development of a Product Family structure for use in the ENERGY STAR product specification. One goal of the Product Family structure is to reduce total test burden by allowing a group of highly-configurable products to be evaluated for qualification based on test results from one or more representative systems. Ultimately, tested system configurations, the associated test results, and any extrapolations should be representative of the entire range of system performance and provide meaningful insights to purchasers and end-users, in line with the objectives of the ENERGY STAR program.

EPA is currently evaluating several Product Family approaches and will continue to refine its perspective as new test data is received. Two approaches have thus far been proposed:

- Book-ending: Where the 'smallest' and the 'largest' system configuration (based on one or more key variables) are tested as boundaries for ENERGY STAR qualification. All system configurations within the bookends would also be eligible for qualification subject to specific limitations established in the product specification.
- N and *x*N: Where a given system configuration (N, based on one or more key variables) and some multiple of that size (e.g., 2N or 4N) are tested for ENERGY STAR qualification. 'Smaller' and 'larger' system configurations would also be eligible for qualification subject to specific limitations established in the product specification.

[2] http://www.energystar.gov/ia/partners/prod_development/new_specs/downloads/storage/StorageRawDataSet.xls

Priorities for Supplemental Data

To further substantiate the preliminary hypotheses above, EPA has identified the following high-priority topics to be investigated via the supplemental data set. Supplemental data may include new or previous test data or simulated data, as described below.

- Variation Across Taxonomy Categories:
  - Assess performance of Online Group 2, Removable Media Library Group 2, or Removable Media Library Group 4 systems, for which no data has been submitted to date.
  - Substantiate apparent energy performance differences between Online Group 3 and Online Group 4 systems.
  - Substantiate apparent energy performance differences between Removable Media Library Group 2, 3, and 4 systems.
- Effect of Drive Quantity: Determine effects of changes to HDD quantity versus a fixed number of controllers on an otherwise equivalent system. Preferably vary drive quantity in multiples of 2 (e.g., test with N, 2N, and 4N, 8N drives).
- Effect of Drive Technology: Determine impact of drive speed (e.g., 7200 vs. 15k RPM), Small Form Factor (SFF) & Solid State Drives (SSD) versus traditional drive technologies on an otherwise equivalent system.
- Effect of RAS Features: Determine impact of RAS features vs. lack of RAS features on an otherwise equivalent system.
- Isolation of Controller vs. Drawer PSUs: Wherever possible, measure loading of individual power supplies (or groups of power supplies) within a system.

Simulation Option: EPA will consider simulated (modeled) data in addition to actual test data, for use in the ENERGY STAR data set, per the following:

- Simulators must be capable of predicting variations in energy consumption for the various phases of a test sequence, including allowances for different workload stimuli.
- All simulations should be run using the test workload sequences defined for the appropriate taxonomy category.
- In order to assess simulator accuracy, EPA is interested in reviewing simulator data corresponding to each tested system configuration. Further simulations can then be used to assess the impact of an additional single- and multi-variable hardware and software configuration changes.

Test Setup

Environment: Testing may be conducted at normal laboratory ambient temperature and humidity. No special environmental controls are necessary. Temperature and humidity must be recorded at the beginning of each major test sequence.

Power Meter: The power meter shall be capable of measuring and recording UUT input power with an accuracy of 1% and a sampling frequency of no more than 5 seconds.

Power Measurement: UUT input voltage and power should be measured at a location appropriate to capture the total power consumed by all components of the UUT. This may be at the PDU, or some other appropriate location. The power measurement should include all items needed to provide for the integration and operation of the UUT. This includes controllers, drawers, robotic assemblies, power distribution / PDUs as well as data networking used internal to the UUT (e.g., integrated SAN switches).

Input Power: It is anticipated that mains power typical of a normal customer installation will be used during this exercise. Input voltage shall remain consistent to ±5% for the duration of the test period. The power supplied to the UUT shall be consistent with one of the following options:

Table 2: Input Power Requirements

Input Voltage Range	Phases	Input Frequency Range
100-120 VAC RMS	1	47-63 Hz
180-240 VAC RMS	3	47-63 Hz
200-240 VAC RMS	1	47-63 Hz
380-508 VAC RMS	3	47-63 Hz

Test Performance

Preconditioning: The preconditioning phase shall be of sufficient duration to ensure that I/Os are going to the storage media under steady-state conditions and are not being serviced by various system caches or other transients such as uninitialized space. For hybrid systems, preconditioning must be sufficient to allow the system to enter stable operation across the various storage media.

Sustained Duration: All test phases must be of sufficient duration to achieve steady-state system performance and mitigate the impact of system cache. Once stability has been achieved, the measurement period begins and shall continue for the "Sustained Duration" specified in the test sequence.

Slack Time: All steps in the test procedure shall be performed in sequence, with no more than 60 seconds delay between steps.

Data Distribution: For purposes of this exercise, testing should be distributed as evenly as possible across all storage media installed in the UUT[3]. Best practices suggest that test stimuli should be configured to exercise at least 80% of the available logical address space. Short-stroking and other techniques to artificially enhance the energy performance of the UUT are not permitted.

Measurement Reference: The point of reference for read measurements is the receipt of data by the host that initiated a read operation. The point of reference for write measurements is the receipt of an acknowledgment of successful data write by the host that initiated a write operation.

Data Compression & Reduction: For purposes of this exercise, all tests should be performed with uncompressed data. As a baseline, any data reduction features (e.g., compression, deduplication, thin provisioning) should be disabled. Data reduction features may be activated as single-variable changes for testing, and should be documented accordingly, along with details about the data set utilized during the test.

Block I/O vs. File I/O: For purposes of this exercise, it is preferable to test systems using a Block I/O interface. Systems that also support File I/O may optionally be tested using a File I/O interface as resources allow. Systems that support only File I/O may use an external means to convert File I/O requests into Block I/O (e.g., via a virtualized filing system), or test only with File I/O.

[3] This suggestion does not apply to Removable Media Libraries, which can access only a limited quantity of storage media at one time.

Battery Backup: For purposes of this exercise, all batteries internal to the UUT shall be fully charged and in a maintenance or 'float' state at the start and for the duration of testing. Battery configurations should be disclosed on the test data entry sheet.

Test Procedure

ONLINE SYSTEMS

The following test sequence shall be applied to systems in the Online taxonomy category.

Table 3: Test Sequence for Online

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration
Pre-Conditioning	Random 70% Read 30% Write	100%	8 KiB	≥ 30 min
Active "A"	Random Read	100%	8 KiB	10 min
Active "B"	Random Write	100%	8 KiB	10 min
Active "C"	Sequential Read	100%	256 KiB	10 min
Active "D"	Sequential Write	100%	256 KiB	10 min
Active "E"	Random 70% Read 30% Write	25%	8 KiB	10 min
Active "F"	Random 70% Read 30% Write	75%	8 KiB	10 min
Active "G"	Random 70% Read 30% Write	100%	8 KiB	10 min
Ready Idle	n/a	0%	n/a	30 min
Deep Idle	n/a	0%	n/a	10 min

Notes:

- Response Time for Online systems during Active test phases shall not exceed an average of 30 ms, and no single response shall exceed the MaxTTD defined for the system taxonomy category.
- The "Deep Idle" test phase may be omitted for systems which do not offer a Deep Idle feature. If Deep Idle data is collected, sufficient details about the system state (e.g., which subsystems are powered down) should be provided on the test data entry form.

REMOVABLE MEDIA LIBRARIES

The following test sequence shall be applied to systems in the Removable Media Library taxonomy category.

Table 4: Test Sequence for Removable Media Libraries

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration
Pre-Conditioning	Sequential Write → Rewind → Read	≥ 80%	128 KiB	10 min
Active "A"	Sequential Write ("x" drives)	≥ 80%	128 KiB	10 min
Active "B"	Sequential Write ("x + n" drives)	≥ 80%	128 KiB	10 min
Ready Idle	n/a	0%	n/a	30 min
Deep Idle	n/a	0%	n/a	10 min

Notes:

- The "Active A" test phase shall be performed with an arbitrary number of active drives, with robotics energized, idle, and ready to initiate a command. The "Active B" test phase shall be performed with a greater number of active drives than were used for "Active A," under otherwise similar conditions.
- All "Active" test phases shall be performed as near to 100% of maximum sustainable performance as possible, but no less than 80%. Test results should be extrapolated out to 100%, and details of these calculations shall be supplied on the test data entry form. Any start/stop events should also be noted.
- The "Robotics" test phase shall be performed using an average-distance movement from a single robot. Accumulated energy for the entirety of the movement sequence shall be recorded. The sequence is as follows:
    1. Start and with robot in a neutral/central ready position
    2. Retrieve media from rack and load in an empty slot
    3. Unload media and return it to original position on rack
    4. Return robot to neutral/central ready position
- The "Ready Idle" test phase shall be performed with robotics energized, idle, and ready to initiate a command.
- The "Deep Idle" test phase may be omitted for systems which do not offer a Deep Idle feature. If Deep Idle data is collected, sufficient details about the system state (e.g., which subsystems are powered down) should be provided on the test data entry form.