ENERGY STAR® Data Center Storage
DRAFT Initial Data Collection Procedure
November 2009

Introduction

As part of the ENERGY STAR Data Center Storage specification development process, the EPA is inviting interested parties to perform a series of tests and performance modeling exercises on data center storage systems using the protocol specified in this document.

The objective of this first round of data center storage power consumption testing is to understand the relationship between hardware and software configuration and system energy performance in both Active and Idle states. EPA would like to collect a substantial amount of test data and simulated (modeled) test results in order to perform a sensitivity analysis on the effect 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) technologies are all of interest.

Data submissions will be collated, anonymized, and made available to the stakeholder community for analysis in preparation for the first draft Version 1.0 ENERGY STAR Data Center Storage specification.

Definitions & Acronyms

Maximum Time to Data (MaxTTD): The maximum time before an entire data object is accessible within the constraints imposed by its storage media. For random-access media, a data object is accessible when any byte may be accessed. For sequential-access media, a data object is accessible when the requested object has begun streaming from a previously inactive drive.

Maximum Sustainable Performance: Maximum IOPS or GB/s that the UUT is able to deliver under a specified workload. To be considered sustainable, the performance metric must be maintained within 5% for the duration of the test phase.

Response Time: The time required for the UUT to complete an I/O request.

Generic System Configurations (GSC): Four GSC options are proposed as a baseline for discussion:

- **GSC-1:** Transaction-oriented Value Configuration
- **GSC-2:** Transaction-oriented High-reliability Configuration
- **GSC-3:** Archival/Streaming-oriented Value Configuration

GSC-4: Archival/Streaming-oriented High-reliability Configuration

Note that the specifics of each configuration will be up to the individual vendor or test facility and dependent on product options, availability of components, test equipment, and other resources. The data collection process will be sufficiently detailed to capture hardware and software configuration details that are pertinent to energy consumption.

Within many of the taxonomy cells, it will be possible to configure a system that is representative of a typical high-volume customer installation for each of the GSC classifications. For example, in the Online-3 taxonomy category, a GSC-1 system would likely utilize high-speed, high-performance HDDs or SSDs with minimal RAS hardware and software features installed; whereas a GSC-2 system may include additional RAS features such as redundant controllers, mirroring, and other RAS features. There are other taxonomy categories for which certain GSC classifications would not be representative of actual customer implementations (e.g. the Removable Media Library taxonomy category does not lend itself to transaction-oriented GSC-1 and GSC-2 applications), and there is no expectation that these non-realistic configurations be tested.

Scope

This document references the v0.0.18 DRAFT “SNIA Green Storage Power Measurement Specification” Storage Taxonomy Summary, shown here as Figure 1. Additional details may be found at: http://www.snia.org/green/.

Figure 1: SNIA GSI Storage Taxonomy

Storage Taxonomy Summary	**Online Storage**	**Near Online Storage**	**Removable Media Libraries**	**Virtual Media Libraries**	**Infrastructure Appliances**	**Infrastructure Interconnect**
	Prime storage, able to serve random as well as sequential workloads with minimal delay	Intended as second tier storage behind Online Storage. Able to service Random and Sequential workloads, but perhaps with noticeable delay in time to 1st data access.	Archival storage used in a sequential access mode. A Typical example would be Tape based archival, both Stand Along and Robotically assisted libraries.	Storage which simulates removable Media Libraries. Will typically use non tape based storage and as such are able to respond to data requests more quickly	Devices placed in the storage SAN or network adding value through one or more dedicated Storage enhancements. Examples include: SAN Virtualization, Compression, De-duplication, etc.	Devices which enable a SAN or other Storage Network data switching or routing.
Maximum Capacity Guidance Note: Maximum Capacity Guidance reflects the maximum capacity a given offering can be purchased with and/or field upgraded to. It is intended to be used as a guideline as apposed to an absolute value. There will be case where a device may have greater or small capabilities, but otherwise is an appropriate match for a given classification due to other criteria, e.g.: redundancy capabilities	**Max Storage Devices** (Up to 80 Ms MTTD)	**Max Storage Devices** (Over 80Ms MTTD)	**Max Tape Drives**		**Max Storage Devices Supported***	**Max Port Count**
Group 1) SoHo & Consumer Storage which is designed primarily for home (consumer) or home / small office usage. –Often Direct Connected (USB, IP, etc) –No option for redundancy (will contain SPOFs)	Up to 4 Devices	MTTD = Max Time to Data Maximum time needed to access any data stored in any place on the storage system	Stand Alone Drive (No Robotics)		Note: * Infrastructure Appliances by definition have no intrinsic storage, other than what is used for local processing and/or local Cashing of data. Storage Devices Support in this case refers to the number of storage devices controllable down stream of the Appliance	
Group 2) Entry, DAS, or JBOD Storage which is dedicated to one or at most a very limited number of servers. Often will not include any integrated controller, but rely on server host for that functionality. –Often Direct Connected (SATA, IP, etc.) –May optionally offer limited number of redundancy features	More than 4 Devices	Up to 4 Devices	Up to 4 Drives			Up to 32
Group 3) Entry / Midrange SAN or NAS connected storage which places a higher emphasis on value than scalability and performance. This is often referred to as ‘Entry Level’ storage. –Network connected (IP, SAN, etc.) –Has options for redundancy features	More than 20 Devices	More than 4 Devices	More than 4 Drives	Up to 100 Devices	Support for up to 20 Devices	Up to 128
Group 4) Midrange / Enterprise SAN or NAS connected storage which delivers a balance of performance and features. Offers higher level of management as well as scalability and reliability capabilities. –Network connected (IP, SAN, etc.) –Has options for and often delivered with full redundancy (no SPOF)	More than 100 Devices	More than 100 Devices	More than 24 Drives	More than 100 Devices	Support for more than 20 Devices	More than 128
Group 5) Enterprise / Mainframe Storage which exhibits large scalability and extreme robustness associated with Mainframe deployments, though are not restricted to Mainframe only deployments. –Mainframe connectivity with optional network connection (IP, SAN..) –Always delivered with full redundancy (no SPOF) –Often Capable of non-disruptive serviceability	More than 1000 Devices		More than 11 Drives	More than 100 Devices	Support for more than 100 Devices	© SNIA 2009

For purposes of Round 1 data collection, EPA is interested in the following taxonomy classifications:

- ONLINE: Groups 2, 3, 4, and 5
- NEAR-ONLINE: Groups 2, 3, 4
- REMOVABLE MEDIA LIBRARY: Groups 2, 3, 4 and 5
- VIRTUAL MEDIA LIBRARY: Groups 3, 4, and 5

Objective

Stakeholders are encouraged to submit power consumption data for as many products, in as many Generic System Configurations (GSCs), as can be reasonably accomplished during the data collection period. If possible, testing and modeling should be repeated on similar systems after making single-variable changes to the hardware or software configuration. Examples of single-variable changes include:

- Change disk drive type or technology
- Add or remove storage capacity
- Modify RAID configuration
- Change from single to redundant controllers

Any simulation models used to generate data must be capable of predicting variations in energy consumption for the various phases of the test sequence. All models should be run using the test workload sequences defined below for the appropriate taxonomy category.

In order to assess model accuracy, stakeholders should model any system configurations that are tested so that results can be easily compared. Further modeling exercises can then be used to simulate the impact of a wide variety of additional single- and multi-variable hardware and software configuration changes.

To summarize, EPA is interested in collecting the following, in order of priority:

1. Data from testing of as many GSCs as possible.
2. Data from re-testing of GSCs after single-variable changes to hardware or software configuration.
3. Data from modeling of any tested system configurations
4. Data from modeling of additional single- or multi-variable changes to hardware or software configuration.

Test Setup

TEST EQUIPMENT

UUT input voltage and power should be measured with a power meter at an appropriate location 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).

The power meter shall have the following capabilities:

- Measure and record UUT input voltage with an accuracy of 1% and a measurement frequency of no more than 5 seconds.
- Measure and record UUT instantaneous power consumption with an accuracy of +/- 1.0 Watts and a measurement frequency of no more than 5 seconds.

INPUT VOLTAGE

The power supplied to the UUT shall be consistent with one of the options presented below:

Table 1: Input Power Requirements

Input Voltage Range	Phases	AC 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 Procedure

ONLINE & NEAR-ONLINE CATEGORIES

The following test sequence shall be applied to systems in the Online and Near-Online taxonomy categories.

Table 2: Test Sequence for Online & Near-Online

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration[2]
Pre-conditioning	Random 70% Read 30% Write	100%	8K	10 min
Active “A”	Random Read	100%	8K	10 min
Active “B”	Random Write	100%	8K	10 min
Active “C”	Sequential Read	100%	256K	10 min
Active “D”	Sequential Write	100%	256K	10 min

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration[2]
Active "E"	Random 70% Read 30% Write	25%	8K	10 min
Active "F"	Random 70% Read 30% Write	80%	8K	10 min
Active "G"	Random 70% Read 30% Write	100%	8K	10 min
Ready Idle	n/a	0%	n/a	30 min
Deep Idle (Optional)[3]	n/a	0%	n/a	10 min

Notes:

1. Response Time for Online and Near-online systems during Active test phases must not exceed 30 ms.
2. Each test phase must be of sufficient duration to exercise the entire storage system and achieve stable system performance. Once stability has been achieved, the measurement period begins and shall continue for the "Sustained Duration" specified in the table.
3. The Deep Idle test phase is included to allow storage systems to demonstrate advanced power saving capability by placing one or more storage devices into a reduced power state. Deep Idle should only be performed on systems which offer a Deep Idle feature that may be enabled and configured by the system operator. Storage systems which by design require a subset of storage devices to be in a reduced power state and as a result cannot meet the 30 ms Response Time requirement should be considered to be in the Near-online category.

REMOVABLE & VIRTUAL MEDIA LIBRARY CATEGORIES

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

Table 3: Test Sequence for Removable & Virtual Media Libraries

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration[5]
Pre-conditioning	Sequential Write → Rewind → Read	100%	128K	10 min

Phase	Workload	% of Max Sustainable Performance	Block Size	Sustained Duration[5]
Active "A"	Sequential Write	100%	128K	10 min
Active "B"	Sequential Read	100%	128K	10 min
Ready Idle	n/a	0%	n/a	30 min
Deep Idle (Optional)[6]	n/a	0%	n/a	10 min

Notes:

4. Response Time for Virtual and Removable Media Library systems during Active test phases must not exceed 30 ms.
5. Each test phase must be of sufficient duration to exercise the entire storage system and achieve stable system performance. Once stability has been achieved, the measurement period begins and shall continue for the "Sustained Duration" specified in the table.
6. The Deep Idle test phase is included to allow storage systems to demonstrate advanced power saving capability by placing one or more storage devices into a reduced power state. Deep Idle should only be performed on systems which offer a Deep Idle feature that may be enabled and configured by the system operator.