SNIA Emerald™ Power Efficiency Measurement Specification

Version 3.0.3

ABSTRACT: This document describes a standardized method to assess the energy efficiency of commercial storage products in both active and idle states of operation. A taxonomy is defined that classifies storage products in terms of operational profiles and supported features. Test definition and execution rules for measuring the power efficiency of each taxonomy category are described; these include test sequence, test configuration, instrumentation, benchmark driver, IO profiles, measurement interval, and metric stability assessment. Qualitative heuristic tests are defined to verify the existence of several capacity optimization methods. Resulting power efficiency metrics are defined as ratios of idle capacity or active operations during a selected stable measurement interval to the average measured power.

This document has been released and approved by the SNIA. The SNIA believes that the ideas, methodologies and technologies described in this document accurately represent the SNIA goals and are appropriate for widespread distribution. Suggestions for revisions should be directed to http://www.snia.org/feedback/.

SNIA Technical Position

October 11, 2018

USAGE

DISCLAIMER

The information contained in this publication is subject to change without notice. The SNIA makes no warranty of any kind with regard to this specification, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The SNIA shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this specification.

Revision History

Revision	Date	Changes
0.0.1	August 19, 2015	First Draft
0.0.2	Sept. 30, 2015	Starting with V2.1.0, the V2.1.0 Errata list was incorporated: Revised sections 7.3.7 (deleted reference to Vdbench version) 7.4.5.3 (specified COM test data) and 7.4.5.5 (revised steps). Added revised Data Sets table.
V2.1.1	Dec. 2, 2015	Published as a SNIA Technical Position, per SNIA Technical Council approval
V3.0.0	June 28, 2017	Add tests for file access devices of Online and Near-Online taxonomy categories. Eliminate tests for Parity RAID as a COM. Improve tests for Data Deduplication and Compression COMs. Eliminate taxonomy categories Adjunct Product and Interconnect Element. Add clause on disclosure requirements. Add Stable Storage. Align power requirements with ENGERY STAR Data Center Storage. Add temperature meters to Annex A. Reorganize document. Change power meter collection interval to match operation rate collection interval. Add clarifications. Make terminology more consistent.
V3.0.0	Sept. 11, 2017	Adopted by SNIA Members as SNIA Technical Position.
V3.0.1	Sept. 11, 2017	Correct inconsistency in collection interval operations rate for file access data collection in Table B.3 and subclause 7.4.2.5 by making collection interval consistently 10 seconds. Approved by SNIA Technical Council effective this date.
V3.0.2	May 2, 2018	Address inconsistencies and errors in sub clauses 7.2.9, 7.2.11, 7.4.2.6, and Annex B.
V3.0.2	May 23, 2018	Approved by SNIA Technical Council as a corrected SNIA Technical Position.
V3.0.3	Oct. 3, 2018	Update editorially to prepare for ISO Publicly Available Specification submission. Make additional editorial corrections and clarifications.
V3.0.3	Oct. 11, 2018	Approved by SNIA Technical Council as a corrected SNIA Technical Position.

Suggestions for revisions should be directed to http://www.snia.org/feedback/.

CONTACTING SNIA

SNIA Emerald™ Program Web Site

The SNIA Emerald™ Program web site is https://www.snia.org/emerald. SNIA Emerald™ Program-related downloads are available at https://www.snia.org/emerald/download, the *Download Materials for SNIA Emerald™ Testing* web page.

SNIA Address

Requests for interpretation, suggestions for improvement and addenda, or defect reports are welcome. They should be sent via the SNIA Feedback Portal at http://www.snia.org/feedback/ or by mail to the Storage Networking Industry Association, 4360 ArrowsWest Drive, Colorado Springs, Colorado 80907, U.S.A.

INTENDED AUDIENCE

This document is intended for use by individuals and companies engaged in assessing the power utilization of storage products.

Acknowledgements

Contents

List of Tables, Figures and Equations

1 Introduction

There is a growing awareness of the environmental impact of IT equipment use. This impact takes several forms: the energy expended in equipment manufacture and distribution, the impact of materials reclamation, and the energy consumed in operation and cooling of the equipment. IT equipment users of all kinds now wish to make their IT operations as energy efficient as possible. This new priority can be driven by one or more of several requirements:

- Rising energy costs have made power and cooling expenses a more significant percentage of total cost of ownership of server and storage equipment;

- Some data centers are physically unable to add more power and cooling load, which means that new applications and data can only be brought on if old ones are retired or consolidated onto new, more efficient configurations;

- Increased regulatory and societal pressures provide incentives for companies to lower their total energy footprints. For many companies, IT is a significant portion of overall energy consumption, and corporate Green goals can only be achieved by reducing IT's energy needs or by making operations more efficient.

IT equipment users will seek advice on the most energy efficient approach to getting their work done. It is not practical for customers to test a wide range of storage products and architectures for themselves. A more effective approach is to create a collection of standard metrics that allow IT architects to objectively compare a range of possible solutions. This objective, metric-based approach has a dual impact:

- Users can select the mode of storage usage that accomplishes their work objectives with the lowest overall energy consumption;

- Companies will be driven to innovate and compete in the development of energy efficient products as measured by the standard yardsticks.

2 Scope

2.1 Abstract

This document describes a standardized method to assess the energy efficiency of commercial storage products in both active and idle states of operation. A taxonomy is defined that classifies storage products in terms of operational profiles and supported features. Test definition and execution rules for measuring the power efficiency of each taxonomy category are described; these include test sequence, test configuration, instrumentation, benchmark driver, IO profiles, measurement interval, and metric stability assessment. Qualitative heuristic tests are defined to verify the existence of several capacity optimization methods. Resulting power efficiency metrics are defined as ratios of idle capacity or active operations during a selected stable measurement interval to the average measured power.

2.2 Introduction

This document defines methodologies and metrics for the evaluation of the related performance and energy consumption of storage products in specific active and idle states.

Storage products and components are said to be in an "active" state when they are processing externally initiated, application-level requests for data transfer between host(s) and the storage product(s). For purposes of this document, idle is defined as "ready idle", in which storage systems and components are configured, powered up, connected to one or more hosts and capable of satisfying externally initiated, application-level initiated IO requests within normal response time constraints, but no such IO requests are being submitted.

2.3 Current Revision

This document addresses storage products supporting block or file data access. Block access and file access refer to the type of service provided typically by Storage Area Network (SAN) and Network Attached Storage (NAS) systems, respectively. It is not appropriate to use this document to ascertain power efficiency for anything other than these two access modes. This document includes:

- A generalized taxonomy for storage products (clause 5);

- An assessment mechanism for software-based Capacity Optimization Methods (clause 6);

- Measurement and data collection guidelines for assessing the power efficiency of block- and file-based storage products in both active and ready idle states (clause 7);

- Metrics describing storage product power efficiency (clause 8);

- Required disclosures for a test result published as a SNIA Emerald™ Power Efficiency Measurement test result (clause 9).

2.4 Purpose

The purpose of a SNIA Emerald™ Power Efficiency Measurement is to provide a reproducible and standardized assessment of the energy efficiency of commercial storage products in both active and ready idle states.

1. Tested systems shall be comprised of commercially released products and components;

2. Tested systems shall employ settings, parameters, and configurations that would allow end-users to achieve power efficiency levels equivalent to the published result;

3. All data published as an SNIA Emerald™ Power Efficiency Measurement test result shall be gathered from test execution conducted according to this document;

4. Test execution shall complete in its entirety and without test failure or test error messages;

5. Software features which invoke, generate, or use software designed specifically for the test shall not be used. Configuration options chosen for test execution shall be options that are generally recommended for the customer;

6. Energy for powering the equipment shall be provided by the power mains that are being monitored (not by an internal UPS).

A SNIA Emerald™ Power Efficiency Measurement shall be a good faith effort to accurately characterize the power requirements of the tested system. The precise configuration used in a SNIA Emerald™ Power Efficiency Measurement is left to the sponsor of a test. Any commercially released components may be used, and a focus on new or emerging components or technologies is encouraged.

2.5 Disclaimer

A SNIA Emerald™ Power Efficiency Measurement test result provides a high-level assessment of the energy efficiency of the tested system in specific ready idle and active states. It is not an attempt to precisely model or reproduce any specific installation.

Actual performance and energy consumption behavior is highly dependent upon precise workload, environmental and usage parameters. While a SNIA Emerald™ Power Efficiency Measurement test result is intended to provide a realistic and reproducible assessment of the relative power efficiency of a system across a broad range of configurations and usage patterns, it cannot completely match the precise needs of any one specific installation.

3   Normative References

The following documents are referred to in the text in a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

Table 1 lists these documents.

Table 1 – Normative References

Author/Owner	Title	Revision	URL
ISO/IEC	*ISO/IEC Directives Part II*	Eighth edition, 2018	https://www.iso.org/directives-and-policies.html

4 Definitions, Symbols, Abbreviations, and Conventions

4.1 Overview

For the purposes of this document, the terms and definitions given in The SNIA Dictionary[2] and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

• ISO Online browsing platform: available at https://www.iso.org/obp

• IEC Electropedia: available at http://www.electropedia.org/

The terms and definitions defined in this document are based on those found in The SNIA Dictionary[2]. They have been extended, as needed, for use in this document. In cases where the current definitions in the SNIA dictionary conflict with those presented in this document, the definitions in this document shall apply.

4.2 Definitions

4.2.1

auto-tiering

policy-based system that automatically places and moves data across tiers to optimize performance service levels, cost targets, and overall energy consumption

Note 1 to entry: Each storage tier may comprise different storage technologies, offering varying performance, cost, and energy consumption characteristics.

4.2.2

cache

temporary storage used to transparently store data for expedited access to or from slower media, and not directly addressable by end-user applications

4.2.3

capacity optimization method (COM)

subsystem, whether implemented in hardware or software, which reduces the consumption of space required to store a data set

4.2.4

committed data

data that has been written to stable storage

4.2.5

compression

the process of encoding data to reduce its size

4.2.6

count-key-data (CKD)

disk data organization model in which the disk is assumed to consist of a fixed number of tracks, each having a maximum data capacity

Note 1 to entry: The CKD architecture derives its name from the record format, which consists of a field containing the number of bytes of data and a record address, an optional key field by which particular records can be easily recognized, and the data itself.

4.2.7

data deduplication

replacement of multiple copies of data—at variable levels of granularity—with references to a shared copy in order to save storage space and/or bandwidth

4.2.8

dedupable

property that a collection of data is said to possess if the needed storage capacity for the data is reduced significantly by data deduplication

4.2.9

delta snapshot

type of point in time copy that preserves the state of data at an instant in time, by storing only those blocks that are different from an already existing full copy of the data

4.2.10

direct-connected

storage designed to be under the control of a single host, or multiple hosts in a non-shared environment

4.2.11

efficiency

ratio of useful work to the power required to do the work

4.2.12

file

abstract data object made up of a) an ordered sequence of data bytes stored on a disk or tape, b) a symbolic name by which the object can be uniquely identified, and c) a set of properties, such as ownership and access permissions that allow the object to be managed by a file system or backup manager

4.2.13

file system

software component that imposes structure on the address space of one or more physical or virtual disks so that applications may deal more conveniently with abstract named data objects of variable size (files)

4.2.14

fixed block architecture (FBA)

model of disks in which storage space is organized as linear, dense address spaces of blocks of a fixed size

Note 1 to entry: Fixed block architecture is the disk model on which SCSI is predicated.

4.2.15

fixed content addressable storage (FCAS)

storage optimized to manage content addressable data that is not expected to change during its lifetime

4.2.16

formatted capacity

total number of bytes available to be written after a system or device has been formatted for use, e.g., by an object store, file system or block services manager

Note 1 to entry: Formatted capacity, also called usable capacity, is less than or equal to raw capacity. It does not include areas set aside for system use, spares, RAID parity areas, checksum space, host- or file system-level remapping, "right sizing" of disks, disk labeling and so on. However, it may include areas that are normally reserved—such as snapshot set-asides—if they can alternatively be configured for ordinary data storage by the storage administrator.

4.2.17

free space

amount of additional irreducible data that can be written to the product under test as configured

4.2.18

hot band

simulation of naturally occurring hot spots

4.2.19

hot spot

area of storage more frequently accessed across the addressable space

4.2.20

IO intensity

measure of the number of IOPS requested by a load generator

Note 1 to entry: IO intensity is phrased as a percentage of selected maximum IOPS level that satisfies the timing requirement(s) for a taxonomy category.

4.2.21

irreducible data

data that is neither compressible nor dedupable

4.2.22

load generator

hardware and software environment executing the workload generator to drive the product under test during measurements

4.2.23

Logical Unit (LU)

entity within a SCSI target that executes IO commands

4.2.24

Logical Unit Number (LUN)

synonym for logical unit

4.2.25

Maximum Time to First Data (MaxTTFD)

maximum time required to start receiving data from a storage system to satisfy a read request for arbitrary data

4.2.26

network-connected

storage designed to be connected to a host via a network protocol (e.g., TCP/IP, IB, and FC)

4.2.27

non-disruptive serviceability

support for continued availability of data during all FRU service operations, including break/fix, code patches, software/firmware upgrades, configuration changes, data migrations, and system expansion

Note 1 to entry: Service operations may impact performance, but shall not result in a loss of access.

4.2.28

parity RAID

collective term used to refer to Berkeley RAID Levels 3, 4, 5 and 6

4.2.29

permanent storage

data storage media which can retain data indefinitely without a power source

4.2.30

product under test

customer-orderable system or component that is the subject of a SNIA Emerald™ Power Efficiency Measurement

4.2.31

raw capacity

sum of the raw, unformatted, uncompressed capacity of each storage device in the product under test

4.2.32

ready idle

operational state in which a system is capable of satisfying an arbitrary IO request within the response time and MaxTTFD constraints of its selected taxonomy category, but no user-initiated IO requests are being submitted to the system. In the ready idle state, background I/O activity, autonomously initiated by the solution under test, may take place

4.2.33

sequential read

IO load consisting of consecutively issued read requests to logically adjacent data

4.2.34

sequential write

IO load consisting of consecutively issued write requests to logically adjacent data

4.2.35

Single Point of Failure (SPOF)

one component or path in a system, the failure of which would make the system inoperable or data inaccessible

4.2.36

SNIA Emerald Power Efficiency Measurement

test performed according to this document

4.2.37

SNIA Emerald Power Efficiency Measurement test result

result of a test performed according to this document

4.2.38

solution under test

All hardware and software components that are exercised during a test to verify functional behavior or determine performance characteristics, including all of the hardware and software components involved in a test

Note 1 to entry: It includes the physical and virtual components of the load generator, storage media, and the entire data path between the load generator and the storage media.

4.2.39

stable storage

storage that retains its content over power failures

Note 1 to entry: See Annex C.

4.2.40

storage controller

device for handling storage requests that includes a processor or sequencer programmed to autonomously process a substantial portion of IO requests directed to storage devices

4.2.41

storage device

collective term for disk drives, solid state drives and modules, tape cartridges, and any other mechanisms providing non-volatile data storage

Note 1 to entry: This definition is specifically intended to exclude aggregating storage elements such as RAID array subsystems, robotic tape libraries, filers, and file servers. Also excluded are storage devices which are not directly accessible by end-user application programs, and are instead employed as a form of internal cache.

4.2.42

storage protection

any combination of hardware and software (e.g., RAID, NVRAM, disk sparing and background disk scrubbing or media scan) that assures that all IO operations will be preserved in the event of power loss or storage device failure

4.2.43

system crash

hardware or software failure which causes data to no longer be available, at least temporarily, and which requires a reboot of one or more hardware components and/or re-initialization of one or more software components in order for data access to be restored

4.2.44

test sponsor

individual, company, or agent that performs a test according to this document

4.2.45

thin provisioning

technology that allocates the physical capacity of a volume or file system as applications write data, rather than pre-allocating all the physical capacity at the time of provisioning

4.2.46

virtual drive

removable media storage device (e.g., tape drive) that is emulated using other storage devices

4.2.47

workload generator

software used in the load generator to drive the product under test during measurement

4.3   Symbols and Abbreviated Terms

CKD	Count-Key-Data
COM	Capacity Optimization Method
FBA	Fixed Block Architecture
FCAS	Fixed Content Addressable Storage
FRU	Field-Replaceable Unit
IT	Information Technology
LU	Logical Unit
LUN	Logical Unit Number
MAID	Massive Array of Idle Disks
MaxTTFD	Maximum Time to First Data
NVRAM	Non-Volatile Random Access Memory
RAS	Reliability, Availability, and Serviceability
SCSI	Small Computer System Interface
SNIA	Storage Networking Industry Association
SPOF	Single Point of Failure
UPS	Uninterruptible Power Supply

4.4   Expression of Provisions

This document uses the verbal forms for expressions of provisions as defined by ISO/IEC Directives Part 2 (Eighth edition, 2018). These verbal forms include shall, shall not, should, should not, may, may not, can, cannot, and must.

4.5   Conventions

Certain words and terms used in this document have a specific meaning beyond their normal English meaning. These words and terms are defined either in subclause 4.2 or in the text where they first appear.

Storage capacities are represented in base-10. IO transfer sizes and offsets are represented in base-2.

The associated units and abbreviations used in this document are:

- A kilobyte (KB) is equal to 10^3 B.

- A megabyte (MB) is equal to 10^6 B.

- A gigabyte (GB) is equal to 10^9 B.

- A terabyte (TB) is equal to 10^{12} B.

- A petabyte (PB) is equal to 10^{15} B.

- An exabyte (EB) is equal to 10^{18} B.

- A kibibyte (KiB) is equal to 2^{10} B.

- A mebibyte (MiB) is equal to 2^{20} B.

- A gibibyte (GiB) is equal to 2^{30} B.

- A tebibyte (TiB) is equal to 2^{40} B.

- A pebibyte (PiB) is equal to 2^{50} B.

- An exbibyte (EiB) is equal to 2^{60} B.

5   Taxonomy

5.1   Introduction

This clause defines a market taxonomy that classifies storage products or subsystems in terms of operational profile and supported features.

While this taxonomy is broad and defines a framework for products that range from consumer solutions to enterprise installations, it is not intended to address all storage devices. For example, this taxonomy does not address storage devices that rely on a Universal Serial Bus (USB) connection for their power.

Further, while this document includes definitions for its entire taxonomy, it does not include testing methodologies for the entire taxonomy. Both individual sections of the taxonomy (e.g., Near-Online-1) and broader and more general groups of sections are not addressed beyond taxonomy definition in this document. Their cells in Table 2 are shaded to make it clear that this omission in intentional.

Table 2 – Taxonomy Overview

Level	Category			
	Online (see 5.5)	Near-Online (see 5.6)	Removable Media Library (see 5.7)	Virtual Media Library (see 5.8)
Consumer/ Component[a]	Online 1	Near-Online 1	Removable 1	Virtual 1
Low-end	Online 2	Near-Online 2	Removable 2	Virtual 2
Mid-range	Online 3	Near-Online 3	Removable 3	Virtual 3
	Online 4			
High-end	Online 5	Near-Online 5	Removable 5	Virtual 5
Mainframe	Online 6	Near-Online 6	Removable 6	Virtual 6
[a] Entries in this level of taxonomy include both consumer products and data-center components (e.g., stand-alone tape drives)				

5.2 Taxonomy Categories

5.2.1 General

Taxonomy categories define broad market segments that can be used to group products that share common functionality or performance requirements, and within which meaningful product comparison can be undertaken. This document defines four broad taxonomy categories (summarized in Table 2):

- Online, defined in subclause 5.5;

- Near-Online, defined in subclause 5.6;

- Removable Media Library, defined in subclause 5.7;

- Virtual Media Library, defined in subclause 5.8.

Within a taxonomy category, a specific model or release of a product will support different feature sets, whether focused on capacity, reliability, performance, functionality, or another differentiator. Feature and functionality differences within a category are addressed with attributes. Each taxonomy category defines a set of attributes that are common to all products within the category.

5.2.2 Category Attributes

Where a taxonomy category requires a specific, fixed setting or range for a given attribute, that setting is summarized in Table 3 to assist a test sponsor in initial category selection. The full set of attributes for each category is provided in subclauses 5.5 through 5.8.

Table 3 – Common Category Attributes

Attribute	Category			
	Online	**Near-Online**	**Removable Media Library**	**Virtual Media Library**
Access Pattern	Random/ Sequential	Random/ Sequential	Sequential	Sequential
MaxTTFD (t)	t < 80 ms	t > 80 ms	t > 80 ms t < 5 min	t < 80 ms

The attribute Consumer/Component refers to either

- Any of a wide array of manufactured goods which are purchased primarily for personal, family and/or household purposes;

- A data-center component (e.g., a stand-alone tape drive or an individual hard disk drive) that is a part of a data-center storage product.

5.3 Taxonomy Classifications

Classifications define combinations of settings or values for the attributes within a category.

5.4 Taxonomy Rules

A product shall satisfy all the attributes for its designated category and designated classification. If a product satisfies the attributes of multiple classifications, the product may be considered to be in any or all of these classifications. In cases where storage devices within a product fall within more than one

category or classification, the product shall be considered a member of the category and classification whose requirements can be met by all of its storage devices.

Maximum Supported Configuration bounds the number of storage devices that the product is capable of supporting.

A product shall be considered to be of a taxonomy category and classification if, in the Classification Table for the category (Table 4, Table 5. Table 6, or Table 7):

1. The product has each of the features and functionalities listed as *required* or described in a table entry;
2. The product has no feature or functionality listed as *prohibited.*

Note1: The presence or absence of attributes label as Optional has no impact on the category and classification of a product.

Note 2: The nature of features and functionalities listed as Not Specified is not material to the categorization of a product.

Note 3: Based on these rules, it is possible that a product is considered to be of multiple taxonomy categories and/or multiple taxonomy classifications of a category.

5.5 Online Category

This category defines the features and functionalities for an online, random-access storage product. Products in this profile may provide any combination of block, file, or object interfaces. Table 4 defines the requirements for the taxonomy classifications defined in this category.

Table 4 – Online Classifications

Attribute	Classification					
	Online 1	Online 2	Online 3	Online 4	Online 5	Online 6
Access Pattern	Random/ Sequential	Random/ Sequential	Random/ Sequential	Random/ Sequential	Random/ Sequential	Random/ Sequential
MaxTTFD (t)	t < 80 ms	t < 80 ms	t < 80 ms	t < 80 ms	t < 80 ms	t < 80 ms
Connectivity	Not Specified	Direct-connected to single or multiple hosts	Network-connected	Network-connected	Network-connected	Network-connected
Consumer/ Component	Yes	No	No	No	No	No
Integrated Storage Controller	Optional	Optional	Required	Required	Required	Required
Storage Protection	Optional	Optional	Required	Required	Required	Required
No SPOF	Optional	Optional	Optional	Required	Required	Required
Stable storage support	Optional, unless Required by protocol	Optional, unless Required by protocol	Required	Required	Required	Required
Non-Disruptive Serviceability	Optional	Optional	Optional	Optional	Required	Required
FBA/CKD Support	Optional	Optional	Optional	Optional	Optional	Required
Maximum Supported Configuration[a]	≥ 1	≥ 4	≥ 12	> 100	> 400	> 400

[a] Maximum Supported Configuration does not apply to an all solid-state system that is not based on replaceable storage devices.

SNIA Emerald™ Power Efficiency Measurement Specification V3.0.3
SNIA Technical Position

5.6 Near-Online Category

This category defines the features and functionalities for a near-online, random-access storage product. Products in this profile employ MAID or FCAS architectures as well as any combination of block, file, or object interfaces. Table 5 defines the requirements for this taxonomy classifications defined in this category.

Table 5 – Near-Online Classifications

Attribute	Classification					
	Near-Online 1	Near-Online 2	Near-Online 3	Near-Online 4	Near-Online 5	Near-Online 6
Access Pattern	Random/ Sequential	Random/ Sequential	Random/ Sequential		Random/ Sequential	Random/ Sequential
MaxTTFD (t)	t > 80 ms	t > 80 ms	t > 80 ms		t > 80 ms	t > 80 ms
Connectivity	Not Specified	Network connected	Network connected		Network connected	Network connected
Consumer/ Component	Yes	No	No		No	No
Integrated Storage Controller	Optional	Optional	Required		Required	Required
Storage Protection	Optional	Optional	Required		Required	Required
No SPOF	Optional	Optional	Optional		Optional	Required
Non-Disruptive Serviceability	Optional	Optional	Optional		Optional	Required
FBA/CKD Support	Optional	Optional	Optional		Optional	Optional
Maximum Supported Configuration	≥ 1	≥ 4	≥ 12		> 100	> 1 000

5.7 Removable Media Library Category

This category defines the features and functionalities for storage products that rely on automated or manual media loaders (e.g., tape or optical libraries). Table 6 defines the requirements for the taxonomy classifications defined in this category.

Table 6 – Removable Media Library Classifications

Attribute	Classification					
	Removable 1	Removable 2	Removable 3	Removable 4	Removable 5	Removable 6
Access Pattern	Sequential	Sequential	Sequential		Sequential	Sequential
MaxTTFD (t)	t > 80 ms t < 5 min	t > 80 ms t < 5 min	t > 80 ms t < 5 min		t > 80 ms t < 5 min	t > 80 ms t < 5 min
Robotics	Prohibited	Required	Required		Required	Required
No SPOF	Optional	Optional	Optional		Optional	Required
Non-disruptive Serviceability	Optional	Optional	Optional		Optional	Required
Maximum Supported Drive Count	Not Specified	4	≥ 5		≥ 25	≥ 25

5.8 Virtual Media Library Category

This operational profile defines the features and functionalities for sequential-access storage products that rely on non-removable storage media to provide a Virtual Media Library. Table 7 defines the requirements for the taxonomy classifications defined in this category.

Table 7 – Virtual Media Library Classifications

Attribute	Classification					
	Virtual 1	Virtual 2	Virtual 3	Virtual 4	Virtual 5	Virtual 6
Access Pattern	Sequential	Sequential	Sequential		Sequential	Sequential
MaxTTFD (t)	t < 80 ms	t < 80 ms	t < 80 ms		t < 80 ms	t < 80 ms
Storage Protection	Optional	Optional	Required		Required	Required
No SPOF	Optional	Optional	Optional		Optional	Required
Non-Disruptive Serviceability	Optional	Optional	Optional		Optional	Required
Maximum Supported Configuration	12	> 12	> 48		> 96	> 96

6   Capacity Optimization

6.1   Introduction

Hardware efficiencies are essential for reducing the amount of power used by storage, but equally real savings are obtained by capacity optimization. Capacity optimization refers to a set of techniques which collectively reduce the amount of storage necessary to meet storage objectives. Reduced use of storage (or increased utilization of raw storage) will result in less energy usage for a given task or objective.

Each of these techniques is known as a capacity optimization method (COM). The COMs are largely, though not completely, independent. In other words, they provide benefit in any combination, though their combined effect does not precisely equal the sum of their individual impacts. Nonetheless, since data sets vary greatly, a hybrid approach using as many techniques as possible is more likely to minimize the capacity requirements of any given data set, and therefore is also likely to achieve the best results over the universe of data sets. In addition, the space savings achievable through the different COMs are sufficiently close to one another that they are roughly equivalent in storage capacity impact.

6.2   Space Consuming Practices

A central assumption of this arena is that certain space consuming practices are essential to the storage of data at a data center class service level:

- Disk-based redundancy. When one or more drives (or other storage devices) fail—the number depending on service level—no interruption in service or loss of data occurs;

- Sufficient space. An application shall have enough space provisioned for it, so that no downtime shall be incurred during normal operation;

- Point-In-Time (PIT) copy availability. Data center applications under test need access to PIT copies of data sets that they can manipulate without fear of interference with live data;

- Disaster recovery. When data corruption or loss does occur, good copies of the data must be available to restore service.

6.3   COMs Characterized

In this document, tests for the presence of the following COMs are defined:

- **Delta Snapshots**: applicable to backup, PIT copy availability and disaster recovery. Both read-only and writeable delta snapshots are featured in shipping systems, but there are fundamental technical differences between them, and some systems implement only the read-only version;

- **Thin Provisioning**: a technology that allocates the physical capacity as applications write data, rather than preallocating all the physical capacity at the time of provisioning;

- **Data Deduplication:** addresses issues caused by multiple backups of the same data sets, and the tendency of large data sets, due to human usage patterns, to contain many copies of the same data (not necessarily on file boundaries)[1];

- **Compression**: takes advantage of the inherent compressibility of many data sets.

The following COMs and additional COMs may exist, however no test for presence is defined in this document.

- **Parity RAID:** addresses the need for disk-based redundancy.

7 Test Definition and Execution Rules

7.1 Overview

This clause defines the data collection and testing methodology that shall be used in a SNIA Emerald™ Power Efficiency Measurement. The data collected using the procedures defined in this clause becomes the basis for the metrics defined in clause 8.

Online and Near Online products shall be tested as either block access or file access systems. Removable Media Library and Virtual Media Library products shall be tested as block access systems. Throughout this clause the block access execution rules are listed first followed by the file access execution rules.

Materials and information to support testing according to this document are available at the SNIA Emerald™ Program web site[12].

7.1.1 Block Access Execution Overview

A SNIA Emerald™ Power Efficiency Measurement consists of a sequence of tests:

1. Pre-fill test, which puts data on the product under test;

2. Conditioning test, which assures accurate and reproducible measurements;

3. Active test, the basis for the active metrics;

4. Ready idle test, the basis of the ready idle metric;

5. Capacity optimization test (if defined), the basis of the secondary, capacity optimization metrics.

Some tests involve a timed sequence of defined measurements taken over defined intervals. Some tests have subordinate test phases. Clause 7 details the precise requirements for completing a given test for each taxonomy category, as well as any subordinate test phases defined within a given test. A valid measurement shall adhere to all requirements that are specific to the taxonomy category selected for the result (see subclause 7.2.2) as well as any general requirements for test execution (see subclause 7.2).

7.1.2 File Access Execution Overview

A SNIA Emerald™ Power Efficiency Measurement consists of a sequence of tests:

1. Execution of the four workloads, the basis for the active metrics;

2. Ready idle test, the basis of the ready idle metric;

3. Capacity optimization test (if defined), the basis of the secondary, capacity optimization metrics.

7.2 General Requirements and Definitions

7.2.1 Configuration Guidelines

This document does not constrain the precise configuration and interconnection of the hardware necessary to complete a SNIA Emerald™ Power Efficiency Measurement. Figure 1 and Figure 2 are provided as guidelines, but test sponsors are free to modify their configuration to suit their particular needs and equipment, provided no other requirement of this document is violated.
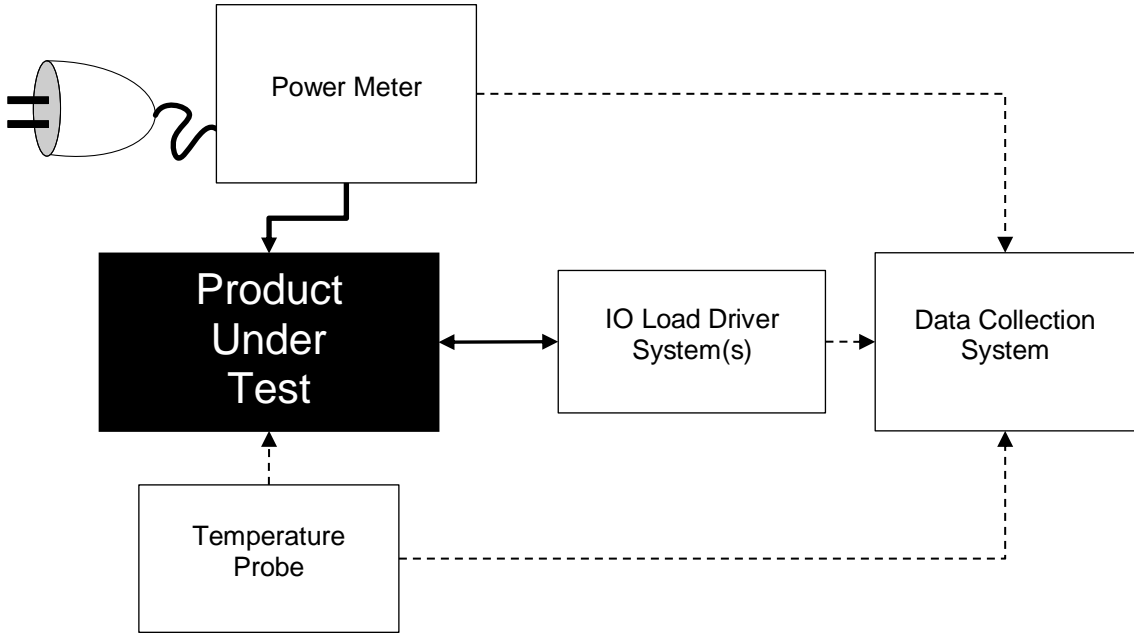
Figure 1 - Sample Configuration Block Access
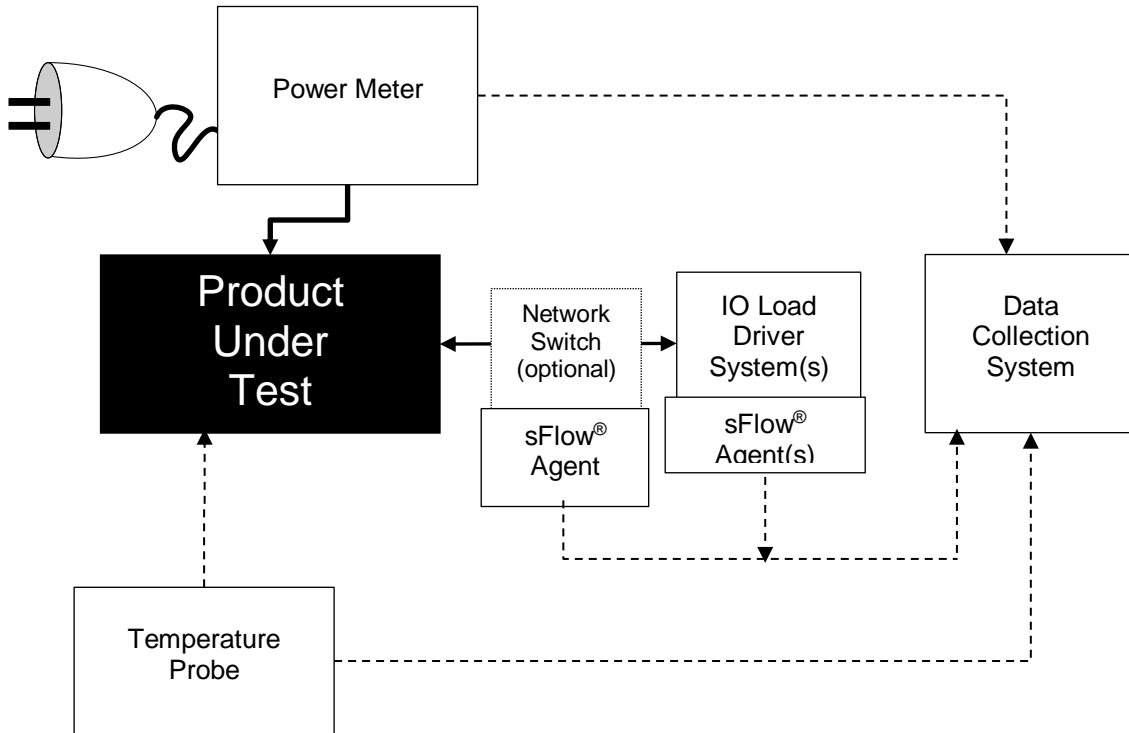
Figure 2 - Sample Configuration File Access

7.2.2 Product Under Test Configuration

The test sponsor shall identify one taxonomy classification for the product under test.

The product under test shall be configured to satisfy the requirements of the selected taxonomy classification.

The product under test shall represent a customer orderable configuration whose use within the selected taxonomy category is supported by the test sponsor.

For a product under test in the Removable Media Library category, all drives shall provide the same stated maximum data rate.

7.2.3 RAS

RAS features can have a significant impact on the power consumption of the product under test. Typical RAS features are summarized in Table 8.

Table 8 – Example RAS Features

Example RAS Features
Dual Controller (No SPOF Controller)
Mirroring (Local or remote, sync or async)
RAID 1, 4/5, 6
Snapshots (Full or Delta)
Disk Scrubbing
Multi-pathing
Disk Sparing
Dual Robotics
Drive-level Maintenance
Dual Power Supply
Variable-speed Fans

Any RAS features required to satisfy the requirements of the selected taxonomy category shall be enabled. The choice of what additional RAS features to enable in a product under test is left to the test sponsor.

If the product under test includes RAS features that are enabled for any test or test phase, then they shall be enabled for all tests and test phase, unless disabling of RAS features is explicitly allowed in the definition of a given test or test phase.

7.2.4 Product Under Test Consistency

The physical and logical configuration of the product under test, including its configuration and tuning parameters, shall not be changed between or during a test or test phase unless explicitly allowed in the definition of the test or test phase.

7.2.5 No Non-Test Activity

Other than booting/starting the product under test and any test equipment employed during the test, no substantive work shall be performed on the product under test between the tests or test phases defined in this document, unless explicitly allowed in the definition of the test or test phase.

7.2.6   Test Sequence

All tests shall be executed as an uninterrupted sequence, except as explicitly allowed by the execution requirements defined for a given test or test phase.

7.2.7   Power

The power supplied to the product under test shall match one of the power profiles outlined in Table 9 or Table 10.

Table 9 – Input Power Requirements (Products with Nameplate Rated Power ≤ 1 500 W)

Supply Voltage (AC RMS)	Phases	Voltage Tolerance	Frequency Tolerance	Maximum Total Harmonic Distortion
100 V, 115 V, 230 V	1	± 1 %	± 1 %	2 %
200 V, 208 V, 400 V	3	± 1 %	± 1 %	2 %

Table 10 – Input Power Requirements (Products with Nameplate Rated Power > 1 500 W)

Supply Voltage (AC RMS)	Phases	Voltage Tolerance	Frequency Tolerance	Maximum Total Harmonic Distortion
100 V, 115 V, 230 V	1	± 5 %	± 1 %	5 %
200 V, 208 V, 400 V	3	± 5 %	± 1 %	5 %

The power supplied to the product under test shall conform to the selected profile throughout test execution.

All batteries present in the product under test shall be fully charged at the start of testing.

7.2.8   Environmental

7.2.8.1   General Environmental Measurement Requirements

All measurements shall be conducted in a climate-controlled facility.

Note: The environmental conditions specified in subclauses 7.2.8.2 and 7.2.8.3 satisfy ASHRAE Class A1 standards for data centers as described in *Thermal Guidelines for Data Processing Environments*[20][21].

7.2.8.2   Temperature Requirements

Ambient temperature shall be within the range 18 °C to 28 °C for the duration of the test.

7.2.8.3   Humidity Requirements

Humidity shall be within the range 15 % relative humidity to 80 % relative humidity for the duration of the test.

7.2.9   Power and Temperature Measurement Instrumentation

The solution under test shall include a recommended power meter (sometimes called an *analyzer*). If the selected power meter does not gather environmental data, the solution under test shall include an environmental meter. See Annex A for information regarding recommended meters.

The power meter shall be active throughout all tests and test phases and shall record:

- Input voltage to the product under test, to an accuracy within 2 % at the 95 % confidence level;

- Power consumption by the SUT, to the resolution summarized in Table 11.

The power and voltage measurements shall be recorded to durable media using a period of not more than 5 s and shall use a timestamp that is synchronized with the other components of the solution under test to a resolution of at least 1 s.

Table 11 – Power Meter Resolution

Power Consumption (p)	Minimum Resolution
p ≤ 10 W	± 10 mW
10 < p ≤ 100 W	± 100 mW
p > 100 W	± 1 W

The temperature, measured in degrees Celsius, to an accuracy of ± 0,5 °C, as measured no more than 50 mm in front of (upwind of) the primary air inlet port for the product under test, shall be recorded to durable media using a reading rate of one reading every 10 s.

7.2.10   Workload Generator

7.2.10.1 Block Access Workload Generator

Vdbench (see Annex D) shall be used as the workload generator for Block Access testing. Information on access to Vdbench is provided in Annex D. Annex D identifies the version(s) of Vdbench that shall be used.

Test sponsors shall use a provided Vdbench Script (see Annex E). The script contains user adjustable parameters that the Test Sponsor may adjust within the constraints stated in the provided Vdbench Script.

This document takes precedence over any script if there is a conflict.

7.2.10.2 File Access Workload Generator

Test Sponsors shall use SPEC SFS® 2014 (see Annex F) as the workload generator for file access testing. Information on access to SPEC SFS® 2014 is provided in Annex F. Annex F identifies the version(s) of SPEC SFS® 2014 that may be used.

Test sponsors shall use a provided File Access IO Load Driver Configuration File (see Annex G). A configuration file contains user adjustable parameters.

This document takes precedence over any script if there is a conflict.

7.2.11   Instrumentation

7.2.11.1 Block Access Instrumentation

Vdbench shall be used as the workload generator and for collecting I/O rates and throughput data.

The load generator shall be active throughout all tests and test phases of the active metrics and shall record:

- IO rate and throughput in IO/sec and MiB/s respectively, reported every minute.

7.2.11.2 File Access Instrumentation

SPEC SFS® 2014 (see Annex F) shall be used as the workload generator.

An sFlow® (see Annex H) collector shall be used to collect storage performance data. sFlow® data shall be acquired by an sFlow® agent or agents located either in a network switch or the load generating host(s). The workload generator shall be active throughout all tests and test phases of the active metrics and shall record:

- The workload transition events in the log file for use during the data reduction process.

sFlow® agent(s) and an sFlow® collector shall be active and recording data throughout all tests and test phases of the active metrics and shall record:

- IO throughput in MiB/sec to/from product under test, reported every 10 seconds.

7.2.12 IO Profiles

7.2.12.1 Block Access IO Profiles

7.2.12.1.1 Overview

The particular IO stimuli used to drive the product under test during a test or test phase are specified in terms of an IO profile (a.k.a. workload) made up of multiple attributes:

- Name: the name of the IO pattern for this stimulus. The identifier for the associated test phase is included parenthetically, when appropriate;

- IO Size: the number of bytes requested by a given read or write operation;

- Read/Write Percentage: the mixture of read/write IO requests within an IO profile;

- Transfer Alignment: Minimum granularity of IO transfer addresses. All transfer addresses within an IO stream shall be a multiple of this value;

- Access Pattern: either one or the other of the following two alternatives:

    – Random: Randomly distributed throughout the address space of the product under test;

    – Sequential, as defined in 4.2;

- Data Pattern: compression ratio of 2:1.

7.2.12.1.2 Sequential Access

The first IO within an IO Stream with a sequential access pattern shall use an offset randomly distributed throughout the address range provided to the workload generator, and rounded down to satisfy the Transfer Alignment requirement. Each subsequent IO request shall be sent to and satisfied by the product under test in sequence using an offset that satisfies Equation 7-1.

Equation 7-1: Sequential Transfer Offset

$$O_{n+1} = \left(O_n + S\right) \; MOD \; R$$

Where:

- O_n is an IO offset;
- S is the IO size;
- R is the formatted capacity of the product under test.

7.2.12.1.3 Hot Band IO Profile

7.2.12.1.3.1 Overview

The goal of the hot band IO profile is to provide a workload that considers the contribution of auto-tiering mechanisms, e.g., read caching. This workload consists of a mix of different IO sizes and access patterns with a skewed access across a range of blocks. For example, this skewed access tends to hold data in cache and creates "cache hits" for improved throughput and reduced power consumption.

7.2.12.1.3.2 Exponential Access Pattern

Within a hot band, the probability of block access is skewed. Not all blocks are accessed equally. For example, this can result in an access pattern that creates a cache-friendly workload. The larger the cache size, the better the cache hit rate, as shown in Figure 3.
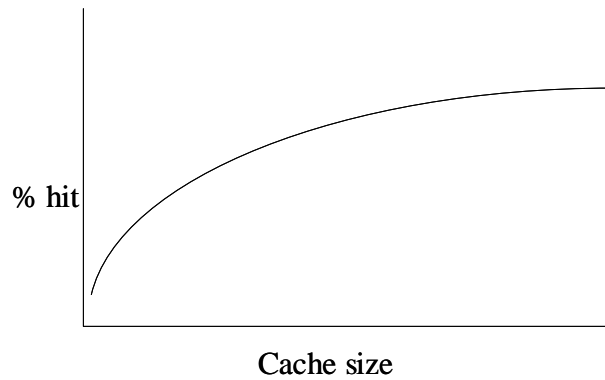
Figure 3 - Percentage of Address Hit vs. Cache Size

7.2.12.1.3.3 Workloads within the Hot Band IO Profile

Table 12 shows information concerning workloads within the hot band IO profile.

SNIA Emerald™ Power Efficiency Measurement Specification V3.0.3
SNIA Technical Position

Table 12 – Workloads within the Hot Band IO Profile

IO Profile	% of workload (Vdbench skew)	Read/Write Percentage	IO Size (KiB)	Access Pattern	Usable Address Range
Write Stream 1	5	0/100	See Table 13/14[a]	Sequential	0 % to 100 %
Write Stream 2	5	0/100	See Table 13/14[a]	Sequential	0 % to 100 %
Write Stream 3	5	0/100	See Table 13/14[a]	Sequential	0 % to 100 %
Read Stream 1	5	100/0	See Table 13/14[a]	Sequential	0 % to 100 %
Read Stream 2	5	100/0	See Table 13/14[a]	Sequential	0 % to 100 %
Read Stream 3	5	100/0	See Table 13/14[a]	Sequential	0 % to 100 %
Read Stream 4	5	100/0	See Table 13/14[a]	Sequential	0 % to 100 %
Read Stream 5	5	100/0	See Table 13/14[a]	Sequential	0 % to 100 %
Uniform Random	6	50/50	See Table 13/14[a]	Random	0 % to 100 %
Hot Band 1	28	70/30	See Table 13/14[a]	Random	10 % to 18 %
Hot Band 2	14	70/30	See Table 13/14[a]	Random	32 % to 40 %
Hot Band 3	7	70/30	See Table 13/14[a]	Random	55 % to 63 %
Hot Band 4	5	70/30	See Table 13/14[a]	Random	80 % to 88 %

[a] For storage products using native 512 B sectors, see Table 13. For products using native 4 KiB sectors, see Table 14.

7.2.12.1.3.4 Variable IO

The IO transfer (xfer) size used within the hot band IO profile is listed in Table 13 and Table 14.

Table 13 – IO Transfer Size within the Hot Band IO Profile for 512 Byte Native Devices

Xfer	Streaming Write	Streaming Read	Uniform	Hot Band
512 B			2 %	2 %
1 KiB			2 %	2 %
4 KiB	29 %	29 %	27 %	27 %
8 KiB	33 %	33 %	31 %	31 %
16 KiB	6 %	6 %	5 %	5 %
32 KiB	5 %	5 %	5 %	5 %
48 KiB			1 %	1 %
56 KiB			1 %	1 %
60 KiB			2 %	2 %
64 KiB	22 %	22 %	20 %	20 %
128 KiB	3 %	3 %	2 %	2 %
256 KiB	2 %	2 %	2 %	2 %

Table 14 – IO Transfer Size within the Hot Band IO Profile for 4 KiB Native Devices

Xfer	Streaming Write	Streaming Read	Uniform	Hot Band
512 B				
1 KiB				
4 KiB	29 %	29 %	31 %	31 %
8 KiB	33 %	33 %	31 %	31 %
16 KiB	6 %	6 %	5 %	5 %
32 KiB	5 %	5 %	5 %	5 %
48 KiB			1 %	1 %
56 KiB			1 %	1 %
60 KiB			2 %	2 %
64 KiB	22 %	22 %	20 %	20 %
128 KiB	3 %	3 %	2 %	2 %
256 KiB	2 %	2 %	2 %	2 %

7.2.12.2 File Access IO Profiles

7.2.12.2.1 Overview of Workloads

The SPEC SFS® 2014 Benchmark is the workload generator used for all active tests. SPEC SFS® 2014 workloads are used to exercise the product under test during all active tests.

There are four workloads executed in the File Access sequence of tests. These are the VDA, Database, VDI and SW build workloads. The warmup interval of these workloads may be adjusted by the test sponsor to ensure stable operation throughout the measurement period.

For more information about the workloads, see the SPEC SFS® 2014 User's Guide[3].

7.2.12.2.2 Video Data Acquisition (VDA) Workload

The workload generally simulates applications that store data acquired from a temporally volatile source (e.g., surveillance cameras). A stream refers to an instance of the application storing data from a single source (e.g., one video feed). The storage admin is concerned primarily about maintaining a minimum fixed bit rate per stream and secondarily about maintaining the fidelity of the stream. The goal of the storage admin is to provide as many simultaneous streams as possible while meeting the bit rate and fidelity constraints.

The business metric for the workload is STREAMS. The workload consists of two workload objects: VDA1 (data stream) and VDA2 (companion applications). Each stream corresponds to a roughly 36 Mb/s bit rate, which is in the upper range of high definition video.

7.2.12.2.3 Database (DATABASE) Workload

This workload represents the typical behavior of a database. The complete workload is a mixture of DB_TABLE and DB_LOG workloads. The DB_TABLE workload is the database component, and DB_LOG represents the log writer component of a database operation.

7.2.12.2.4 Virtual Desktop Infrastructure (VDI) Workload

This workload simulates a steady-state high-intensity knowledge worker in a VDI environment that uses full clones. This workload does not simulate a linked-clone environment. This is the behavior that was seen in traces between the hypervisor and storage when the VM's were running on ESXi, Hyper-V, KVM and Xen environments.

7.2.12.2.5 Software Build (SW Build) Workload

The software build type workload is a classic meta-data intensive build workload. This workload was derived from analysis of software builds, and traces collected on systems in the software build arena. Conceptually, these tests are similar to running Unix 'make' against several tens of thousands of files. The file attributes are checked (metadata operations) and if necessary, the file is read, compiled, then data is written back out to storage.

7.2.13 IO Modes

All IO requests on a product under test shall be classified as either:

- Foreground IO, an externally-initiated, application-level request for data transfer between the load generator and the product under test, or;

- Background IO, a system-initiated request for data transfer within the product under test.

7.2.14 Average Response Time

The average response time for a test or test phase i, $RTA_i(T)$, is calculated over a specified time interval T in seconds.

7.2.15 Average Power

The average power for a test or test phase i, $PA_i(T)$, is the arithmetic average of sampled power measurements taken over a specified time interval T in seconds, as illustrated in Equation 7-2.

Equation 7-2: Average Power

$$PA_i(T) = \frac{\sum W_s}{n}$$

Where:
- $PA_i(T)$ is the average power during test or test phase i, taken over a time interval of T seconds;
- W_s is power in watts measured at each sampling interval s taken during the time interval T;
- n is the number of samples gathered by the power meter during the time interval T.

7.2.16 Operations Rate

The operations rate for a test or test phase i, $O_i(T)$, is a measure of the average rate of completed work over a specified time interval T. It is different for random workloads and sequential workloads.

For random workloads (i.e., random read, random write, and a mix of random read and random write), the operations rate is the average rate of IO operation completions during time interval T.

For sequential workloads (i.e., sequential read or sequential write) and file access workloads, the operations rate is the average rate of data transfer in mebibytes per second (MiB/s) within time interval T.

To provide a uniform basis for the metrics of a SNIA Emerald™ Power Efficiency Measurement, these two different measures of operations rate are both represented by $O_i(T)$.

7.2.17 Periodic Power Efficiency

The periodic power efficiency for a test or test phase i, $EPP_i(T)$, is the ratio of operations rate over a specified time interval T and the average power during the same time interval T as illustrated by Equation 7-3.

Equation 7-3: Periodic Power Efficiency

$$EPP_i(T) = \frac{O_i(T)}{PA_i(T)}$$

Where:
- EPP_i (T) is the periodic power efficiency during test or test phase i, taken over a time interval of T seconds;
- $O_i(T)$ is the operations rate during test or test phase i, taken over the same time interval of T seconds;
- $PA_i(T)$ is the average power during test or test phase i, taken over the same time interval of T seconds.

7.2.18 Measurement Interval

All tests and test phases state a minimum duration for their measurement interval. A measurement interval is a subset of a test or test phase during which the data underlying a specific metric or calculation is gathered.

A test sponsor shall ensure that a test's metric is stable throughout the measurement interval by analyzing the test output as specified in subclause 7.2.19, and may extend some or all conditioning tests or test phases to achieve stability of that test's metric.

7.2.19 Metric Stability Assessment

7.2.19.1 General

Metric M stability is assessed over the specified measurement interval defined in subclause 7.2.18, where M represents a selected performance/watt metric (e.g., periodic power efficiency $EPP_i(T)$; see subclause 7.2.17. Particular test phases result in the generation of a continuous sequence of J fixed period samples of metric M values beginning after a specified warm-up period. The stability assessment evaluates a candidate sequence of K consecutive samples of metric M values where K ≤ J. The metric M sample period T is fixed by the test; hence the measurement interval is equal to K * T (e.g., T = 60 seconds and K = 30 yielding a 30 min measurement interval producing 30 samples of metric *EPPi(60)*). Each stability assessment shall supply a value for K.

The stability of metric M is assessed by testing the flatness of the selected candidate sequence of K metric M values. It is recognized that this sequence, while otherwise flat, can have amplitude dispersal. Hence, the stability assessment is comprised of two tests:

1. A maximum allowed slope of a linear approximation of the K metric M values;
2. A smoothing function applied to the same K metric M values followed by comparison to a defined base reference value and a specified validity range.

The slope in (1) is determined from a least squares linear fit of the K metric M values.

The smoothing function in (2) consists of a weighted moving average S(M) of the same K metric M values. S(M) is based on a weighted average of present M and prior S(M) values.

The sequence of K metric M values shall meet the stability requirements of both tests to be considered stable.

7.2.19.2 Assessment Method

Assessing the stability of metric M values consists of the following test flow:

1. Select the first sample immediately after the warm-up period and set this point as N = 0. There shall be at least J = K metric M samples after this point;
2. Perform both the least squares linear fit test on the ($M_{N+1}..M_{N+K}$) sequence as described in 7.2.19.3 and the weighted moving average test on the same ($M_{N+1}..M_{N+K}$) sequence as described in 7.2.19.4;
3. If either test fails, N is incremented and start again at (2) above as long as incremented N + K ≤ J;
4. If full range J is exhausted without both tests passing, the metric M is deemed not stable;
5. If both tests pass, the metric M is deemed stable.

Once a stable sequence has been found, the assessment may continue looking for more stable sequences by continuing to increment N until incremented N + K = J.

7.2.19.3 Least Squares Linear Fit Test

The least squares linear fit test is performed over the K metric M values ($M_{N+1}..M_{N+K}$). The least squares linear fit calculation over this sequence returns a fitted line slope value and an intercept value as shown in Equation 7-4.

Equation 7-4: Least Squares Linear Fit Calculation

$$Slope(M) = (\sum_{n=1}^{K} M_{N+n}(12n - 6K - 6))/(K(K-1)(K+1))$$

$$Int(M) = (\sum_{n=1}^{K} M_{N+n})/K - Slope(M)(K+1)/2$$

$$Y = (n \cdot Slope(M)) + Int(M)$$

Where:
- Slope(M) is the slope of the least squares fit line;
- Int(M) is the intercept of the least squares fit line;
- M_{N+n} is the $N+n$-th sample value of metric M.

To be stable, the value Y in Equation 7-4 of the least squares fit line at sample point n = K shall not be more than 5 % different than the value Y in Equation 7-4 of the least squares fit line at sample point n = 1.

7.2.19.4 Weighted Moving Average Test

The weighted moving average test is performed over the K metric M values ($M_{N+1}..M_{N+K}$), as defined in Equation 7-5.

Equation 7-5: Weighted Moving Average Calculation

$$Base(M) = (\sum_{n=1}^{K} M_{N+n})/K$$

$$S_n(M) = wM_{N+n} + (1-w)S_{n-1}(M), for\ n = 1..K$$

$$S_0(M) = Base(M)$$

Where:
- Base(M) is the base reference value used to establish stability;
- M_{N+n} is the $N+n$-th sample value of metric M;
- $S_n(M)$ is the n-th value of the weighted moving average of metric M;
- w is the factor that determines how much weight a new sample has in the moving average $S_n(M)$ - each stability assessment shall supply a value for w;
- $S_0(M)$ is the initial value of the weighted moving average.

To be stable, each value of $S_n(M)$ for n = 1..K shall not differ from base reference Base(M) by more than ± 5 %.

7.3 Block Access Online and Near-Online Tests

7.3.1 Solution Under Test Configuration and Protocol Requirements

If a product under test is based on an industry standard storage protocol, the product under test shall adhere to all mandatory parts of the protocol specification.

All transport protocols used by load generator threads to access the product under test shall be disclosed. Transport protocols not involved in the test data flow may be disclosed.

7.3.2 Pre-fill Test

7.3.2.1 Overview

The pre-fill test is intended to provide a working data set on the storage system to be used with the other testing sets.

7.3.2.2 Procedure

- The product under test shall have a minimum of 50 % of the physical formatted storage pre-filled with the IO pattern described in Table 15 before continuing to the Conditioning Test;

- The data used to pre-fill the product under test shall be a 2:1 compression pattern, when compressed by gzip (see Annex J), using the latest version and compression algorithm level 6;

- Each IO stream shall issue each IO request synchronously, with each subsequent IO request issued immediately following the completion of its predecessor;

- The pre-fill test is defined to begin when the first request from the IO streams is issued by the load generator;

- IO transfer size and alignment is picked by the test sponsor;

- The pre-fill test shall last for at least the amount of time required to reach the pre-fill requirements of the product under test;

- The following test phases shall only access the allocated storage space that has pre-filled data.

Table 15 – Pre-fill Test IO Profile

IO Profile	Read/Write Percentage	IO Intensity	Access Pattern	Data Pattern
Sequential Write	0/100	100	Sequential	2:1 compression

7.3.3 Conditioning Test

7.3.3.1 Overview

The conditioning test is intended to provide a uniform initial condition for subsequent measurement(s) and to:

- Demonstrate the ability of the product under test to process IO requests;

- Ensure that each storage device in the product under test is fully operational and capable of satisfying all supported requests within the constraints of the taxonomy classification identified for the product under test;

- Achieve typical operating temperature;

- Optionally, to provide a time period to allow the product under test to monitor/learn the characteristics of the workload and subsequent tiered storage data migration.

The limitations of timely test execution make it impossible to remove all variability between results, or to provide complete pre-testing stability. Test sponsors are encouraged to minimize the impact of certain long-duration or infrequent changes to the product under test that can impact test results, including:

- Cache stability;
- Maintenance cycles.

7.3.3.2 Procedure

The conditioning test shall begin when the first request from the IO streams is issued by the load generator.

Each IO stream shall issue each IO request synchronously, with each subsequent IO request issued immediately following the completion of its predecessor.

The load generator shall initiate a number of independent IO streams equal to or greater than the number of LUNS made available to the workload generator by the product under test.

Each IO stream shall issue a sequence of IO requests with the IO profile shown in Table 16.

Table 16 – Online and Near-Online Testing: Conditioning Test IO Profiles

IO Profile	Read/Write Percentage	IO Intensity	Access Pattern
Hot banding	See Table 12	100	See Table 12

If the product under test includes functionality that requires changes to the IO profile defined in Table 16 in order to meet the intent stated in subclause 7.3.3.1, the changes shall be disclosed.

The conditioning test shall last for a minimum of 12 h any part of which may be used for workload monitoring/learning of a tiered configuration.

7.3.3.2.1 Tiered Storage Configuration Data Migration Phase

If tiered storage is deployed, a migration phase may be included immediately following the monitoring/learning phase to allow data relocation into the appropriate tiers. If the migration phase is included, the IO intensity during the migration phase shall be at least 25 % per the IO profile defined in Table 16. If auto-tiering is deployed, it shall be disclosed.

7.3.3.3   Data to be Collected

During the conditioning test, the following data shall be collected at successive 1 min intervals:

- Average response time to complete an IO, $RTA_{sc}(60)$ (see subclause 7.2.14), reported to a precision of 1 ms.

7.3.3.4   Validity

The conditioning test shall satisfy the following conditions in order to be considered valid:

- All IOs issued shall complete successfully;

- During the final four hours of the conditioning test, $RTA_{sc}(14\ 400)$ shall not exceed 20 ms. This requirement does not apply to Near-Online systems.

7.3.4   Block Access Active Test

7.3.4.1   Overview

The active test collects data for test phases defined in subclause 7.3.4.2.

The active test shall begin immediately following the conditioning test.

7.3.4.2   Test Phases

Table 17 defines the sequence of test phases, and their associated IO profiles, for this test.

Table 17 – Online and Near-Online Testing: Active Test Phase IO Profiles

	IO Profile (Test Phase i)	IO Size (KiB)	Read/Write Percentage	IO Intensity	Transfer Alignment (KiB)	Access Pattern
1	Hot Band Workload (i=HB)	See Table 12	See Table 12	100	See Table 12	See Table 12
2	Random Write (i=RW)	8	0/100	100	8	Random
3	Random Read (i=RR)	8	100/0	100	8	Random
4	Sequential Write (i=SW)	256	0/100	100	256	Sequential
5	Sequential Read (i=SR)	256	100/0	100	256	Sequential

7.3.4.3   Procedure

The active test is composed of a set of test phases, which shall be executed as an uninterrupted sequence, in the order presented in Table 17.

Within the active test, each test phase shall begin when the first request from its IO streams is issued by the load generator.

Each test phase shall launch a test-sponsor-selected number of independent IO streams.

Each IO stream shall issue its IO requests synchronously, with each subsequent IO request issued immediately following the completion of its predecessor. All IO operations shall transfer data, either reading or writing. The IO stream does not include any idle or "think time."

Each IO stream shall issue a sequence of IO requests matching the IO profile defined for the current test phase in Table 17.

Within the active test, each test phase shall last for a minimum of 40 min, comprised of a minimum 10 min warm-up period followed by a minimum 30 min measurement interval.

7.3.4.4 Data to be Collected

During a given test phase, the following data shall be collected at successive 1 min (T = 60 s) intervals:

- Average response time to complete an IO, $RTA_i(60)$ (see 7.2.14 reported to a precision of 1 ms;

- Average power, $PA_i(60)$ (see 7.2.15);

- Operations rate, $O_i(60)$ (see 7.2.16).

7.3.4.5 Validity

Each test phase execution shall satisfy the following conditions in order to be considered valid:

- The access pattern supplied by the load generator shall match the IO profile selected for the test phase from Table 17;

- All IOs issued in a test phase complete successfully;

- The $EPP_i(60)$ (see subclause 7.2.17) shall be stable (see subclause 7.2.19) throughout the measurement interval (see subclause 7.2.18). The test sponsor may use any consecutive 30 min interval that is found to be stable as the measurement interval for the purposes of calculating the primary metric for that phase (see subclause subclause 7.2.18). A value of 0,1 shall be used for the weighing factor w and a value of 30 shall be used for K in subclause 7.2.19;

- Each $RTA_{HB}(60)$, $RTA_{RW}(60)$, and $RTA_{RR}(60)$ within the measurement interval, shall not exceed 80 ms. This requirement does not apply to Near-Online systems;

- $RTA_{HB}(1\,800)$, $RTA_{RW}(1\,800)$ and $RTA_{RR}(1\,800)$, based on the measurement interval, shall not exceed 20 ms. This requirement does not apply to Near-Online systems.

All COM functionality active during the capacity optimization test (see subclause 7.6) may be disabled at the discretion of the test sponsor during the active test.

7.4 File Access Online and Near-Online Active Test

7.4.1 General

This document specifies how the SPEC SFS® 2014 Benchmark is to be run as the workload generator for measuring and reporting storage power efficiency results. These rules have been adapted by the SNIA from the run rules for the SPEC SFS® 2014 Benchmark as established by the SPEC. The rules are intended to ensure that results generated with this suite are meaningful, comparable to other generated results, and are repeatable.

The active tests are managed through the execution of the SfsManager script, using a File Access IO Load Driver Configuration File as input. Refer to the *User Guide for the SNIA Emerald™ Power Efficiency Measurement Specification V3*[13] for details.

All loadpoints of a given workload shall be run as an uninterrupted sequence except that removal of files is allowed between workloads.

The product under test may be rebooted between workloads, but shall not be rebooted between loadpoints of a workload.

Each workload constitutes a test phase. Each workload produces 10 results, each for a different level of intensity. The test sponsor shall choose the most favorable of those results as the Measurement Interval to process and use for validation and calculation of the primary metric for that test phase.

7.4.2   Solution Under Test Configuration Requirements

7.4.2.1   Software Requirements

7.4.2.1.1   Solution Under Test Software

The solution under test shall have sufficient software installed to be able to access stable storage and writing of data to stable storage shall be enabled.

7.4.2.1.2   Workload Generator Source Code Changes

The SPEC SFS® 2014 Benchmark shall be used in an unmodified form as provided by SPEC, with the exception that errata patches and service packs for the SPEC SFS® 2014 Benchmark provided by the Storage Performance Evaluation Corporation[5] may be applied to the workload generator source code.

Note: By SPEC policy, these errata patches and service packs are performance-neutral.

7.4.2.2   Storage Solution Configuration and Protocol Requirements

7.4.2.2.1   Storage protocol requirements

If a product under test is based on an industry standard storage protocol, for example, NFS[10] or SMB[11], the product under test shall adhere to all mandatory parts of the protocol specification.

The workload generator automatically conducts a minimal functional validation test at the start of every test iteration, and will halt execution if the validation test fails.

All transport protocols used by load generator threads to access the product under test shall be disclosed. Transport protocols not involved in the test data flow may be disclosed.

7.4.2.2.2   Load Generator configuration requirements

The load generators in the solution under test shall be homogenous with respect to their operating system type, version, and patch level.

7.4.2.2.3   Stable Storage Requirements

Data written to the API and acknowledged as stable shall be in stable storage when acknowledged.

See Annex C for a full definition of stable storage.

7.4.2.3   Test Execution Requirements

7.4.2.3.1   General

This section details the requirements governing how the test is to be executed.

7.4.2.3.2   Valid methods for test execution

The test shall be executed by using the SfsManager on the prime client (see *SPEC SFS® 2014 User's Guide[3]*).

7.4.2.3.3   File System Creation and Configuration

The components of the solution that hold the data and/or metadata for the file systems under test shall follow the stable storage requirements detailed in subclause 7.4.2.2.3.

The file systems of the product under test may be re-initialized between workloads.

Note: The load generator re-initialization of data between tests might have an initializing effect on the underlying storage, based on the design of the product under test.

7.4.2.3.4   Data Point Specification Requirements

The result of test execution is the set of data points for the product under test which are used to calculate the test metrics. The measurement of all data points shall be made within a single test execution run, starting with the lowest requested load level and proceeding to the highest requested load level.

Test results shall be based on at least 10 load points (excluding a business metric of zero). The load points should be as uniformly spaced as possible. Each load point shall be within 30 % of its nominal uniformly-spaced value. The nominal interval spacing is the maximum requested load divided by the number of requested load points. This means the distance between zero and the first requested load point shall also fall within the nominal interval spacing. The product under test shall support a load of a business metric of at least 10.

Any invalid data points will invalidate the entire test. A data point is invalid if the workload generator reports it as invalid.

The following types of changes shall not be made to the solution under test during execution of the test or between data point collection:

- Manual change of the configuration of either the product under test or the testbed configuration;

- Reboots of the load generator or the product under test;

- File system initialization (e.g., "newfs/format").

If any requested load level or data point is rerun for any reason, the entire test execution shall be restarted, i.e., the series of requested load levels repeated in whole.

7.4.2.3.5   SPEC SFS® 2014 Benchmark Modifiable Parameters

7.4.2.3.5.1   General

Each workload has a number of parameters which are configurable. These configuration parameters are set by using a File Access IO Load Driver Configuration File (see Annex G) on the prime client (see *SPEC SFS® 2014 User's Guide[3]*). Parameters outside of the set specified in subclause 7.4.2.3.5 shall not be modified.

Parameters which may be modified for test execution are indicated in subclause 7.4.2.3.5.

7.4.2.3.5.2   BENCHMARK

Name of the workload to run. Valid values are: DATABASE, SWBUILD, VDA, or VDI.

7.4.2.3.5.3   LOAD

Each workload has an associated business metric as a unit of workload. The magnitude of the workload to run is specified with the LOAD parameter in units of the workload's business metric. Valid values for LOAD are either a starting number or a list of values, all positive integers. If a single value is specified, it is interpreted as a starting value and used in conjunction with INCR_LOAD and NUM_RUNS. The Table 18 shows the name for the business metric corresponding to each workload type.

Table 18 – Business Metrics for Workload Type

Workload	Business Metric (LOAD parameter)
DATABASE	DATABASES
SWBUILD	BUILDS
VDA	STREAMS
VDI	DESKTOPS

If a list of values if specified, at least 10 uniformly spaced business metric load values shall be specified for valid test execution.

Note: Some rules of thumb for the resources required per business metric load unit for the different workloads:

Minimum storage capacity needed for product under test per business metric load unit:

DATABASE = 24 GB per DATABASE

SWBUILD = 5 GB per BUILD

VDA = 24 GB per STREAM

VDI = 12 GB per DESKTOP

Minimum load generation RAM needed per business metric load unit:

DATABASE = 55 MiB per LOAD increment

SWBUILD = 400 MiB per LOAD increment

VDA = 10 MiB per LOAD increment

VDI =   8 MiB per LOAD increment

7.4.2.3.5.4 INCR_LOAD
Incremental increase in load for successive data points in a test run. This parameter is used only if LOAD consists of a single (initial) value. To ensure equally spaced points, the value of LOAD and INCR_LOAD shall be equal.

7.4.2.3.5.5 NUM_RUNS
The number of load points to run and measure (minimum of 10). This parameter is used only if INCR_LOAD is specified.

7.4.2.3.5.6 CLIENT_MOUNTPOINTS
The list of local mount points, local directories, or shares, to use in the testing. The value of CLIENT_MOUNTPOINTS can take several different forms:

- UNIX style: client:/exportfs1 client:/exportfs2 …:

  - Used for local storage or mounted network shares;

- Windows style: client:\\server\exportfs1 client:\\server\exportfs2 …;

- Use a file that contains the mount points: mountpoints_file.txt:

  - For examples of the format for this external file, see the *SPEC SFS® 2014 User's Guide*[3].

The business metric values are spread among the client mount points in the following way. If the number of items N in the CLIENT_MOUNTPOINTS list is greater than the business metric value L (the current value for LOAD), then the first L items from the list are used, one business metric value per client/mountpoint. If L > N, then the N + 1 business metric value will wrap around to the beginning of the list and allocation proceeds until all L business metrics have been allocated, wrapping around to the beginning of the list as many times as is necessary.

Reminder: If using a Windows-based load generator, the prime client (see *SPEC SFS® 2014 User's Guide*[3]) must not be listed in the CLIENT_MOUNTPOINTS list.

7.4.2.3.5.7 EXEC_PATH
The full path to the SPEC SFS® 2014 executable. Currently the executable is called netmist for POSIX systems and netmist_pro.exe for Windows systems. The same path will be used on all clients, so the executable must be at the same path on all clients.

7.4.2.3.5.8 USER
The user account name, which must be configured on all clients, to be used for the test execution. To specify a domain, prefix the user with the domain name, separated by a backslash, e.g., DOMAIN\User33

7.4.2.3.5.9 WARMUP_TIME
The amount of time, in seconds, that the test will spend in WARMUP before initiating the measurement ("RUN") phase. The minimum is 300 s (5 min). The maximum value is 604 800 s (7 d).

7.4.2.3.5.10 IPV6_ENABLE
Set to "1" or "Yes" when IPv6 will be used.to communicate with other load generator processes.

7.4.2.3.5.11 NETMIST_LOGS
Set the path at which netmist will store client log files. The same path will be used on all clients. If this path is not set, /tmp/ or c:\tmp\ will be used.

7.4.2.3.5.12 PASSWORD
The password for the user specified in USER. (Only applicable when running on Windows platforms)

7.4.2.4 File Access Active Test Sequence and corresponding Business Metrics

Table 19 shows the business metrics for various workloads in the File Access Active Test Sequence.

Table 19 – Business Metrics for File Access Active Test Sequence

Workload	Business Metric (LOAD parameter)	Primary Metric
DATABASE	DATABASES	EP_{DB}
SWBUILD	BUILDS	EP_{SWB}
VDA	STREAMS	EP_{VDA}
VDI	DESKTOPS	EP_{VDI}

7.4.2.5 Data to be Collected

During a given test phase, the following data shall be collected at 10-second intervals:

- The average throughput, $O_i(10)$, from the sFlow® agent(s) in MiB/s;
- Average power, $PA_i(10)$ (see subclause 7.2.15).

7.4.2.6 Validity

Each test phase execution shall satisfy the following conditions in order to be considered valid:

- All IOs issued in a test phase complete successfully;

- The $EPP_i(10)$ (see subclause 7.2.17) shall be stable (see subclause 7.2.19) throughout the measurement interval (see subclause 7.2.18). The test sponsor may use any consecutive 300 s interval that is found to be stable as the measurement interval for the purposes of calculating the primary metric for that phase (see subclause 7.2.18). A value of 0,1 shall be used for the weighing factor w and a value of 30 shall be used for K in subclause 7.2.19.

A COM function active during the capacity optimization test (see subclause 7.6) may be disabled at the discretion of the test sponsor during the active test.

7.5 Block and File Access Ready Idle Test

7.5.1.1 Overview

The ready idle test collects data for the ready idle metric.

The ready idle test shall begin immediately following the active test.

7.5.1.2 Procedure

No foreground IO shall be initiated on the product under test during the ready idle test other than that required to satisfy the instrumentation requirements in subclause 7.2.9.

The test sponsor shall select a duration of at least two hours for the ready idle test. The test sponsor shall use the final two hours of the test as the measurement interval for the purposes of calculating the average power for the ready idle test primary metric (see subclause 8.4.1).

7.6 Block and File Access Capacity Optimization Test

7.6.1 Overview

This section defines qualitative heuristics for validating the existence and activation of COMs that are present on the product under test. Each heuristic is a simple pass/fail test, intended only to verify the presence and activation of a particular capacity optimization method.

7.6.2 Testing Requirements

The heuristics assess the impact on free space to determine whether or not a COM is present and active. The assessments rely on:

- FS_{sot}: free space at the start of a test;

- FS_{eot}: free space at the end of a test;

- S_{ds}: size of a data set (see subclause 7.6.3);

- I_{com} the impact of the COM on overall space utilization within the heuristic. The precise formulation of this value is defined by each heuristic.

For the purpose of the COM tests, free space, as reported by the product under test, shall be used.

Test sponsors shall select which of the heuristics they wish to execute. Only heuristics which are executed can be marked as passed (see subclause 8.7.1).

If test sponsors choose to execute a heuristic, they shall execute all of its steps in sequence. The testing of storage equipment may have minimal interruptions between heuristic COMs as needed to

set up for each selected heuristic test. Configuration changes shall not be made between heuristic tests.

Storage device(s) shall not be added or removed, nor changed in state (taken on or offline, made a spare or incorporated, etc.) and RAID groups shall not be changed during any particular COM test. In the event of an automated disk failure and subsequent RAID rebuild at any time during a test, the test shall be restarted after the rebuild has completed and the failed disk has been replaced per manufacturers guidelines for installed and working systems.

Some of the following subclauses use the term "container" meaning a collection of logical blocks, e.g., a LUN or file system.

7.6.3   Generating Data Sets

Most heuristics require the generation and use of specific data sets as part of their existence test. There are three different data set categories depending on the needs of the particular heuristic:

- Completely irreducible;

- Dedupable but not easily compressible;

- Compressible but not dedupable.

Note: The required exclusivity of compressible and dedupable data sets comes from situations where certain systems do not have the ability to individually disable other COMs features during a particular heuristic test.

Each data set is to be approximately 2 GB in size and shall be generated by the COM Test Data Set Generator (see Annex I). Each data set will be created in a directory named by the user. There are numerous files in each directory, and the order in which they are presented depends on the operating system where the program is run.

Table 20 lists the data sets; these can be represented by a single file or by a directory (folder).

Table 20 – Data Sets

Category	Generator Output	Comments
Irreducible	Directory Name: Irreducible Content: *filename*.dat	A data set that is neither compressible nor dedupable. Use different salt values as needed to produce multiple non-duplicated irreducible data sets.
Compressible	Directory Name: Compressible Content: *filename*.dat	A data set that is compressible but not dedupable. Supports multiple compression methods.
Dedupable	Directory Name: Dedupable Content: *filename*.dat	A data set that is dedupable but not easily compressible. Supports multiple data deduplication methods.

7.6.4   Delta Snapshot Heuristics

7.6.4.1   General

Delta snapshots in a storage system can be detected using a straightforward algorithm:

1. Query the free space before taking a snapshot;

2. Attempt to create a snapshot;

3. In the case of a writeable snapshot, write something to the snapshot;

4. Query the free space after that snapshot to determine whether significant storage space has been used.

Note: Read-only and writeable delta snapshots are treated separately so that systems that only do read-only snapshots get credit for them.

7.6.4.2   Heuristic 1: Read-only delta snapshots

The method varies according to where the product under test places snapshots. Follow steps 1 or 2, and then proceed to step 3:

1. For a product under test that places snapshots in separate containers:

    a. On the product under test, create two containers, each 15 GB, with a tolerance of ± 10 %, in size. The amount of actual physical storage that is committed as a result of creating the containers will vary by product under test, and should be determined by the test sponsor in trials before a formal run of the test;

    b. Mount the first container on a host, via any chosen protocol;

    c. Determine the amount of free space FS_{sot} available on the product under test as seen by the product under test;

    d. Write the irreducible data set to the first container;

    e. Perform a read-only delta snapshot of the first container and expose it through the second container, disabling any background copying mechanism. As an example, the snapshot of lun1 may be exposed as lun2;

    f. Perform whatever steps are necessary to mount the second container as a file system. Open a small file on this file system (i.e., one of the files in the irreducible data set), read some data from it, and close the file. Confirm that the file has been successfully read;

2. For a product under test that places snapshots on the originating container:

    a. On the product under test, create a container of 15 GB, with a tolerance of ± 10 %, in size. The amount of actual physical storage that is committed as a result of creating the container will vary by product under test, and should have been determined by the test sponsor in trials before a formal run of the test;

    b. Mount the container on the host via any chosen protocol;

    c. Determine the amount of free space FS_{sot} available on the container as seen by the product under test;

    d. Write the irreducible data set to the container;

    e. Perform a read-only snapshot of the container, disabling any optional background copying mechanism;

    f. Perform whatever steps are necessary to mount the container as a file system. Open a file in the snapshot, read some data from it, and close the file. Confirm that the file portion has been successfully read;

3. Determine the amount of free space, FS_{eot}, available on the container containing the snapshot as seen by the product under test. Calculate the space required for the snapshot $I_{com} = FS_{sot} - FS_{eot}$.

If I_{com} is less than 2,5 GB, then the product under test shall be considered to have passed the test.

7.6.4.3   Heuristic 2: Writeable delta snapshots

This method varies according to where the product under test places snapshots. Follow steps 1 or 2, and then proceed to step 3:

1. For a product under test that places snapshots in separate containers:

    a. On the product under test, create two containers, each 15 GB, with a tolerance of ± 10 %, in size. The amount of actual physical storage that is committed as a result of creating the containers will vary by product under test, and should have been determined by the test sponsor in trials before a formal run of the test;

    b. Mount the first container on a host, via any chosen protocol;

    c. Determine the amount of free space FS_{sot} available on the product under test as seen by the product under test;

    d. Write the irreducible data set to the first container;

    e. Perform a writeable snapshot of the first container and expose it through the second container, disabling any optional background copying mechanism. As an example, the snapshot of lun1 may be exposed as lun2;

    f. Perform whatever steps are necessary to mount the second container as a file system. Open a small file on this file system (i.e., one of the files in the irreducible data set), write a few characters to it, and close the file. Confirm that the file has been successfully written with its new contents;

2. For an product under test that places snapshots on the originating container:

    a. On the product under test, create a container of 15 GB, with a tolerance of ± 10 %, in size. The amount of actual physical storage that is committed as a result of creating the container will vary by product under test, and should have been determined by the test sponsor in trials before a formal run of the test;

    b. Mount the container on the host via any chosen protocol;

    c. Determine the amount of free space FS_{sot} available on the container as seen by the product under test;

    d. Write the irreducible data set to the container;

    e. Perform a writeable snapshot of the container, disabling any optional background copying mechanism;

    f. Perform whatever steps are necessary to mount the container as a file system. Open a small file in the snapshot (i.e., one of the files in the irreducible data set), write a few characters to it, and close the file. Confirm that the file has been successfully written;

3. Determine the amount of free space FS_{eot} available on the container containing the snapshot as seen by the product under test. Calculate the space required for the snapshot $I_{com} = FS_{sot} - FS_{eot}$.

If I_{com} is less than 2,5 GB and the small file was successfully written onto the writeable delta snapshot destination, then the product under test shall be considered to support writable delta snapshots.

7.6.4.4 Thin Provisioning Heuristics

7.6.4.4.1 General

The goal of this heuristic is not to highlight differences of thin provisioning implementations between vendors; it is to be used simply to ensure that the product under test does have some sort of thin provisioning capability.

7.6.4.4.2 Thin Provisioning Heuristics – Block Access Device

A test sponsor seeking credit for thin provisioning shall:

1. Establish the total usable space or establish a pool of usable space as seen by the product under test;

2. Enable thin provisioning if not already enabled;

3. Request allocation of N LUNs of capacity M such that N · M is at least 20 % greater than the total (or pooled) usable space. This shall result in an allocation of all requested LUNs.

The product under test shall be considered to support thin provisioning if step 3 is successful.

7.6.4.4.3 Thin Provisioning Heuristics – File Access Device

A test sponsor seeking credit for thin provisioning shall:

1. Establish the total usable space or establish a pool of usable space as seen by the product under test;

2. Enable thin provisioning if not already enabled;

3. Create N file systems of capacity M such that N · M is at least 20 % greater than the total free space.

The product under test shall be considered to support thin provisioning if step 3 is successful.

7.6.4.5 Data Deduplication Heuristics

A test sponsor seeking credit for data deduplication shall:

1. On the product under test, create a container large enough in size to engage the product under test data deduplication mechanism. The amount of actual physical storage that is committed as a result of creating the container will vary by product under test, and should have been determined by the test sponsor in trials before a formal run of the test;

2. Perform whatever steps are necessary to make the container visible on the host from which tests are being run, and create and mount a local file system on that container;

3. Determine the amount of free space FS_{sot} available on the container as seen by the product under test;

4. For cases in which the 2 GB dedupable data set is sufficient, write the 2 GB dedupable data set to the container;

5. For cases in which a larger data set is required:
   a. Establish the necessary data set size in 2 GB increments;
   b. Write the 2 GB dedupable data set to the container;
   c. In the same container, write as many (N) 2 GB irreducible data sets as are required to meet system requirements for data deduplication. For each:
      1. Create a new directory for each data set;
      2. Invoke the COM Test Data Set Generator tool on each directory with the same "salt" value each time;

6. Wait a suitable amount of time as specified by the test sponsor for any non-inline data deduplication processes to have completed;

7. Determine the amount of free space FS_{eot} available on the container as seen by the product under test;

8. Calculate the fraction of formatted capacity saved by data deduplication $I_{com} = (1 − ((FS_{sot} - FS_{eot}) / ((N+1) · S_{ds})))$ where S_{ds} is the size of the single dedupable data set.

If I_{com} is greater than 0,1 then the product under test shall be considered to support data deduplication.

Note: Data set size is not necessarily important for the purposes of data deduplication detection. However, larger data sets may be necessary to activate existing data deduplication functionality. This heuristic allows the building of a larger data set, consisting of a single 2 GB dedupable data set and the option of a test sponsor-determined integer number of 2 GB irreducible data sets, collectively

used to demonstrate data deduplication capability. The addition of identical 2 GB irreducible data sets can themselves be dedupable so it is a requirement to generate each with the same "salt" value.

7.6.4.6   Compression Heuristics

Data set size might be important for the purposes of compression detection. However, larger data sets might be necessary to activate existing compression functionality. This heuristic allows the building of a larger data set, consisting of a single 2 GB compressible data set and the option of a test sponsor-determined integer number of 2 GB compressible data sets, collectively used to demonstrate compression capability. The addition of identical 2 GB compressible data sets can themselves be dedupable, therefore it is a requirement to generate each with a different "salt" value.

A test sponsor seeking credit for compression of primary storage shall:

1. On the product under test, create a container large enough in size to engage the product under test compression mechanism. The amount of actual physical storage that is committed as a result of creating the container will vary by product under test, and should have been determined by the test sponsor in trials before a formal run of the test;

2. Perform whatever steps are necessary to make the container visible on the host from which tests are being run, and create and mount a local file system on that container;

3. Determine the amount of free space FS_{sot} available on the container as seen by the product under test;

4. For cases in which the 2 GB compressible data set is sufficient:

   a. Write the 2 GB compressible data set to the container;

5. For cases in which a larger data set is required:

   a. Establish the necessary data set size in 2 GB increments;

   b. Write the 2 GB compressible data set to the container;

   c. In the same container, write as many additional (N) 2 GB compressible data sets as are required to meet system requirements. For each:

      1. Create a new directory for each data set;

      2. Invoke the COM Test Data Set Generator tool on each directory with a unique "salt" value each time;

6. Wait a suitable amount of time as specified by the test sponsor for any non-inline compression processes to have completed;

7. Determine the amount of free space FS_{eot} available on the container as seen by the product under test;

8. Calculate the fraction of formatted capacity saved by compression $I_{com} = (1 - ((FS_{sot} - FS_{eot}) / ((N+1) \cdot S_{ds})))$ where S_{ds} is the size of the single compressible data set.

If I_{com} is greater than 0,1 then the product under test shall be considered to support compression.

7.7   Removable Media Library Testing

7.7.1   Product Under Test Configuration Requirements

Removable media testing has additional configuration requirements.

7.7.1.1 Drive Counts

For removable media testing the product under test shall be configured to the requirements listed in Table 21. This table lists the number of tape/optical drives that shall be used during the Emerald testing. The drive counts are not allowed to change during the tests

Table 21 – Drive Counts

Category	Drive Count
Removable-1	Maximum supported
Removable-2	Maximum Supported
Removable-3	Maximum Supported
Removable-4	24
Removable-5	24

7.7.1.2 Drive-level Compression

If the product under test supports drive-level compression, it shall be disabled throughout the tests.

7.7.1.3 Tape Robotics

If the product under test contains robotics, they shall be enabled and ready to process a tape/optical manipulation commands throughout the tests.

7.7.2 Pre-fill Test

There are no pre-fill requirements for removable media library testing.

7.7.3 Conditioning Test

7.7.3.1 Overview

The conditioning test is intended to provide a uniform initial condition for subsequent measurement(s) and to:

- Demonstrate the ability of the product under test to process IO requests;
- Ensure that each storage device in the product under test is fully operational and capable of satisfying all supported requests within the constraints of the taxonomy classification identified for the product under test;
- Achieve typical operating temperature.

The limitations of timely test execution make it impossible to remove all variability between results, or to provide complete pre-testing stability. Test sponsors are encouraged to minimize the impact of certain long-duration or infrequent changes to the product under test that can impact test results, including:

- Cache stability;
- Maintenance cycles.

7.7.3.2 Procedure

The conditioning test shall begin when the first request from the IO streams is issued by the load generator.

Each IO stream shall issue each IO request synchronously, with each subsequent IO request issued immediately following the completion of its predecessor. The workload generator shall uniformly

distribute the required IO requests among the IO streams, such that the maximum number of IO requests serviced by an IO stream is no more than 10 % greater than minimum number of IO requests serviced by an IO stream.

The test sponsor shall ensure that all tape/optical drives are accessed during the conditioning test.

The workload generator shall initiate a number of independent IO streams equal to or greater than the number of tape/optical drives made available to the load generator by the product under test.

The test sponsor shall ensure that all tape/optical drives are:

- Loaded prior to the conditioning test;

- Rewound between the two phases of the conditioning test;

- Rewound at the end of the conditioning test.

The conditioning test shall consist of two phases, each lasting for a minimum of seven minutes shown in Table 22.

Table 22 – Removable Media Library Testing: Conditioning Test IO Profiles

IO Profile	IO Size (KiB)	Read/ Write Percentage	IO Intensity	Transfer Alignment (KiB)	Access Pattern
Sequential Write (i=C1)	256	0/100	100	256	Sequential
Sequential Read (i=C2)	256	100/0	100	256	Sequential

Each IO stream shall issue a sequence of IO requests satisfying the Sequential Write IO profile shown in Table 22 during the first phase of the conditioning test, and the Sequential Read IO profile shown in Table 22 during the second phase of the conditioning test.

If the product under test includes functionality that requires changes to the IO profile defined in Table 22 in order to meet the intent stated in subclause 7.7.3.1, the changes shall be disclosed.

7.7.3.3 Data to be Collected

During the conditioning test, the following data shall be collected at successive 1 min intervals:

- Average data rate reported for each drive, reported in MiB/s;

- Average power, $PA_i(60)$ (see 7.2.15);

- Operations rate, $O_i(60)$ (see 7.2.16).

7.7.3.4 Validity

The conditioning test shall satisfy the following conditions in order to be considered valid:

- All IOs issued shall complete successfully;

- The overall data rate for each drive present in the product under test for a given test phase shall be greater than or equal to 80 % of the maximum published data rate for that drive type.

7.7.4 Active Test

7.7.4.1 Overview

The active test collects data for the test phases defined below and is used to generate the power efficiency metrics.

The active test shall begin immediately following the conditioning test.

7.7.4.2 Test Phases

The active test is composed of two test phases, which shall be executed as an uninterrupted sequence, and separated by a rewind to Beginning of Tape (BOT), in the order presented in Table 23.

Table 23 – Removable Media Library Testing: Active Test Phase IO Profiles

IO Profile (Test Phase i)	IO Size (KiB)	Read/Write Percentage	IO Intensity	Transfer Alignment (KiB)	Access Pattern
Sequential Write (i=SW)	256	0/100	100	256	Sequential
Sequential Read (i=SR)	256	100/0	100	256	Sequential

Table 23 defines the sequence of test phases, and their associated IO profiles, used during this test.

7.7.4.3 Procedure

Within the active test, each test phase shall begin when the first requested IO is issued by the load generator.

Each IO stream shall issue a sequence of IO requests matching the profile defined for the current test phase in Table 23.

Each test phase shall launch a number of independent IO streams equal to the number of drives present in the product under test. All drives in the product under test shall be capable of processing IO requests at the start of a test phase.

Each IO stream shall issue its IO requests synchronously, with each subsequent IO request issued immediately following the completion of its predecessor. All IO operations transfer data, either reading or writing. The IO stream does not include any idle or "think time."

The first IO within a sequential IO Stream shall occur at Beginning of Tape (BOT). Each subsequent IO request shall be sent to and satisfied by the product under test in sequence using a transfer offset that satisfies Equation 7-6.

Equation 7-6: Sequential Transfer Offset

$$O_{n+1} = (O_n + S) \ MOD \ R$$

Where:
- O_n is an IO offset;
- S is the IO size;
- R is the formatted capacity of the product under test.

The workload generator shall uniformly distribute the required IO requests among the IO streams, such that the maximum number of IO requests serviced by an IO stream is no more than 10 % greater than minimum number of IO requests serviced by an IO stream.

The test sponsor shall ensure that all tape drives are accessed throughout the active test.

7.7.4.4 Data to be Collected

During a given test phase, the following data shall be collected at successive 1 min (T = 60 s) intervals:

- Average data rate reported for each drive, reported in MiB/s;

- Average power, $PA_i(60)$ (see subclause 7.2.15);

- Operations rate, $O_i(60)$ (see subclause 7.2.16).

7.7.4.5 Validity

Each test phase execution shall satisfy the following conditions in order to be considered valid:

- All IOs issued in a test phase complete successfully;

- The access pattern supplied by the workload generator shall match the IO profile selected for the test phase from Table 23;

- The $EPP_i(60)$ shall be stable (see subclauses 7.2.17 and 7.2.19), based on the 1 min data collected according to subclause 7.7.4.4, throughout the measurement interval;

- The $EPP_i(60)$ (see subclause 7.2.17) shall be stable (see subclause 7.2.19) throughout the measurement interval (see subclause 7.2.18). The test sponsor may use any consecutive 30-minute interval that is found to be stable as the measurement interval for the purposes of calculating the primary metric for that phase (see subclause 7.2.18). A value of 0,1 shall be used for the weighing factor w and a value of 30 shall be used for K in subclause 7.2.19;

- Each $O_{SW}(1\ 800)$ and $O_{SR}(1\ 800)$ rate for each drive present in the product under test for a given test phase shall be greater than or equal to 80 % of the maximum published data rate for that drive type.

7.7.5 Ready Idle Test

7.7.5.1 Overview

The ready idle test collects data for the ready idle metric

The ready idle test shall begin immediately following the active test.

7.7.5.2 Procedure

No foreground IO shall be initiated on the product under test during the ready idle test other than that required to satisfy the instrumentation requirements in subclause7.2.9.

The ready idle test shall begin after any loaded storage devices have been unloaded from the tape drives and any robotics activity has completed.

The test sponsor shall select a duration of at least two hours for the ready idle test. The test sponsor shall use the final two hours of the test as the measurement interval for the purposes of calculating the average power for the ready idle test primary metric (see 8.5.1).

7.7.6 Capacity Optimization Test

This document does not define a capacity optimization method test for the Removable Media Library taxonomy category.

7.8 Virtual Media Library Testing

7.8.1 Pre-fill

There are no pre-fill requirements for removable media library testing.

All writes by the workload generator shall be 8 bit random data.

7.8.2 Conditioning Test

7.8.2.1 Overview

The conditioning test is intended to provide a uniform initial condition for subsequent measurement(s) and to:

- Demonstrate the ability of the solution under test to process IO requests;

- Ensure that each storage device in the product under test is fully operational and capable of satisfying all supported requests within the constraints of the taxonomy classification identified for the product under test;

- Achieve typical operating temperature.

The limitations of timely test execution make it impossible to remove all variability between results, or to provide complete pre-testing stability test sponsors are encouraged to minimize the impact of certain long-duration or infrequent changes to the product under test that can impact test results, including:

- Cache stability;

- Maintenance cycles.

7.8.2.2 Procedure

The conditioning test shall begin when the first requested IO is issued by the load generator.

Each IO stream shall issue each IO request synchronously, with each subsequent IO request issued immediately following the completion of its predecessor.

The workload generator shall uniformly distribute the required IO requests among the IO streams, such that the maximum number of IO requests serviced by an IO stream is no more than 10 % greater than minimum number of IO requests serviced by an IO stream.

The test sponsor shall ensure that all storage devices are accessed at some time during the conditioning test.

The workload generator shall initiate a number of independent IO streams equal to or greater than the number of virtual drives made available to the load generator by the product under test.

The test sponsor shall ensure that all virtual drives are:

- Loaded prior to the conditioning test;

- Rewound between the two phases of the conditioning test;

- Rewound at the end of the conditioning test.

The conditioning test shall consist of two phases, as shown in Table 24, each lasting for a minimum of seven minutes.

Table 24 – Virtual Media Library Testing: Conditioning Test IO Profiles

IO Profile	IO Size (KiB)	Read/ Write Percentage	IO Intensity	Transfer Alignment (KiB)	Access Pattern
Sequential Write (i=C1)	256	0/100	100	256	Sequential
Sequential Read (i=C2)	256	100/0	100	256	Sequential

SNIA Emerald™ Power Efficiency Measurement Specification V3.0.3

Each IO stream shall issue a sequence of IO requests satisfying the Sequential Write IO profile shown in Table 24 during the first phase of the conditioning test, and the Sequential Read IO profile shown in Table 24 during the second phase of the conditioning test.

If the product under test includes functionality that requires changes to the IO profile defined in Table 24 in order to meet the intent stated in subclause 7.8.2.1, the changes shall be disclosed.

7.8.2.3 Data to be Collected

During the conditioning test, the following data shall be collected at successive 1-minute intervals:

- Average data rate reported for each drive, reported in MiB/s;
- Average power, $PA_i(60)$ (see subclause 7.2.15);
- Operations rate, $O_i(60)$ (see subclause 7.2.16).

7.8.2.4 Validity

The conditioning test shall satisfy the following conditions in order to be considered valid:

- All IOs issued shall complete successfully;
- The overall data rate for each drive present in the product under test for a given test phase shall be greater than or equal to 90 % of the maximum published native (without compression) data rate for that drive type.

7.8.3 Active Test

7.8.3.1 Overview

The active test collects data for the test phases defined below and is used to generate the power efficiency metrics.

The active test shall begin immediately following the conditioning test.

7.8.3.2 Test Phases

The active test is composed of a set of test phases, which shall be executed as an uninterrupted sequence, separated by a rewind to Beginning of Tape (BOT), in the order presented in Table 25.

Table 25 – Virtual Media Library Testing: Active Test Phase IO Profiles

IO Profile (Test Phase i)	IO Size (KiB)	Read/Write Percentage	Transfer Alignment (KiB)	Access Pattern
Sequential Write (i=SW)	256	0/100	256	Sequential
Sequential Read (i=SR)	256	100/0	256	Sequential

Table 25 defines the sequence of test phases, and their associated IO profile for this test.

7.8.3.3 Procedure

Within the active test, each test phase shall begin when the first request from its IO streams is issued by the load generator. During the active test, the product under test shall include a number of virtual drives that is sufficient to reach 90 % of maximum published data rate for the drive type being emulated. There shall be no change to the number of virtual drives in the product under test between the sequential write and sequential read test phases.

Each test phase shall launch a number of independent IO streams equal to the number of virtual drives present in the product under test. All virtual drives in the product under test shall be capable of processing IO requests prior to the start of a test phase.

Each IO stream shall issue its IO requests synchronously, with each subsequent IO request issued immediately following the completion of its predecessor. All IO operations shall transfer data, either reading or writing. The IO stream does not include any idle or "think time."

The first IO within an IO Stream shall use occur at Beginning of Tape (BOT). Each subsequent IO request shall be sent to and satisfied by the product under test in sequence using a transfer offset that satisfies Equation 7-7.

Equation 7-7: Sequential Transfer Offset

$$O_{n+1} = \left(O_n + S\right) \ MOD \ R$$

Where:
- O_n is an IO offset;
- S is the IO size;
- R is the formatted capacity of the product under test.

The workload generator shall uniformly distribute the required IO requests among the IO streams, such that the maximum number of IO requests serviced by an IO stream is no more than 10 % greater than minimum number of IO requests serviced by an IO stream.

The test sponsor shall ensure that all storage devices that comprise the formatted capacity are accessed at some time during the active test.

If the product under test supports drive-level compression, it shall be disabled throughout the active test.

7.8.3.4 Data to be Collected

During a given test phase, the following data shall be collected at successive 1 min (T = 60 s) intervals:

- Average data rate reported for each virtual drive in the product under test, reported in MiB/s;

- Average power, $PA_i(60)$ (see subclause 7.2.15);

- Operations rate, $O_i(60)$ (see subclause 7.2.16).

7.8.3.5 Validity

Each test phase execution shall satisfy the following conditions in order to be considered valid:

- All IOs issued in a test phase complete successfully;

- The access pattern supplied by the workload generator shall match the IO profile selected for the test phase from Table 25;

- The $EPP_i(60)$ (see subclause 7.2.17) shall be stable (see subclause 7.2.19) throughout the measurement interval (see subclause 7.2.18). The test sponsor may use any consecutive 30 min interval that is found to be stable as the measurement interval for the purposes of calculating the primary metric for that phase (see subclause 7.2.18). A value of 0.1 shall be used for the weighing factor w and a value of 30 shall be used for K in subclause 7.2.19;

- Each $O_{SW}(1\ 800)$ and $O_{SR}(1\ 800)$ rate for each drive present in the product under test for a given test phase shall be greater than or equal to 90 % of the maximum published data rate for that drive type.

7.8.4   Ready Idle Test

7.8.4.1   Overview

The ready idle test collect data for the ready idle metric.

The ready idle test shall begin immediately following the active test.

7.8.4.2   Procedure

No foreground IO shall be initiated on the product under test during the ready idle test other than that required to satisfy the instrumentation requirements in 7.2.9.

The test sponsor shall select a duration of at least two hours for the ready idle test. The test sponsor shall use the final two hours of the test as the measurement interval for the purposes of calculating the average power.

7.8.5   Capacity Optimization Test

This document does not define a capacity optimization method test for the Virtual Media Library taxonomy category.

8 Metrics

8.1 Taxonomy Considerations

This document defines metrics for the Online, Near-Online, Removable Media Library, and Virtual Media Library taxonomy categories.

8.2 Block Access Primary Metrics

This document defines the following primary metrics:

- Power efficiency for each test phase for Online and Near-Online systems (see subclause 8.4):
  - EP_{HB} for Hot band (IO/s/W);
  - EP_{RR} for Random Read (IO/s/W);
  - EP_{RW} for Random Write (IO/s/W);
  - EP_{SR} for Sequential Read (MiB/s/W);
  - EP_{SW} for Sequential Write (MiB/s/W);
  - EP_{RI} for Ready Idle (GB/W).
- Power efficiency for each test phase for Removable Media Library systems (see subclause 8.5):
  - EP_{SW} for Sequential Write (MiB/s/W);
  - EP_{SR} for Sequential Read (MiB/s/W);
  - EP_{RI} for Ready Idle (GB/W).
- Power efficiency for each test phase for Virtual Media Library systems (see subclause 8.6):
  - EP_{SW} for Sequential Write (MiB/s/W);
  - EP_{SR} for Sequential Read (MiB/s/W);
  - EP_{RI} for Ready Idle (GB/W).

8.3 File Access Primary Metrics

This document defines the following primary metrics:

- Power efficiency for each test phase for Online systems (see subclause 8.4) is:
  - EP_{VDA} for Video Data Acquisition (MiB/s/W);
  - EP_{DB} for Database (MiB/s/W);
  - EP_{VDI} for Virtual Desktop Integration (MiB/s/W);
  - EP_{SWB} for Software Build (MiB/s/W);
  - EP_{RI} for Ready Idle (GB/W).

8.4 Power Efficiency Metric for Online and Near-Online Systems

8.4.1 Ready Idle Test

For the ready idle test, the power efficiency metric represents the amount of raw capacity supported per watt of power required by the product under test. It is calculated as shown in Equation 8-1, as the ratio of:

- The raw capacity of the product under test, measured in GB;

- The average power, from the ready idle test, measured in watts.

Equation 8-1: Power Efficiency, Ready Idle

$$EP_{RI} = \frac{C_R}{PA_{RI}(7\ 200)}$$

Where:
- EP_{RI} is the power efficiency metric for the ready idle test;
- C_R is the raw capacity of the product under test (see subclause 4.2.31);
- $PA_{RI}(7\ 200)$ is the average power over the 2 h measurement interval for the ready idle test.

8.4.2   Block Access Active Test

For each test phase of the active test, the power efficiency metric represents the rate of data transfer supported per watt of power required by the product under test during a selected stable measurement interval. It is calculated, as shown in Equation 8-2, as the ratio of:

- The operations or throughput rate, during the measurement interval of the active test, expressed in IO/s or MiB/s;

- The average power, during the measurement interval of the active test, measured in watts.

Equation 8-2: Power Efficiency, Active (Block Access)

$$EP_i = \frac{O_i(1\ 800)}{PA_i(1\ 800)}$$

Where:
- EP_i is the power efficiency metric for Active Test test phase i;
- $PA_i(1\ 800)$ is the average power over the 30 min measurement interval for Active Test test phase i;
- $O_i(1\ 800)$ is the operations rate over the 30 min measurement interval for Active Test test phase i.

8.4.3   File Access Active Test

For each test phase of the active test, the power efficiency metric represents the rate of data transfer supported per watt of power required by the product under test during a selected stable measurement interval. It is calculated, as shown in Equation 8-3, as the ratio of:

- The throughput rate, during the measurement interval of the active test, measured in MiB/s;

- The average power, during the measurement interval of the active test, measured in watts.

Equation 8-3: Power Efficiency, Active (File Access)

$$EP_i = \frac{O_i(300)}{PA_i(300)}$$

Where:
- EP_i is the power efficiency metric for Active Test test phase i;
- $PA_i(300)$ is the average power over the 5 min measurement interval for Active Test test phase i;
- $O_i(300)$ is the operations rate over the 5 min measurement interval for Active Test test phase i.

8.4.4 Reporting

The power efficiency metric shall be reported to three significant digits.

8.5 Power Efficiency Metric for Removable Media Library Systems

8.5.1 Ready Idle Test

For the ready idle test, the power efficiency metric represents the amount of raw capacity supported per watt of power required by the product under test. It is calculated as shown in Equation 8-1, as the ratio of:

- The raw capacity of the product under test, measured in GB;

- The average power, from the ready idle test, measured in watts.

Note: Since tape cartridges themselves do not impact energy consumption, not all tape cartridges that can be present in a given configuration need to be present at the time of the test.

8.5.2 Active Test

For each test phase of the active test, the power efficiency metric represents the rate of data transfer supported per watt of power required by the product under test during a selected stable measurement interval. It is calculated, as shown in Equation 8-2, as the ratio of:

- The operations rate, during the measurement interval of the active test, measured in MiB/s;

- The average power during the measurement interval of the active test, measured in watts.

8.5.3 Reporting

The power efficiency metric shall be reported to three significant digits.

8.6 Storage Power Efficiency Metric for Virtual Media Library Systems

8.6.1 Ready Idle Test

For the ready idle test, the power efficiency metric represents the amount of raw capacity supported per watt of power consumed by the product under test. It is calculated as shown in Equation 8-1, as the ratio of:

- The raw capacity of the product under test, measured in GB;

- The average power, from the ready idle test, measured in watts.

8.6.2 Active Test

For each test phase of the active test, the power efficiency metric represents the rate of data transfer supported per watt of power required by the product under test during a selected stable measurement interval. It is calculated, as shown in Equation 8-2, as the ratio of:

- The operations rate, during the measurement interval of the active test, reported in MiB/s;

- The average power during the measurement interval of the active test, measured in watts.

8.6.3 Reporting

The power efficiency metrics, when reported, shall be reported to three significant digits.

8.7 Secondary Metrics

8.7.1 Capacity Optimization Metrics

This document defines capacity optimization tests for the Online and Near-Online taxonomy categories. The tests result in five binary, secondary metrics, which are given a value of 1 (Yes) if the product under test satisfies the named COM heuristic, and 0 (No) if it does not. The secondary metrics are:

- COM_{RD} based on the delta snapshots heuristic;

- COM_{WD} based on the delta snapshots heuristic;

- COM_{TP} based on the provisioning heuristic;

- COM_{DD} based on the data deduplication heuristic;

- COM_C based on the compression heuristic.

This document defines no relationship between the binary, secondary metrics defined in this subclause and the quantitative, primary metrics defined in subclauses 8.2 and 8.3.

9 Disclosure Requirements

9.1 General

This clause lists the information that shall be disclosed in every published test result identified as an SNIA Emerald™ Power Efficiency Measurement test result. Test results identified as an SNIA Emerald™ Power Efficiency Measurement test result shall be generated according to this document.

When units are specified after an item, that that shall be reported in terms of the specified unit.

SNIA Emerald™ Power Efficiency Measurement test results should be reported using the *Test Data Report Template*[17] provided for this purpose.

9.2 Product Identification

The information identified in this subclause shall be disclosed:

1. Product name;
2. Product model identification.

9.3 Test Metrics

9.3.1 Online and Near-Online, Block Access

For a test of a product in the taxonomy category Online or Near-Online using the block access test method, the information identified in this subclause shall be disclosed:

9.3.1.1 Primary Metrics

9.3.1.1.1 Ready Idle Test

1. Average power consumption (W);
2. Raw capacity of product under test (GB);
3. EP_{RI} for Ready Idle (GB/W).

9.3.1.1.2 Active Tests

9.3.1.1.2.1 Hot Band

1. EP_{HB} (IO/s/W).

9.3.1.1.2.2 Random Read

1. EP_{RR} (IO/s/W).

9.3.1.1.2.3 Random Write

1. EP_{RW} (IO/s/W).

9.3.1.1.2.4 Sequential Read

1. EP_{SR} (MiB/s/W).

9.3.1.1.2.5 Sequential Write

1. EP_{SW} (MiB/s/W).

9.3.1.2 Secondary Metrics (COMs)

9.3.1.2.1 Data Deduplication

1. COM_{DD} - Data deduplication heuristic indicated data deduplication present (Yes or No);

2. Data deduplication present and enabled in product under test during active tests (Yes or No).

9.3.1.2.2 Compression

1. COM_C - Compression heuristic indicated compression present (Yes or No);

2. Compression present and enabled in product under test during active tests (Yes or No).

9.3.1.2.3 Thin Provisioning

1. COM_{TP} - Thin provisioning heuristic indicated thin provisioning present (Yes or No);

2. Thin provisioning present and enabled in product under test during active tests (Yes or No).

9.3.1.2.4 Read-Only Delta Snapshots

1. COM_{RD} - Read-only delta snapshot heuristic indicated read-only snapshot present (Yes or No);

2. Read-only delta snapshort present and enabled in product under test during active tests (Yes or No).

9.3.1.2.5 Writable Delta Snapshots

1. COM_{WD} - Writable delta snapshot heuristic indicated writable delta snapshot present (Yes or No);

2. Writable delta snapshot present and enabled in product under test during active tests (Yes or No).

9.3.2 Online and Near-Online, File Access

For a test of a product in the taxonomy category Online or Near-Online using the file access test method, the information identified in this subclause shall be disclosed:

9.3.2.1 Primary Metrics

9.3.2.1.1 Ready Idle Test

1. Average power consumption (W);

2. Raw capacity of product under test (GB);

3. EP_{RI} for Ready Idle (GB/W).

9.3.2.1.2 Active Tests

9.3.2.1.2.1 Video Data Acquisition

1. EP_{VDA} (MiB/s/W).

9.3.2.1.2.2 Database

1. EP_{DB} (MiB/s/W).

9.3.2.1.2.3 Virtual Desktop Integration

1. EP_{VDI} (MiB/s/W).

9.3.2.1.2.4 Software Build

1. EP_{SWB} (MiB/s/W).

9.3.2.2   Secondary Metrics (COMs)

9.3.2.2.1   Data deduplication

1. COM_{DD} - Data deduplication heuristic indicated deduplication present (Yes or No);

2. Data deduplication present and enabled in product under test during active tests (Yes or No).

9.3.2.2.2   Compression

1. COM_C - Compression heuristic indicated compression present (Yes or No);

2. Compression present and enabled in product under test during active tests (Yes or No).

9.3.2.2.3   Thin Provisioning

1. COM_{TP} - Thin provisioning heuristic indicated thin provisioning present (Yes or No);

2. Thin provisioning present and enabled in product under test during active tests (Yes or No).

9.3.2.2.4   Read-Only Delta Snapshots

1. COM_{RD} - Read-only delta snapshot heuristic indicated read-only snapshot present (Yes or No);

2. Read-only delta snapshort present and enabled in product under test during active tests (Yes or No).

9.3.2.2.5   Writable Delta Snapshots

1. COM_{WD} - Writable delta snapshot heuristic indicated writable delta snapshot present (Yes or No);

2. Writable delta snapshot present and enabled in product under test during active tests (Yes or No).

9.3.3   Removable Media Library and Virtual Media Library

For a test of a product in the taxonomy category Removable Media Library or Virtual Media Library (using the block access test method), the information identified in this subclause shall be disclosed:

9.3.3.1   Primary Metrics

9.3.3.1.1   Ready Idle Test

1. Average power consumption (W);

2. Raw capacity of product under test (GB);

3. EP_{RI} for Ready Idle (GB/W).

9.3.3.1.2   Active Tests

9.3.3.1.2.1   Sequential Read

1. EP_{SR} (MiB/s/W).

9.3.3.1.2.2   Sequential Write

1. EP_{SW} (MiB/s/W).

9.4   Test Characterization

The information characterizing the test environment identified in this subclause shall be disclosed:

1. Version of this document according to which testing was performed;

2. Taxonomy category of the product under test;

3. Taxonomy classification of the product under test;

4. Access method tested (block access or file access);

5. Workload generator name;

6. Workload generator version identifier;

7. Workload script name;

8. Workload script version identifier;

9. For file access tests, identity of the sFlow® agent;

10. For file access tests, sFlow® agent version identifier;

11. For file access tests, identity of the sFlow® collector used;

12. For file access tests, sFlow® collector version identifier;

13. Transport protocol(s) used during testing to access data on product under test.

9.5 Product Under Test Description

9.5.1   Product Under Test Description – General

The information describing the product under test identified in this subclause shall be disclosed:

1. Power input – nominal voltage (V);

2. Power input – nominal frequency (Hz);

3. Power input – number of phases;

4. Average air inlet temperature during test (°C);

5. Is storage protection present and enabled? (Yes or No);

6. Is product under test configured to be no SPOF (Yes or No)

7. Is auto-tiering present and enabled? (Yes or No).

9.5.2   Product Under Test Description – Storage Controllers

The information describing each storage controller component of the product under test identified in this subclause shall be disclosed:

1. Number of controllers of this type;

2. Controller part nmber;

3. Controller description;

4. Controller main memory size (GiB);

5. Controller NVRAM size (GiB);

6. Is storage protection present and enabled? (Yes or No).

9.5.3   Product Under Test Description – Power Supplies

The information describing each power supply component of the product under test identified in this subclause shall be disclosed:

1. Number of power supplies of this type;

2. Power supply part nmber;

3. Power supply 80 PLUS[16] rating.

9.5.4   Product Under Test Description – Media Shelves / Drawers

The information describing each media shelf / drawer component of the product under test identified in this subclause shall be disclosed:

1.   Number of media shelves / drawers of this type;

2.   Media shelf / drawer part nmber.

9.5.5   Product Under Test Description – Storage Media

The information describing each storage media component (storage device) of the product under test identified in this subclause shall be disclosed:

1.   Number of storage media of this type;

2.   Storage media vendor name;

3.   Media part number;

4.   Media type;

5.   Media form factor;

6.   Raw, unformatted, uncompressed capacity of the storage media (GB);

7.   Media rotational speed;

8.   Media cache size (MiB);

9.   Media interface.

9.5.6   Product Under Test Description – Removable Storage Media

The information describing each removable storage media component (storage device) of the product under test identified in this subclause shall be disclosed:

1.   Number of removable storage media drive slots in product under test;

2.   Number of drives for removable storage media;

3.   Number of robotics units for removable storage media;

4.   Number of removable media cartridge slots in product under test;

5.   Number of removable media cartridges present in product under test;

6.   Removable media cartridge type(s).

9.5.7   Product Under Test Description– Host Protocol Controllers

The information describing each host protocol controller component of the product under test identified in this subclause shall be disclosed:

1.   Number of host protocol controllers of this type;

2.   Host protocol controller part number;

3.   Host protocol(s) supported;

4.   Host protocol controller nominal data rate (GiB/s).

Annex A   (Normative) Suggested Power and Environmental Meters

A.1 Overview

The power and temperature meters that should be used are those approved for use with the SPEC Power® benchmark[18] and appearing in the *SPEC PTDaemon Accepted Measurement Devices*[19] on the *List of Accepted Power Analyzers* and *List of Accepted Temperature Sensors*.

For more information on the appropriate use of a power meter, see the *User Guide for the SNIA Emerald™ Power Efficiency Measurement Specification V3*[13]

A.2 Alternate Meter Usage

A test sponsor may perform a SNIA Emerald™ Power Efficiency Measurement using a meter not on the list ("alternate meter"). If a measurement uses an alternate meter, sufficient documentation about the meter and its configuration and calibration to ensure that its use provides measurement data equivalent to that provided by the recommended meters shall be disclosed.

Test sponsors using an alternate meter shall attest to its equivalence to the meters on the recommended list at the time of the test in the following areas:

* Accuracy;

* Resolution;

* Calibration.

Annex B (Normative) Measurement Requirements

B.1 Online and Near-Online Block Access Data Collection and Processing Requirements

A summary of the data collection and processing requirements for Online and Near-Online block access testing is provided in Table B-1.

Table B-1   Online and Near-Online Block Access Summary

Test	Power and Temperature		Workload Generator Data Collection		Minimum Test Duration (minutes)
	Power PA_i (·) (seconds)	Temperature Recording Interval (seconds)	Metric	Collection interval (seconds)	
Conditioning	60	10	Average Response Time RTA_{sc} (ms)	60	720
Active	60	10	1) Operations Rate O_i (IO/s or MiB/s) 2) Average Response Time RTA_i (ms)	60	40
Ready Idle	60	10	N/A	N/A	120

B.2 Removable and Virtual Media Libary Block Access Data Collection and Processing Requirements

A summary of the data collection and processing requirements for Removable Media Library and Virtual Media Library testing is provided in Table B-2.

Table B-2   Removable and Virtual Media Library Summary

Test	Power and Temperature		Workload Generator Data Collection		Minimum Test Duration (minutes)
	Power PA_i (·) (seconds)	Temperature Recording Interval (seconds)	Metric	Collection interval (seconds)	
Conditioning	60	10	1) Average throughput for each drive (MiB/s) 2) Operations Rate O_{sc} (MiB/s)	60	14
Active	60	10	1) Average throughput for each drive (MiB/s) 2) Operations Rate O_i (MiB/s)	60	30
Ready Idle	60	10	N/A	N/A	120

B.3 Online and Near-Online File Access Data Collection and Processing Requirements

A summary of the data collection and processing requirements for Online and Near-Online file access testing is provided in Table B-3.

Table B-3 Online and Near-Online File Access Summary

Test	Power and Temperature		sFlow® Data Collection		Minimum Test Duration (minutes)
	Power PA_i (·) (seconds)	Temperature Recording Interval (seconds)	Metric	Collection interval (seconds)	
INIT	10	10	Operations Rate O_i (MiB/s)	10	N/A
Warm-up – per load point	10	10	Operations Rate O_i (MiB/s)	10	5
Active – per load point	10	10	Operations Rate O_i (MiB/s)	10	5
Ready Idle	10	10	N/A	N/A	120

Annex C   (Normative) Stable Storage

Stable storage is storage that retains data, for a minimum of 72 h without external power, in the case of:

1. Repeated power failures, including cascading power failures;

2. Hardware failures (of any board, power supply, etc.);

3. Repeated system crashes, including reboot cycle.

This definition does not address failure of the storage itself. For example, failures of disks or nonvolatile RAM modules are not addressed in the definition of stable storage.

We further clarify the definition of the term "stable storage" to resolve any potential ambiguity. This clarification is necessary since the definition of stable storage has been, and continues to be, a point of contention. Therefore, for the purposes of this document, stable storage is defined in terms of the following operational description:

To offer stable storage, the product under test shall be able to tolerate without loss of committed data:

1. Power failures of the product under test's primary power source, including cascading power failures, with a total duration of no longer than 72 h;

2. Non-destructive transient failures (temporary failure which does not require replacement or upgrade of the failed hardware or software component) of any hardware or software component in the product under test which result in a system crash. Multiple and/or cascading failures are excluded;

3. Manual reset of the entire product under test, or of any of its components involved in providing services, if required to recover from transient failures.

If the product under test allows data to be cached in intermediate storage, after a response to the client indicating that the data has been committed, but before the data is flushed to permanent storage, then there shall be a mechanism to ensure that the cached data survives failures of the types defined above.

Committed data may or may not be preserved in the face of unbounded numbers of cascading hardware or software errors that combine to prevent the system from performing any significantly useful work.

Annex D   (Normative) Vdbench

Vdbench is an IO performance test tool developed by Oracle Corporation.

The Vdbench software specified on the *Download Material for SNIA Emerald™ Testing*[12] web page for use with this document shall be used for block access testing. The web page also provides access to the Vdbench software and related materials.

Annex E   (Normative) Vdbench Test Scripts

The Vdbench test tool is directed by a test script that parameterizes a run of Vdbench.

Two Vdbench Test Scripts are provided for block access testing. Scripts are provided for:

- Block access testing of Online and Near-Online category devices;
- Block access testing of Removable Media Library and Virtual Media Library category devices.

The Vdbench Test Scripts provided on the *Download Material for SNIA Emerald™ Testing*[12] web page for use with this document shall be used for block access testing.

Annex F   (Normative) SPEC SFS® 2014 Benchmark

SPEC SFS® 2014 is a performance benchmark developed by the Standard Performance Evaluation Corporation (SPEC). This document describes the use of SPEC SFS® 2014 Benchmark as the workload generator for file access testing.

The SPEC SFS® 2014 software specified on the *Download Material for SNIA Emerald™ Testing*[12] web page for use with this document shall be used for file access testing. The web page also provides access to the SPEC SFS® 2014 software and related materials.

SPEC SFS® 2014 benchmarking software is orderable for a fee from the SPEC website *Purchase Current SPEC Benchmark Suites*[6].

Annex G   (Normative) File Access IO Load Driver Configuration File

The SPEC SFS® 2014 benchmark tool is directed by configuration file that parameterizes a run of the SPEC SFS® 2014 benchmark.

The File Access IO Load Driver Configuration File provided on the *Download Material for SNIA Emerald™ Testing*[12] web page for use with this document shall be used for file access testing.

Annex H (Informative) sFlow®

sFlow®[7][9] is an industry standard for sampling packet traffic and interface counter statistics of devices participating in high speed networks. The sFlow® protocol is maintained by sFlow.org[8].

An sFlow® Agent is software that runs within a network device, combining data into sFlow® datagrams that are sent across the network to an sFlow® Collector. The sFlow® Agent is typically implemented by the network equipment vendor as a software feature.

An sFlow® Collector is software that accepts sFlow® datagrams and performs processing and analysis on the received data, producing reports for the user. There are multiple sFlow® Collectors available commercially.

A list of network equipment that supports sFlow® and a list of available sFlow® Collectors are available from sFlow.org[8].

Annex I   (Normative) COM Test Data Set Generator

The COM Test Data Set Generator[15] is open-source software that creates data sets used in performing SNIA Emerald™ Power Efficiency Measurements.

The COM Test Data Set Generator software specified on the *Download Material for SNIA Emerald™Testing*[12] web page for use with this document shall be used COM testing.

Annex J   (Informative) gzip

Gzip[14] is an open-source file compression utility.

Bibliography

[1] Storage Networking Industry Association (SNIA) (2008). *Understanding data deduplication ratios.* Available from https://www.snia.org/sites/default/files/Understanding_Data_Deduplication_Ratios-20080718.pdf.

[2] Storage Networking Industry Association (SNIA). *The SNIA Dictionary*. Available from https://www.snia.org/dictionary.

[3] Standard Performance Evaluation Corporation (SPEC). *SPEC SFS® 2014 SP2 User's Guide.* Available from https://www.spec.org/sfs2014/index.html#userguidee.

[4] Standard Performance Evaluation Corporation (SPEC). *Standard Performance Evaluation Corporation*. Available from https://www.spec.org/.

[5] Standard Performance Evaluation Corporation (SPEC). *SPEC's Benchmarks.* Available from https://www.spec.org/benchmarks.html.

[6] Standard Performance Evaluation Corporation (SPEC). *Purchase Current SPEC Benchmark Suites.* Available from https://www.spec.org/order.html.

[7] Internet Engineering Task Force (IETF). RFC 3176. *InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks.* Available from https://www.ietf.org/rfc/rfc3176.txt.

[8] sFlow.org. *sFlow®*. Available from http://www.sflow.org/.

[9] Allied Telesis. *sFlow® Feature and Configuration Guide.* Available from https://www.alliedtelesis.com/sites/default/files/documents/feature-guides/sflow_feature_overview_config_guide_revb.pdf.

[10] Internet Engineering Task Force (IETF). RFC 1813. *NFS Version 3 Protocol Specification*. Available from https://datatracker.ietf.org/doc/rfc1813/.

[11] Microsoft (2017*). MS-SMB2. Server Message Block (SMB) Protocol Versions 2 and 3*. Available from https://msdn.microsoft.com/en-us/library/cc246482.aspx.

[12] Storage Networking Industry Association (SNIA). *Download Material for SNIA Emerald™Testing*. Available from https://www.snia.org/emerald/download.

[13] Storage Networking Industry Association (SNIA) (2018). *User Guide for the SNIA Emerald™ Power Efficiency Measurement Specification V3.* Available from https://www.snia.org/emerald/download.

[14] *The gzip home page*. Available from https://www.gzip.org/.

[15] Storage Networking Industry Association (SNIA). *COM Test Data Set Generator software*. Available from https://www.snia.org/emerald/download.

[16] CLEAResult Plug Load Solutions. *80 PLUS Certified Power Supplies and Manufacturers*. Available from https://www.80plus.org/.

[17] Storage Networking Industry Association (SNIA). *Test Data Report Template.* Available from https://www.snia.org/emerald/download.

[18] Standard Performance Evaluation Corporation (SPEC). *SPECpower_ssj® 2008.* Available from https://www.spec.org/power_ssj2008/.

[19] Standard Performance Evaluation Corporation (SPEC). *SPEC PTDaemon Accepted Measurement Devices.* Available from https://www.spec.org/power/docs/SPECpower-Device_List.html.

[20] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). *Thermal Guidelines for Data Processing Environments.*

SNIA Emerald™ Power Efficiency Measurement Specification V3.0.3

SNIA Technical Position

[21] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) (2011). *2011 Thermal Guidelines for Data Processing Environments – Expanded Data Center Classes and Usage Guidance*. Available from https://ecoinfo.cnrs.fr/IMG/pdf/ashrae_2011_thermal_guidelines_data_center.pdf.