# GREEN STORAGE PRODUCTS: Efficiency with ENERGY STAR & Beyond

Patrick Stanko
Oracle Tape Development, Oracle
Chair of the SNIA Green TWG

# SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the Author nor the Presenter is an attorney and nothing in this presentation is intended to be nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.
- The information presented herein represents the Author's personal opinion and current understanding of the issues involved. The Author, the Presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

# Abstract


- **GREEN STORAGE PRODUCTS: Efficiency with ENERGY STAR & Beyond**
  - This workshop will cover storage-specific topics related to energy-efficiency and outline the current state of the industry. We will discuss the components of modern storage systems that impact energy consumption and a range of technologies that are currently considered “green storage”.  We will discuss the SNIA Emerald program and its corresponding storage power efficiency measurement specification.  This discussion will cover the metrics used for measuring, maintaining and designing  for power in storage systems.  We will also outline ongoing efforts by other organizations to standardize metrics for measuring storage systems.

# Agenda

- Overview, definitions, and what does green storage mean
- SNIA green storage activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald™ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald™ Program
  - SNIA Emerald™ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations' green storage efforts
- Storage technologies for energy savings
- Typical savings

# Acknowledgement

- Initial tutorial
  - Erik Riedel (EMC) and Carlos Pratt (IBM)
- Green Storage –The Big Picture
  - SW Worth, Microsoft
- Best foot forward
  - Jim Espy (EMC)
- Best foot forward example
  - Herb Tanzer (HP)
- Storage technologies for energy savings tutorial
  - Alan Yoder (NetApp)
- Members of SNIA GSI and GreenTechnical Working Group
- Emerald™ Director
  - Dave Thiel

# Overview and Definitions

- **‘Green’** – What does it mean to the IT ‘ecosystem’?
  - Reduction of Total Impact on Environment
    - Systems approach - More than just Energy Use!
  - Defined by Gov’t. (EPA, EU, Kyoto), Orgs, Vendors, etc.
- How does “Green” differ from normal economic considerations, e.g. TCO, efficiency, optimization?
  - Systems viewpoint + Lifecycle analysis
  - Widen scope of action across system/org boundaries, time
  - Rationalize decisions by including **“externalities”**

**Check out the SNIA Tutorials Green Storage – The Big Picture www.snia.org/education/tutorials**

# "Green" effects on I.T.

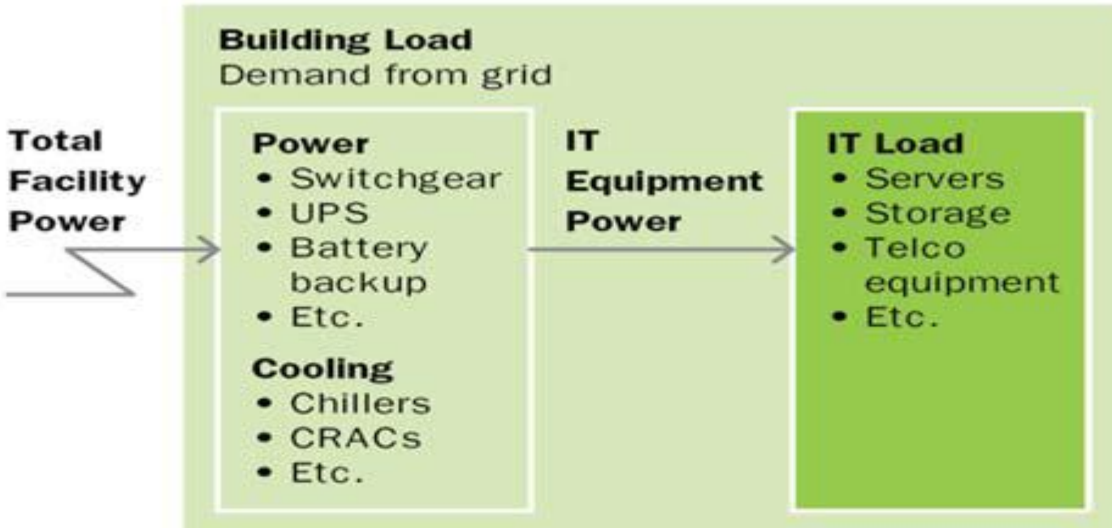
- "TCO" (Total Cost of Ownership) should include Externalities in accounting and purchase decisions
  - In most cases Externalities will evolve to provide clear pricing signals (e.g. RoHS, WEEE, Cap-and-Trade)
- Systems viewpoint *(bigger picture)* is essential!
- Expand scope of decision-criteria and constraints to include (at least) entire datacenter *(entire supply chain?)*
  - Servers, Networking, and **Storage**
  - Power, energy, and cooling (CapEx and OpEx)
  - People: widen their decision-boundaries, -constraints
    - Include your Facilities managers!

# Measuring 'Green' (Energy only)

Education SNIA

## Green Grid metrics www.thegreengrid.org

**PUE: Power Usage Effectiveness**
DCiE: Data Center Infrastructure Efficiency




$$\text{PUE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

$$\text{DCiE} = \frac{1}{\text{PUE}} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}} \longleftarrow$$

*(Multiply both terms by 100%)*

# How we talk about ‘Green’ Storage…

- How much data center Energy Usage is due to Storage?
  - ***It depends…***on Design and Workload (I/O profiles)!
  - Published studies range from <10% - >40%
  - “Rule-of-Thumb” for energy: 60% servers, 20% networking, and 20% **Storage** (but no consistent definition of *‘Storage’*)
- *➔Proportion of Energy used by Storage is increasing, because of…*
  - Facilities improvements (PUE, DCiE)
  - Virtualization – especially of Servers, O.S., Applications

# The Players in Green IT

- I.T. owners / **Data Center operators** (“Customers”)
- **Vendors** of I.T. hardware, software, systems, services
  - Engineers/Developers/Architects – *including Cloud vendors!*
- Energy **Utilities and Regulators**
- **Governments**: local, regional, national, supra-national
  - US-EPA Energy*Star programs
  - Euro. Comm. Code of Conduct on Data Centre Energy Effic.
- **Green Grid** metrics www.thegreengrid.org
  - Focus on Power, Energy, and Cooling used for IT
- **➔SNIA** – org expertise on enterprise STORAGE
- Other interested parties (e.g. Uptime Institute, ASHREA)
- **ISO/IEC/INCITS** Energy Efficient Data Center

# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA green activities
  - Green Storage Initiative
  - Green  Storage Technical Working Group
  - SNIA Emerald™ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald™ Program
  - SNIA Emerald™ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

# SNIA Green Activities GSI



- SNIA Green Storage Initiative (GSI)
  - To conduct research on power and cooling issues confronting storage administrators
  - Educate the vendor and user community about the importance of power efficiency in shared storage environments
  - Leverage SNW and other SNIA and partner conference to focus attention on energy efficiency for networked storage infrastructures
  - Provide input to the SNIA Green Storage TWG on requirements for green storage metrics and standards
  - Provide external advocacy and support of SNIA Green TWG technical work

SNIA Green Storage Initiative
SNIA
towards energy-efficient storage networking solutions

# SNIA Green Activities
# Green Storage TWG

- SNIA Green Storage Technical Working Group (TWG)
  - Technical body working on green storage metrics and standards
  - Gets direction from GSI
  - Wrote the SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification and related documents
  - Supports the SNAI Emerald$^{TM}$ Program
    - White papers
    - Tutorials
    - Training
  - Works with regulatory agencies; i.e. EPA, on green storage specifications

# SNIA Emerald™ Program Overview

- Purpose
  - Provide open access to storage system power efficiency information using a well-defined testing procedure and additional information related to system power characteristics
  - The report data can help IT professionals make storage platform selections as part of an overall Green IT and Sustainability objective
  - Easily identifiable program logo
- Test procedure: SNIA Emerald™ Power Efficiency Measurement Specification
- Public access and submittal is through the sniaemerald.com web site
  - No charge for access to test results, specifications or user guides
  - Submission of results is for a modest fee, discounted or waived for SNIA/GSI members
  - SNIA membership is not required to submit or to access test results
  - Voluntary, non-exclusionary, low cost program for manufacturers - Options for self-measure or third party measurement

# SNIA Emerald™ Program Overview Continued

- **Process**
  - Storage Vendors test their equipment and submit test results to the Emerald Program
  - Emerald Program publishes results on the sniaemerald.com web site
  - IT users (public) download results from the sniaemerald.com web site
  - Vendor gains right to use the SNIA Emerald™ logo in conjunction with tested products
- **Legal protections**
  - Terms of Use: conditions on use of test results agreed to by those downloading results
  - Terms of Submission: agreed to by vendor submitting test results
- **Sign up for the mailing list: sniaemerald.com**

# Why Should Storage Vendors Use the Emerald™ Program?

- SNIA Emerald™ Program seeks to
  - Encourage storage vendors to build better products
  - Stimulate the IT community to more rapidly deploy and operate multi-vendor storage technology efficiently
- SNIA Emerald™ Program
  - Provides a level playing field for test sponsors
  - Produces results that are powerful and yet simple to use
  - Provides value for vendors as well as IT consumers and solution integrators
  - Reports results in a manner that is easy to submit, audit and verify

# Why Should Storage Consumers Use the Emerald™ Program?

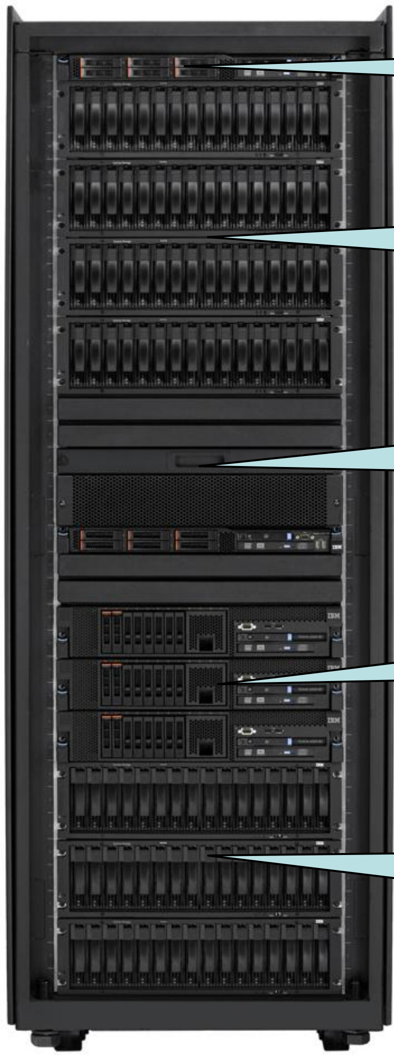
- SNIA Emerald™ Program seeks to
  - Provide a collection of standard metrics and data that allows IT architects to objectively compare a range of possible storage solutions
- SNIA Emerald™ Program
  - Enables users to select the mode of storage usage that accomplishes their work objectives with the lowest overall energy consumption
  - Drives vendor companies to innovate and compete in the development of energy efficient products as measured by the standard yardsticks

# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA green storage activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald$^{TM}$ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald$^{TM}$ Program
  - SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations' Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

# Background

## Basic Anatomy of a Disk Storage System and its Use



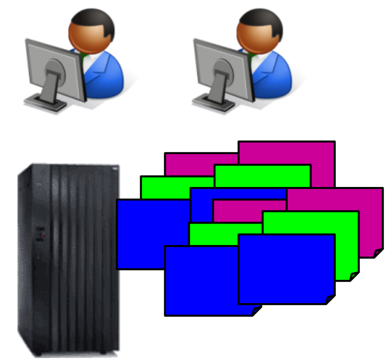
Switches
(sometimes optional)

Disks
(Spin or Solid State)
Diverse technologies
(SAS, FC, SATA)

Power supplies,
Batteries & fans
(may or may not be
redundant)

Controller's
(may or may not be
redundant)

Disks
(Spin or Solid State)
Diverse technologies
(SAS, FC, SATA)



Users and Apps

Other:

Software (firmware & microcode)

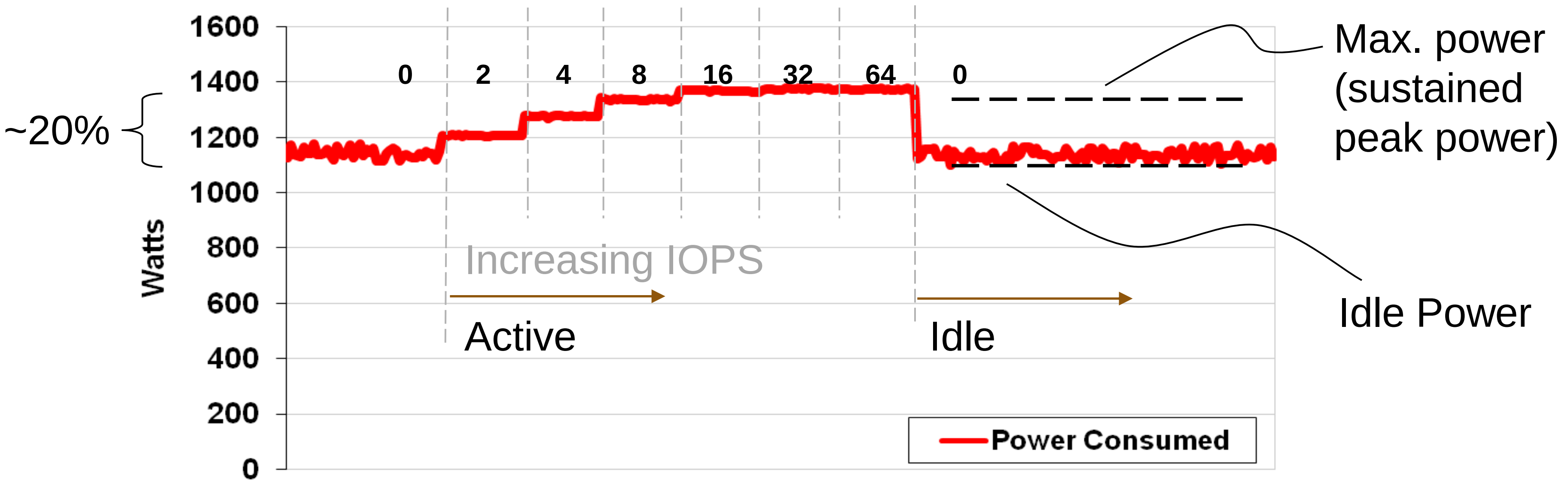
Hardware Design

Environment

# What Impacts Power Consumption

- Storage capacity / usage efficiency
  - increasing data → larger capacity → more disks
  - redundant copies → magnify capacity needs
  - variability in usage and utilization → inefficient allocation of space
  - What is valuable data? What is the retention policy?
- Data transfer rate / access speed
  - high I/O bandwidth → higher rotational speed; striping across many drives
  - low access times → faster actuators; higher rotational speeds; caches
  - How fast and immediate must data be available? (time-to-data)
- Data integrity
  - 25% of "digital universe" is unique, but 75% are replicas / duplicates
  - partly to ensure data integrity and survivability; partly wasteful
- Data availability / system reliability
  - RAID uses extra drives, plus redundant power supplies, fans, controllers,
  - How valuable is data? How likely are failures? How fast must data be available?

# Storage Power Example




- Ideally, systems consume minimum power in all modes
  - Example system consumes **significant power in ready idle** (**80% of max**)
- % of time in Idle versus Active depends on storage type, application and workloads; available optimizations will vary
  - Power itself is only one part of the story it must be reflected as a metric
- Power consumed is not linearly proportional to workload

# Opportunities to Make Storage Green(er)

- Environment
  - Higher system tolerance to high/lower temperatures and humidity
  - In line with cold and hot aisles designs on new data centers
- Improve usage efficiency
  - De-duplication and compression
  - Thin provisioning
- Minimize energy consumption
  - Improved component designs – high-efficiency power supplies, advanced & flexible storage devices
  - Variants of MAID – idle and spin-down
- New technologies
  - Solid state storage
  - Alternative + hybrid system designs (opportunity to rethink)

must be driven by metrics / standards / guidelines

# Green Storage Metrics

- SNIA recommended metrics
  - Capacity metric (ready-idle)
    - Relates the power of the system to its total storage raw capacity. It is reported as GB/watt (or TB/watt)
    - Power required to store and protect the data
  - Workload metric (Active)
    - Relates the power of the system to the maximum possible IOPS generated by a specific random stress load. It is reported as IOPS/watt
    - Power required to randomly supply data to and from a host
  - Bandwidth metric (Active)
    - Relates the power drown by the system to the maximum possible MBPS generated by a specific sequential stress load. It is reported as MBPS/watt
    - Power required to stream data to  and from a host

# What is the Ready Idle Metric and what it Means to an IT Manager?

Education SNIA

- Depending on the systems and their usage their energy usage may be evaluated according to:
  - Is the system idle time at least 12 hours or more a day?
    - You should be interested in the power required to store the data
    - **capacity metric (GB/Watt)** may be your best indicator on how energy efficient your system is
    - The larger this number is the less watts are used to energize the total storage of your system

# What are the Active Metrics and what it Means to an IT Manager?

Education SNIA

- Depending on the systems and their usage their energy usage may be evaluated according to:
  - For systems running more than 12 hours a day
    - You should be interested in the Power to move the data onto and off the storage system
    - Is your load predominantly sequential?
      - **Bandwidth metric (MBS/Watt)** will help you to determine how effective is your power use. The larger this number is, the more data the system is pushing per watt
    - Is your load predominantly random?
      - **Workload metric (IOPS/Watt)** will help you determine how effective is your power use. The larger this number is the system is provides more operations per watt.
    - Independently on how long the system is idle it is always good to know what is your capacity per watt ratio

# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA storage green activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald$^{TM}$ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald$^{TM}$ Program
  - SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations' Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

# Why Have a Storage Taxonomy

Education SNIA

- Need a taxonomy (product classification) to enable fair comparisons among similar storage products
  - e.g. for motor vehicles – motorcycles, cars, trucks
- Similar green metrics may apply to all product categories, but different values establish best-in-class
- Unique considerations apply to special categories
  - e.g. amphibious cars, skid steer loaders, tanks
- Clear taxonomy simplifies comparisons and aid regulatory efforts
- SNIA Storage Taxonomy is defined in SNIA Emerald™ Power Efficiency Measurement Specification

# Taxonomy – Categories

| Attribute | Category | | | | | |
|---|---|---|---|---|---|---|
| | **Online** | **Near Online** | **Removable Media Library** | **Virtual Media Library** | **Adjunct Product** | **Interconnect Element** |
| Access Pattern | Random/ Sequential | Random/ Sequential | Sequential | Sequential | | |
| MaxTTFD (t) | t < 80 ms | t > 80 ms | t > 80 ms<br>t < 5 min | t < 80 ms | t < 80 ms | t < 80 ms |
| User Accessible Data | Required | Required | Required | Required | Prohibited | Prohibited |

- Six categories, covering most storage industry products

# Taxonomy – Categories

| Category / Level | **Online** | **Near Online** | **Removable Media Library** | **Virtual Media Library** |
|---|---|---|---|---|
| Consumer/ Component | 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 |

| **Adjunct Product** | **Interconnect Element** |
|---|---|
| | |
| | |
| | |
| | |
| | |

- 23 total "buckets" covering the breadth of the industry

# Taxonomy – Online

- Most common storage systems

| 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 |
| User-Accessible Data | Required | Required | Required | Required | Required | Required |
| Consumer/ Component | Yes | No | No | No | No | No |
| Connectivity | Not specified | Connected to single or multiple hosts | Network-connected | Network-connected | Network-connected | Network-connected |
| Maximum Configuration | ≥1 | ≥ 4 | ≥ 12 | > 100 | >400 | >400 |
| 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 |
| Non-Disruptive Serviceability | Optional | Optional | Optional | Optional | Required | Required |
| FBA/CKD Support | Optional | Optional | Optional | Optional | Optional | Required |

# Taxonomy – Near Online

| 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 | Random/ Sequential |
| MaxTTFD (t) | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms |
| User-Accessible Data | Required | Required | Required | Required | Required | Required |
| Consumer/ Component | Yes | No | No | No | No | No |
| Connectivity | Not specified | Connected to single or multiple hosts | Network-connected | Network-connected | Network-connected | Network-connected |
| Maximum Configuration | ≥1 | ≥ 4 | ≥ 12 | > 100 | >400 | >400 |
| 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 |
| Non-Disruptive Serviceability | Optional | Optional | Optional | Optional | Required | Required |
| FBA/CKD Support | Optional | Optional | Optional | Optional | Optional | Required |

# Taxonomy – Removable Media Library

| Attribute | Classification | | | | | |
|---|---|---|---|---|---|---|
| | Removable 1 | Removable 2 | Removable 3 | Removable 4 | Removable 5 | Removable 6 |
| Access Pattern | Sequential | Sequential | Sequential | | Sequential | Sequential |
| MaxTTFD (t) | 80ms < t < 5m | 80ms < t < 5m | 80ms < t < 5m | | 80ms < t < 5m | 80ms < t < 5m |
| User-Accessible Data | Required | Required | Required | | Required | Required |
| Maximum Drive Count | Not specified | 4 | ≥ 5 | | ≥ 25 | ≥ 25 |
| Robotics | Prohibited | Required | Required | | Required | Required |
| No SPOF | Optional | Optional | Optional | | Optional | Required |
| Non-disruptive Serviceability | Optional | Optional | Optional | | Optional | Required |

# Taxonomy – Virtual Media Library

| Attribute | Classification | | | | | |
|---|---|---|---|---|---|---|
| | Virtual1 | 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 |
| User-accessible Data | Required | Required | Required | | Required | Required |
| Maximum Configuration | 12 | >12 | > 48 | | > 96 | > 96 |
| Storage Protection | Optional | Optional | Required | | Required | Required |
| No SPOF | Optional | Optional | Optional | | Optional | Required |
| Non-Disruptive Serviceability | Optional | Optional | Optional | | Optional | Required |

# Taxonomy – Adjunct & Interconnect

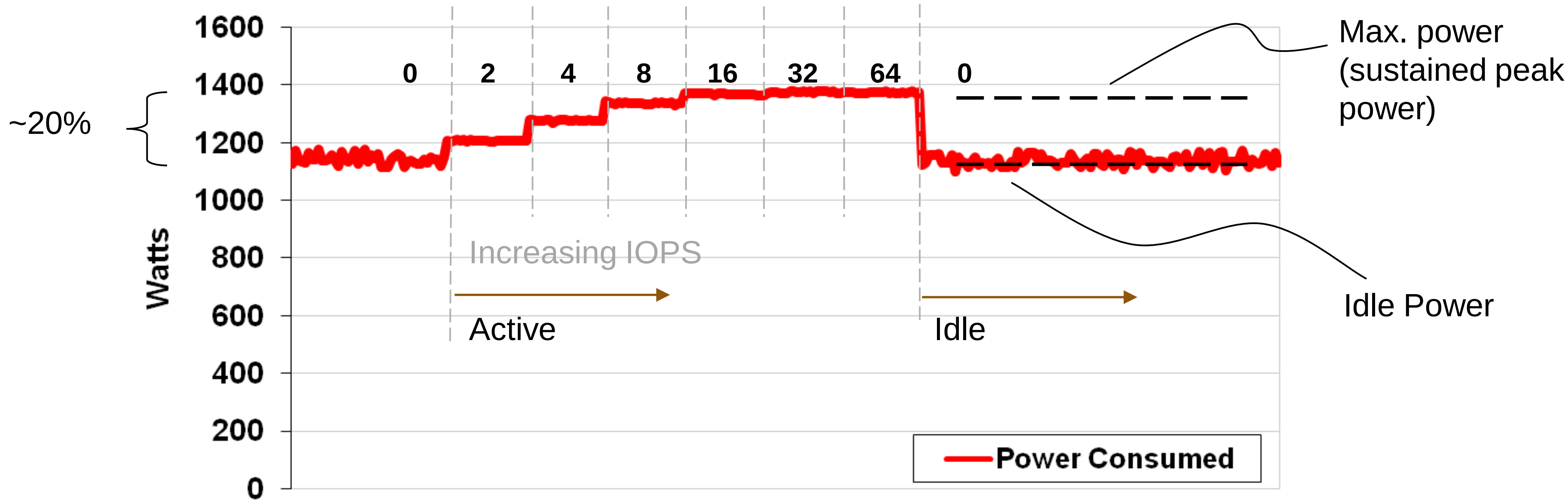
- Adjunct and Interconnect left to be defined

| Adjunct Product | Interconnect Element |
|---|---|
| | |
| | |
| | |
| | |
| | |

# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA storage green activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald$^{TM}$ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald$^{TM}$ Program
  - SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations' Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

# Storage Array Power Example

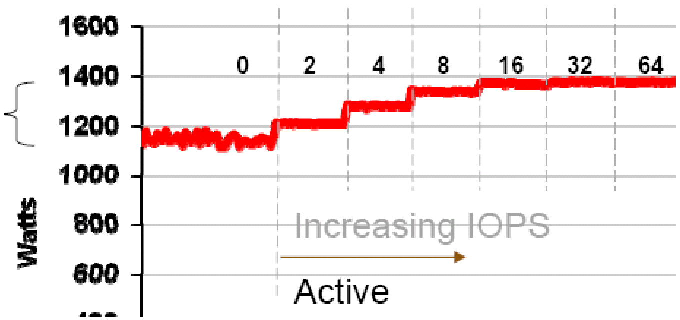



- Many variations in workloads
- Many variations in system configuration
- Need to pick metrics that show storage power efficiency

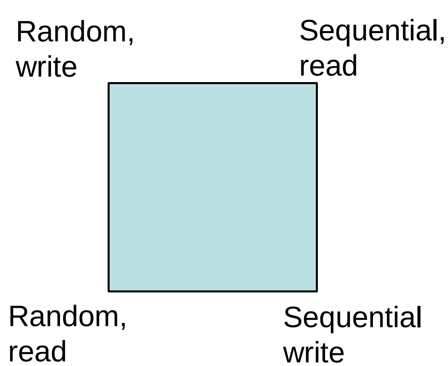
# Desired Metric – “Productivity”

Many possible definitions – must balance simplicity against applicability

- “typical workload”, with levels




- “four corners”, maximum performance, maximum power




- detailed performance benchmarks – results/W

Storage Performance Council

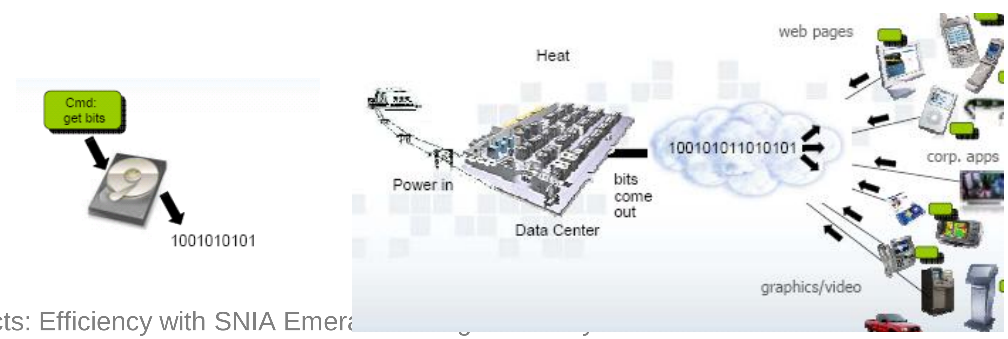
*Defining, administering, and promoting industry-standard, vendor-neutral benchmarks to characterize the performance of storage products*

Standard Performance Evaluation Corporation

spec

the green grid
get connected to efficient IT

- The Green Grid Productivity Proxy Proposals example – Proxy #4 – bits/kilowatt-hour



# Green Storage Metrics

- SNIA recommended metrics
  - Capacity metric (ready-idle)
    - Relates the power of the system to its total storage raw capacity. It is reported as GB/watt (or TB/watt)
    - Power required to store and protect the data
  - Workload metric (Active)
    - Relates the power of the system to the maximum possible IOPS generated by a specific random stress load. It is reported as IOPS/watt
    - Power required to randomly supply data to and from a host
  - Bandwidth metric (Active)
    - Relates the power drown by the system to the maximum possible MBPS generated by a specific sequential stress load. It is reported as MBPS/watt
    - Power required to stream data to  and from a host
  - Capacity Optimization Heuristic

# Emerald$^{TM}$ Measurement Specification Overview

- SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Written by SNIA Green TWG to measure recommended metrics
- What does this measurement specification do
  - Defines a proxy method to measure power efficiency of a storage system
  - Covers Online (disk), Near Online (MAID), Removable Media Library (Tape/Optical Library), Virtual Media Library
  - Supports measurement of block based storage
  - Provides a storage taxonomy
- What it does not yet do (future revisions)
  - Specify how to measure power efficiency of file system or object-based systems
  - Define measurement procedures for adjunct products or interconnect elements

# Measurement Specification Sections

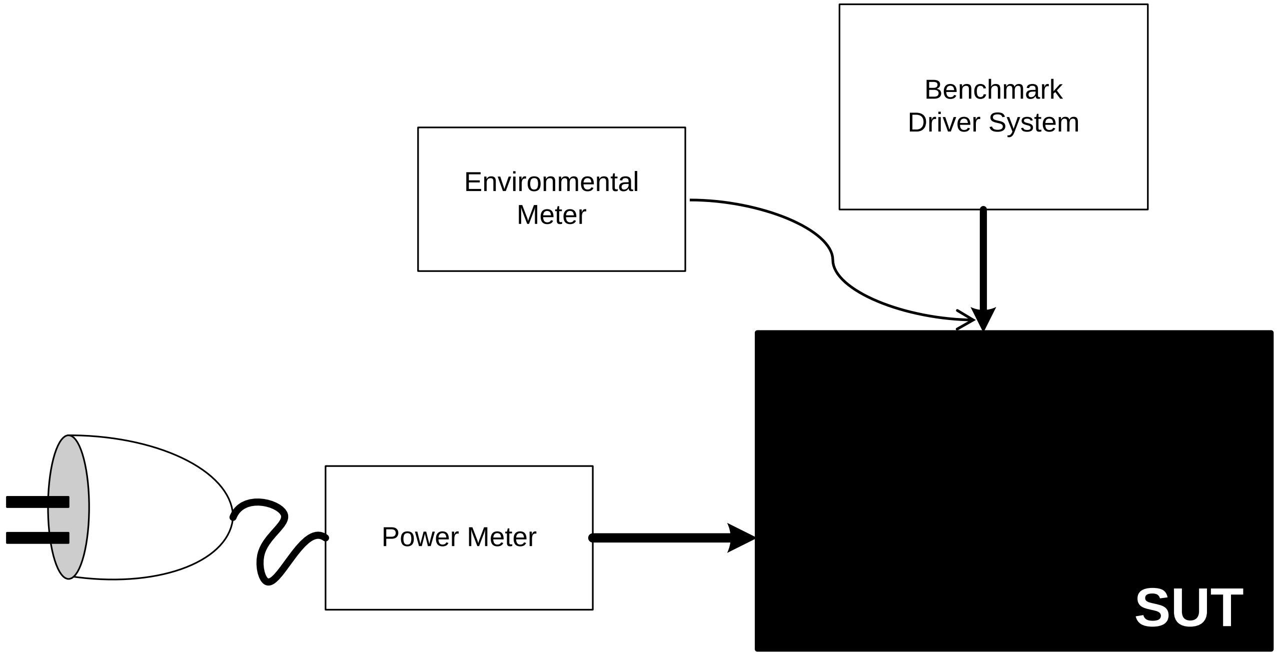
- Eight Sections
  - Sections 1 - 4 cover Overview, References, Scope, Definitions, Symbols, Abbreviations, and Conventions
  - Section 5 defines a storage taxonomy
  - Section 6 provides an top level overview of capacity optimization techniques
  - Section 7 describes the test procedure and requirements
    - Online
    - Near online
    - Removable Media
    - Virtual Media Library
  - Section 8 names the metrics generated from the test procedure

# Measurement Specification Execution

- Section 7 defines the test/measurement procedure
  - Defines configuration guidelines and instrumentation requirements
  - A section for each of the six categories for test execution
- Basic measurement specification procedure
  - Four continuous test phases
    - SUT Conditioning
    - Active test
    - Read Idle test
    - Capacity Optimization test
  - Each category could have a different requirements for each of the test phases

# Test Configuration and Requirements

- Basic configuration
  - Not allowed to change configuration or tune parameters during test phases



# SUT Conditioning Test Phase

- Intended to provide a uniform initial condition for subsequent measurements
- Demonstrate the SUT's ability to process IO requests
- Assure that each storage device in the SUT is fully operational and capable of satisfying any supported request
- Achieve typical operational temperature
- Each taxonomy category will have different measurement interval requirement to demonstrate stability

# Active Test IO Profiles

| IO Profile | IO Size (KiB) | Read/Write Percentage | IO Intensity | Transfer Alignment (KiB) | Access Pattern |
|---|---|---|---|---|---|
| Mixed Workload 1 (i=MW1) | 8 | 70/30 | 100 | 8 | Random |
| Mixed Workload 2 ( i=MW2) | 8 | 70/30 | 25 | 8 | Random |
| Random Write ( i=RW) | 8 | 0/100 | 100 | 8 | Random |
| Random Read ( i=RR) | 8 | 100/0 | 100 | 8 | Random |
| Sequential Write ( i=SW) | 256 | 0/100 | 100 | 256 | Sequential |
| Sequential Read ( i=SR) | 256 | 100/0 | 100 | 256 | Sequential |

- All or some of the IO profiles are used by the defined taxonomy categories
  - Drive enough IOs to reach the required response time or through-put specified in the measurement specification
  - The 25 IO intensity is 25% of the IO defined for MW1

# SUT Active Test Phase

- IO profiles used by taxonomy category
  - Online and Near-Online use all six IO profiles
  - Removable Media and VML use only the sequential IO profiles
- Run as an uninterrupted sequence of workloads
  - Specification defines the order to be run for each taxonomy category

# SUT Ready Idle Test Phase

- Defined as storage systems and components that 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.
- Average power measured in the measurement window
- No external IO given by the host
- Can perform any IO within the taxonomy required response time interval

# SUT Capacity Optimization Method Test Phase

- Heuristic tests
  - Delta snapshots
  - Thin provisioning
  - Data de-duplication
  - Parity RAID
  - Compression
- Run after ready idle test phase
- C program generated by SNIA
  - Download from sourceforge.net/projects/sniadeduptest
  - Used for de-duplication and compression
- Taxonomy dependent

# Metrics

- Active (Primary)
  - Ratio of operations rate over average power for the same measurement interval
    - $EP_{MW1}$ (IOP/s/W) of the 70% mixed workload at maximum response time
    - $EP_{MW2}$ (IOP/s/W) of the 25% of the IO used in MW1
    - $EP_{RR1}$ (IOP/s/W) of the random read workload at maximum response time
    - $EP_{RW1}$ (IOP/s/W) of the random write workload at maximum response time
    - $EP_{SR1}$ (MiB/s/W) of the sequential read workload at maximum throughput
    - $EP_{SW1}$ (MiB/s/W) of the sequential write workload at maximum throughput
  - Number of active metrics generated dependent on the taxonomy category tested

| IO Profile | IO Size (KiB) | Read/Write Percentage | IO Intensity | Transfer Alignment (KiB) | Access Pattern |
|---|---|---|---|---|---|
| Mixed Workload 1 (i=MW1) | 8 | 70/30 | 100 | 8 | Random |
| Mixed Workload 2 ( i=MW2) | 8 | 70/30 | 25 | 8 | Random |
| Random Write ( i=RW) | 8 | 0/100 | 100 | 8 | Random |
| Random Read ( i=RR) | 8 | 100/0 | 100 | 8 | Random |
| Sequential Write ( i=SW) | 256 | 0/100 | 100 | 256 | Sequential |
| Sequential Read ( i=SR) | 256 | 100/0 | 100 | 256 | Sequential |

# Metrics (Continued)

- Ready Idle (Primary)
  - Ratio of raw capacity over average power measured in the defined measurement window (GB/W)
- Capacity Optimization (Secondary)
  - A yes/no for each Capacity Optimization Method tested
  - Do not have to test all COMs but if vendor declares to have a COM it must be tested and on during active test phase

# Flow Needed for Valid Emerald Measurement

| Setup: System to test identification (following slides ) Collection meters, workload drivers | → | System Setup: Pre-runs and tuning of the system under test Verify stability of the system | → | Test of System Run complete Emerald test | → | Review data and generate metrics Verify system stability | → | Submit data to SNIA Emerald™ |
|---|---|---|---|---|---|---|---|---|

- General timeline
  - Tune the system
  - A day to run test
  - A day to generate the required data and review it
  - A few hours to submit the data

# Test Data Report of the Emerald Program

- First page
  - List Vendor information
  - System information

**The SNIA Emerald Test Data Report**

**Disclosure for storage systems and products**

**NOTICE: This document is published and made accessible by SNIA for non-commercial use only and subject to restrictions stated in the Terms of Use contained herein.**

**Product Description**

| | | |
|---|---|---|
| Company | Best Storage Company | |
| Address | 1234 Storage Drive | |
| (address line 2) | | company logo here |
| (address line 3) | | |
| Municipality | Somewhere, World | |
| Product Name | Store100 | |
| Taxonomy Category | online 3 | |
| Product Release Date | 1-Jan-12 | |
| Description | Store100 is a disk based storage system capabil of doing it all | |
| Product Web Page | WWW.Beststoragecompany.org | |
| List Price (optional) | US | |
| Raw capacity | 100000 GB | ( ### GiB) |
| Submission Date | | Last status change |
| Document Status | | SNIA tracking # |
| | Mandatory items | Optional items |

# Test Data Report of the Emerald Program

- Second Page
  - Idle Metric
    - Raw Capacity/Average Power
  - Sequential Metrics
    - MiB per Second/Average Power
  - Random Metrics
    - IO per Second/ Average Power
    - If taxonomy supports Random
  - Capacity Optimization Results
    - List if available and on during test
  - Other disclosures
    - Here vendor should list system configuration optimization

## Operational Power

**Idle power test**

| | | | |
|---|---|---|---|
| Average watts | 3500 W | | |
| Raw capacity tested | 10000 GB | | 2.66086 GiB |
| $EP_{RI}$ | 2.66 GB/W | | 2.420 GiB/W |
| Standard idle metric | GB per Watt | Note: 1 GB = 10^9 bytes; 1 GiB = 2^30 bytes<br>a GiB is about 7.4% larger than a GB | |

**Active power tests**

| | | | | | | |
|---|---|---|---|---|---|---|
| $EP_{RR}$ | 15 | run length (minutes) | 40 | Average latency | 25 | ms |
| Small random reads | I/Os per second per Watt | | | | | |
| $EP_{RW}$ | 10 | run length (minutes) | 40 | Average latency | 12 | ms |
| Small random writes | I/Os per second per Watt | | | | | |
| $EP_{SR}$ | 3 | run length (minutes) | 40 | Average latency | 10 | ms |
| Large sequential reads | MiB per second, per Watt | | | | | |
| $EP_{SW}$ | 2 | run length (minutes) | 40 | Average latency | 14 | ms |
| Large sequential writes | MiB per second, per Watt | | | | | |
| $EP_{MW1}$ | 12 | run length (minutes) | 50 | Average latency | 27 | ms |
| Mixed workload 1 | I/Os per second per Watt | | | | | |
| 70% random, 30% sequential, I/O intensity = 100 | | | | | | |
| $EP_{MW2}$ | 3 | run length (minutes) | 45 | Average latency | 5 | ms |
| Mixed workload 2 | I/Os per second per Watt | | | | | |
| 70% random, 30% sequential, I/O intensity = 25 | | | | | | |

NOTE: power-related numbers are required to be reported to three significant digits

| Capacity Optimizations | On during test? | Available in SUT? |
|---|---|---|
| Deduplication | Yes | Yes |
| Compression | No | Yes |
| Thin provisioning | No | No |
| Parity RAID | Yes | Yes |
| Read-only delta snapshots | No | No |
| Writeable delta snapshots | No | No |

**Other mandatory disclosures, per spec**

System was optimized for the random workloads with the 10000K SAS drives

# Test Data Report of the Emerald Program

- Following pages list
  - Specification and test infrastructure
  - Cabinet information
  - Controller information
  - Power supply efficiency rating
  - Shelve information
  - Storage Media
- Question for You!
  - What information would you like in the report? Let us know at: sniaemerald.com

**Product Configuration - Controller(s)**

| | | |
|---|---|---|
| Part number | BSC-1000 | |
| Description | | |
| Rack "U" height | 1 | |
| # of controller units | 1 | |
| HA configuration | none | |
| Main memory | 4294967296 GiB | (base 2 arithmetic) |
| NVRAM | 1.048576 MiB | (base 2 arithmetic) |
| Supported RAID levels | 0, 1, 4/5, 6 | |
| Default RAID level | RAID 1 | |
| RAID level during tests | RAID 6 | |
| Other RAID info (if any) | | |
| Snapshots supported | no | |
| Max snapshots/container | | |
| Max snapshots/controller | | |
| Disk Scrubbing enabled | no | |
| Multi-pathing enabled: | | |
| to storage | no | |
| to hosts | no | |
| Any single points of failure? | yes | |
| Variable speed fans | yes | |
| Number of hot spares | 1 | |
| Other | | |

**Power Supply ECOS (80+) Ratings**

| | | | |
|---|---|---|---|
| PS #1 Part number | OF100 | Rating | 80+ bronze |
| PS #2 Part number | OF200 | Rating | none |
| PS #3 Part number | | Rating | |
| PS #4 Part number | | Rating | |
| PS #5 Part number | | Rating | |

See http://www.80plus.org for ratings and lists of manufacturers

# Test Configuration Challenges for Vendor

Education SNIA

- Wide Spectrum of Storage-Oriented Products
  - Created a taxonomy to narrow scope
  - Categories: On-Line, Near-Line, etc.
  - Classifications: Further granularity of each Category
- Still too Broad in Scope
  - Vendors may have multiple products in a particular Category/Classification
  - Each product may have many configuration variables
- Requirement/Challenge: Select Appropriate Test Configurations
  - Comprehensive and usable results for customer
  - Minimized, lower cost, but effective testing methods for vendor

# Concept of Product and Family

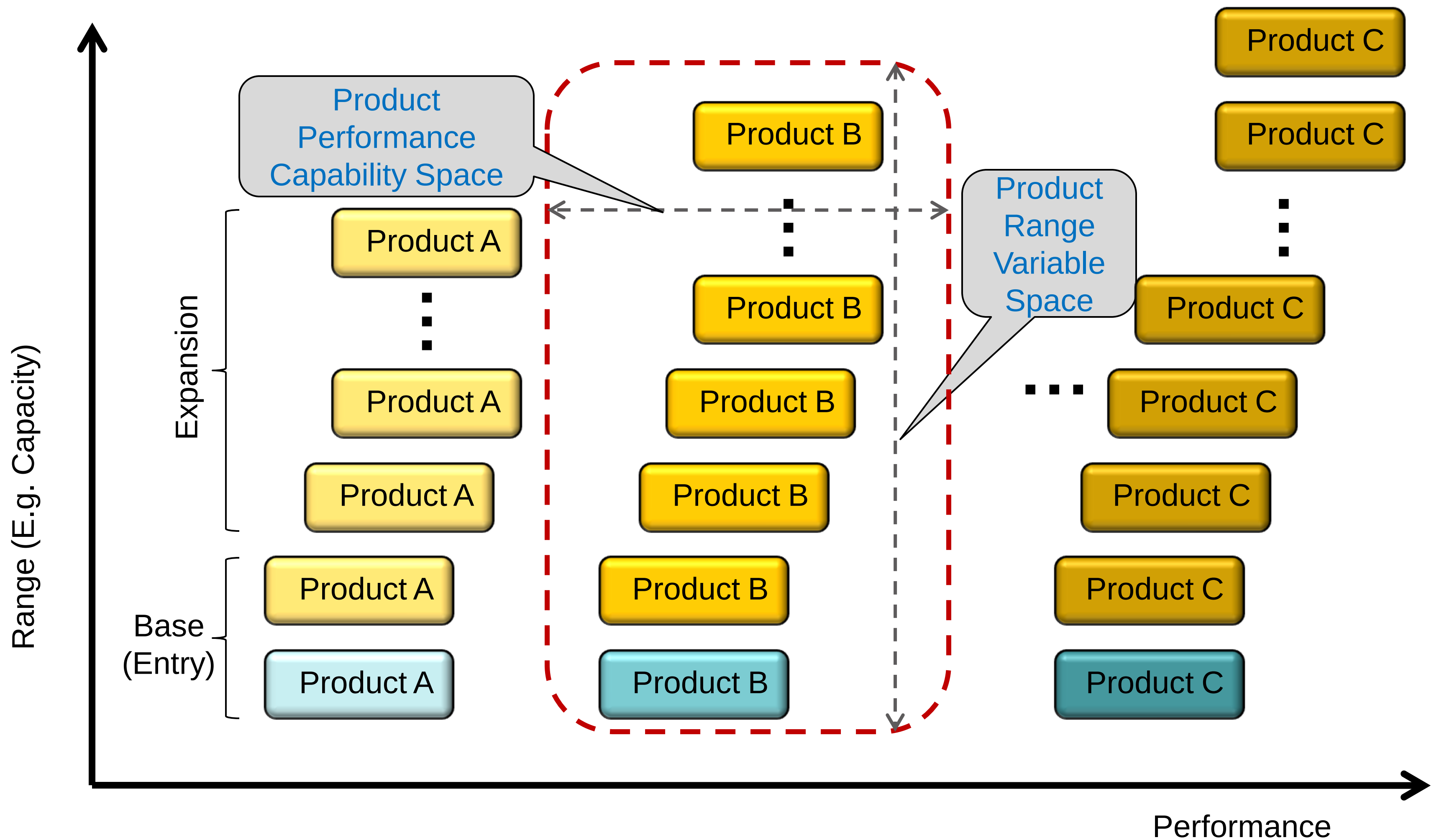
- **Product:**
  - Represents a fundamental performance capability space that separates it from any other potentially related products
- **Product Family:**
  - Represents the full range space of configuration variables and options for a particular product.
- **Term Usage:**
  - Terms family and range are used interchangeably and may include such aspects as number and type of storage device (spinning or solid state drive), cache size, availability levels, etc.

# Approach

Education SNIA

- Vendor Aligns Product(s) with SNIA Taxonomy Category
  - Hopefully straightforward – Taxonomy will adapt over time
- Vendor Aligns Product(s) with Category Classification
  - Will be some boundary gray areas - E.g. OL-3 or OL-4?
- Vendor Further Defines Product/Family Configurations
  - The really hard part…
- Conceptual Representation
  - Next slide depicts a possible product/family (range) differentiation
  - Believed applicable to most storage system architectures

# Simplified Product/Family Representation



# Product/Family Discussion

- Products Could be of Various Architectural Types
  - Monolithic – Little or no scaling but may still have family aspects
  - Scale-up – E.g. base controller + storage expansion
  - Scale-out – E.g. base compute/storage + compute/storage expansion
  - Others TBD
- Product Performance Typically Scales With Expansion
  - Varying degrees
    - Scale-up performance typically rolls off at varying degrees before max configuration
    - Scale-out performance can be linear with increasing configurations
  - Any inter-product performance overlap driven by vendor's market positioning

# Family (Range) Discussion

- Range Variables
  - Example on previous product/family depiction focuses on capacity
  - Could involve other variables
- Range Variable Types
  - Particular Items of highest potential energy consumption impact:
    - Controller or related compute element – Typically defines performance aspect
    - Cache – Also performance oriented - Not considered part of the user-addressable space
    - Number and type of persistent storage devices – Defines user-addressable space
    - RAS items – As necessary to meet reliability, availability, serviceability requirements
    - Capacity optimization – Functionality (typically software) that more effectively utilizes physical storage space such as thin provisioning, compression and de-duplication
  - Many other examples
    - Power supplies, cooling, I/O, etc.

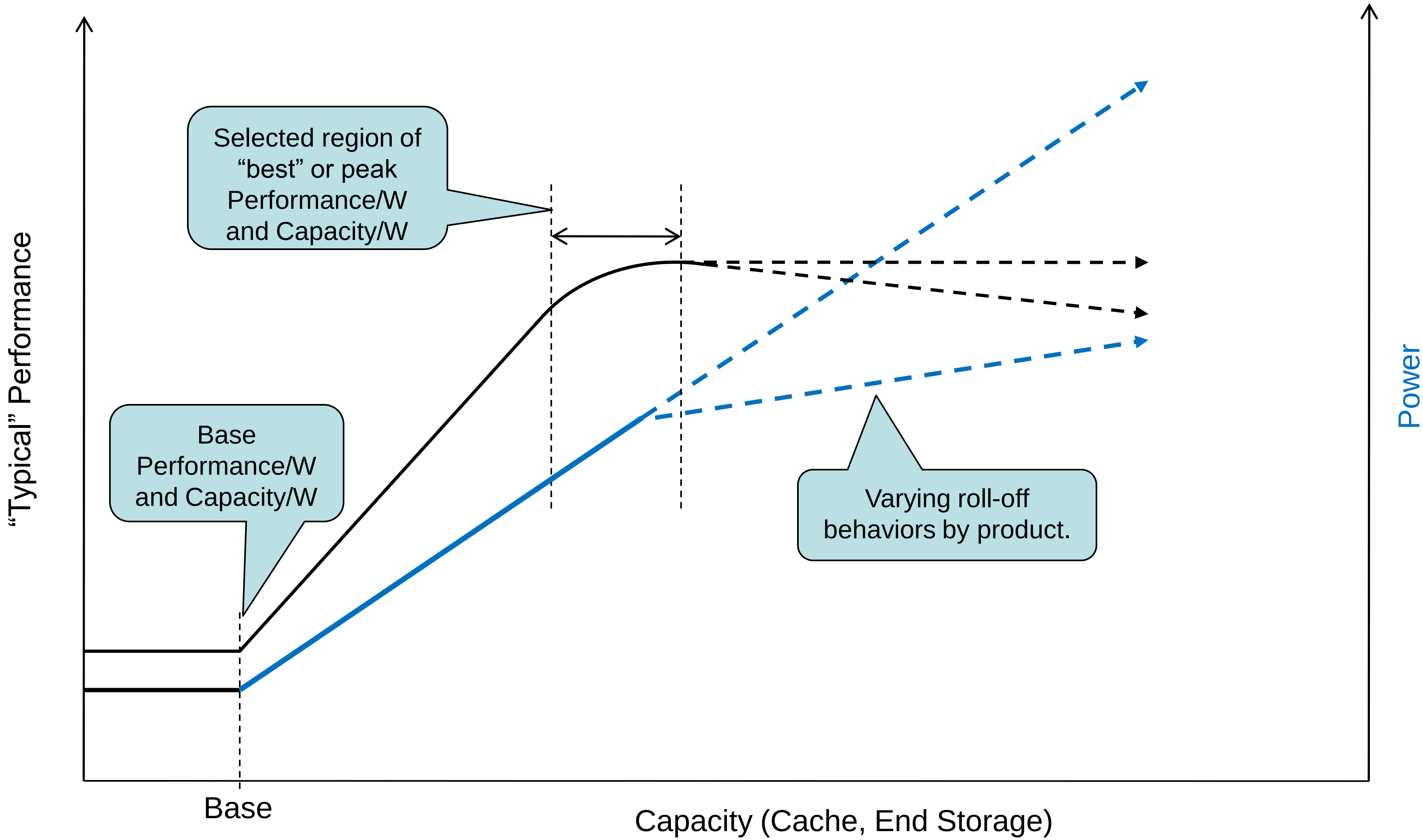
# Approach to Range Variable Reduction

- Range Variable Reduction is Difficult
  - Even with the 5 listed items still too many test cases
    - Significant set-up and execution times
    - Complex results sets
  - Maximum system size testing is expensive and cumbersome to manage
  - Need a simpler alternative…
- "Best Foot Forward" (aka Sweet Spot) - BFF
  - Find proxy family configuration(s)
    - Reasonably representative of the all range variables?
  - Find test point(s) where Measurement Specification active metrics are best
    - The "sweet spot"
  - Suitable for any architecture
    - E.g. scale-up, scale-out, hybrid, …

# Best Foot Forward Approach

- BFF Looks Holistically at Storage System Product/Family
  - Allows vendor to select and test one product/family configuration
    - Or more if desired
  - At operating points near the Measurement Spec metric peak values
    - I.e. the “sweet spot”
  - Results reasonably representative of the entire family
    - Easier and less expensive for the vendor
    - Simple and understandable results for the potential customer
- Scale Up Example on Following Slide
  - Based on notion that Measurement Spec active metrics have peak values
  - Peaks typically located at points well below maximum configurations

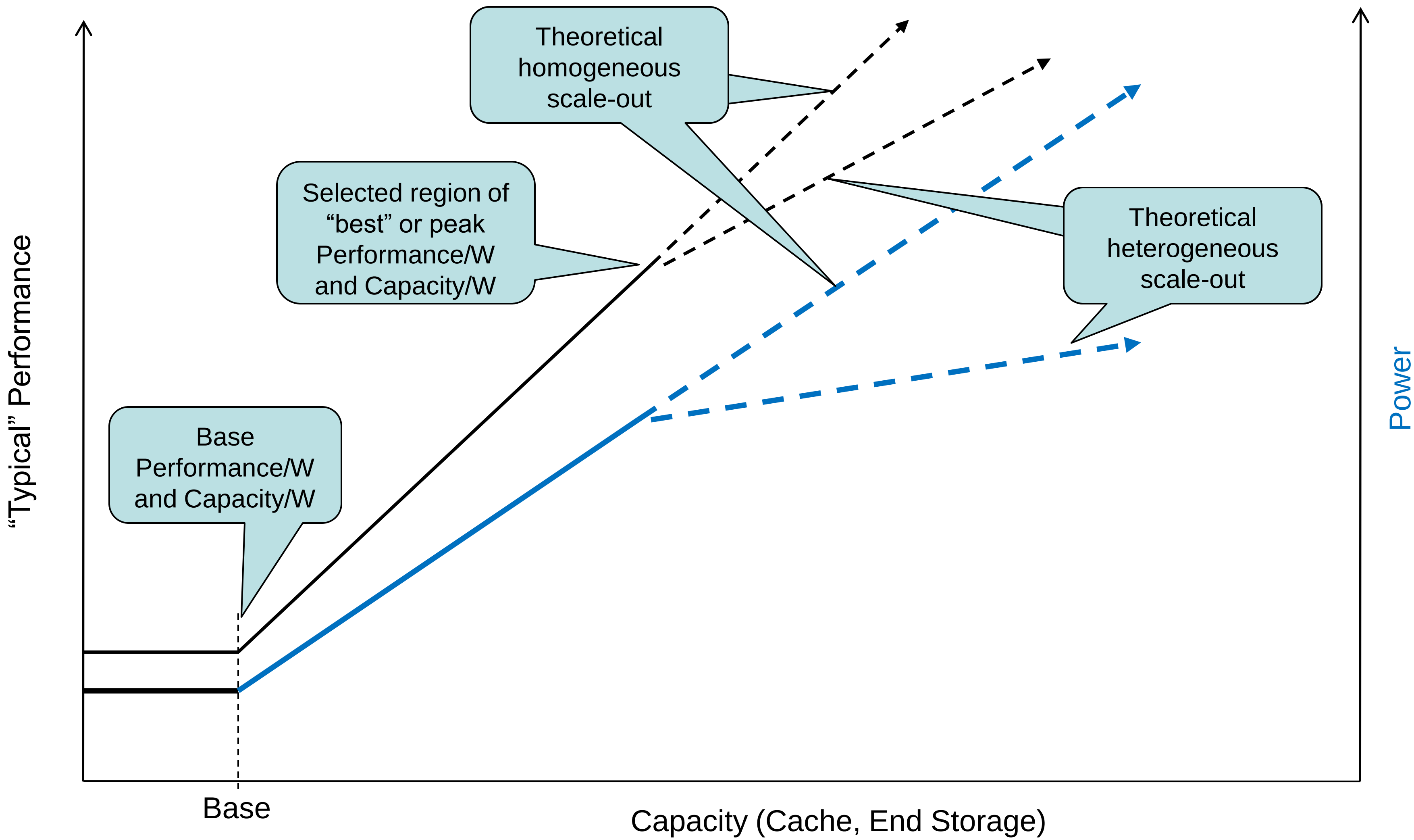
# Best Foot Forward Approach
# Scale-Up System



# Best Foot Forward Approach

- Previous Slide is a Rough Approximation
  - Capacity increases are actually more stepwise
  - Performance roll-off can vary by product
    - Dashed lines attempt to show one (of possibly many) changes due to different storage technology tiers, e.g. scaling capacity w/large SATA drives
  - Regardless, example depicts a smaller test configuration
- What About Other Test Points?
  - Could also test at base (entry point) but not required
  - Key is no requirement to test beyond the peak point
- Scale Out Example on Following Slide
  - What if there is no clearly discernable peak?

# Best Foot Forward Approach Scale-Out System



# Best Foot Forward Approach

- Again a Rough Approximation
  - Capacity increases are actually more stepwise
  - Dashed lines attempt to show one (of possibly many) changes due homogeneous vs heterogeneous scale-out configurations
  - Can still select a smaller test configuration

# What Dose it Mean IT Managers

- Need to know your workload
- Look for a best foot forward that will match your load (sequential or random)
- What kind of growth do you need to match systems configuration
  - May have to choose be scale up or scale out
  - Note in scale out it is the SAN that has the performance can suffer
  - The sweet spot is where you get the most performance/ watt

# Candidate SUT: A shipping Online-3 SAN

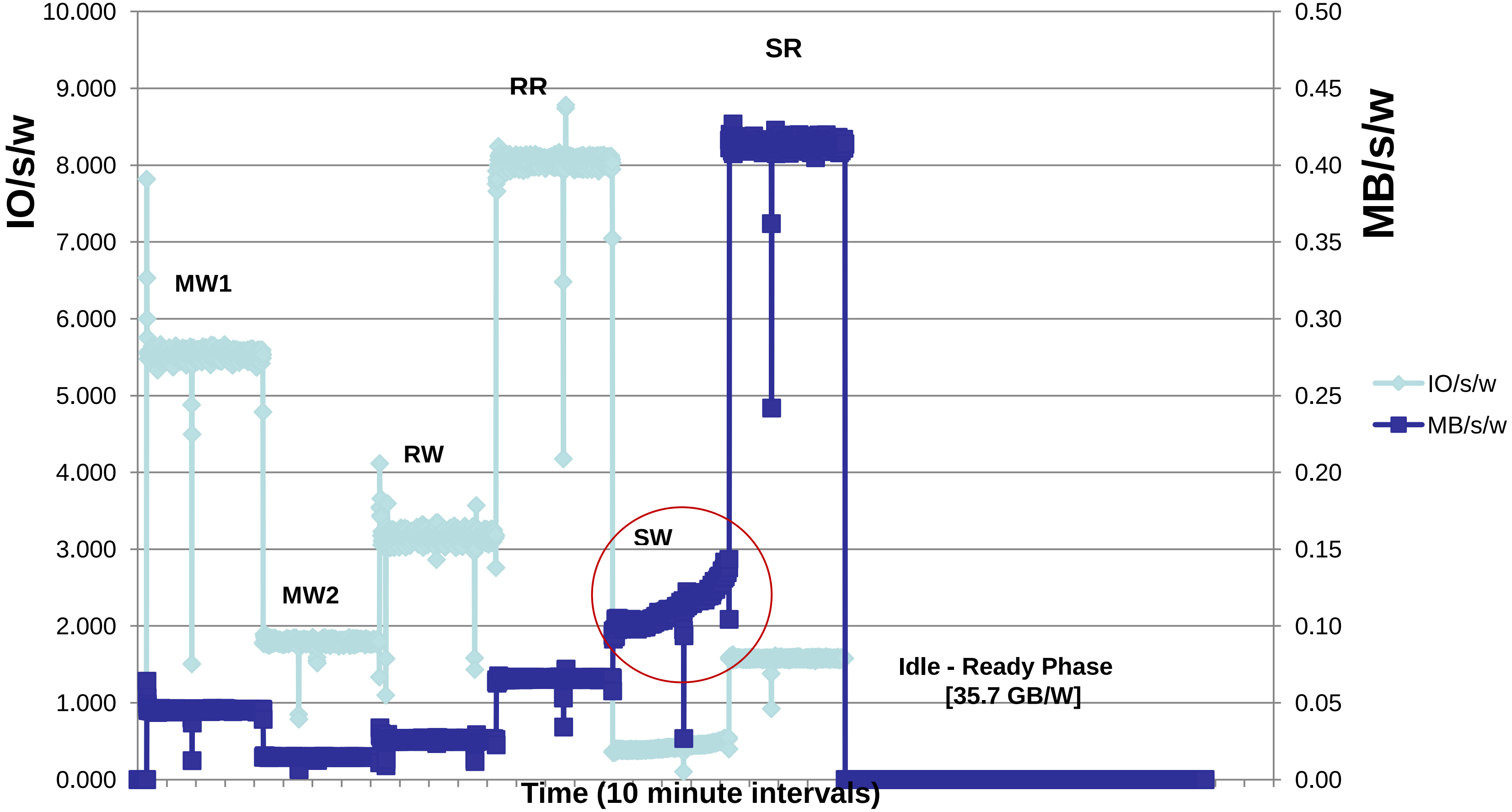
- Full redundancy except for single midplane in dual-controller enclosure
- Two controller performance points, with variable cache and front-end interfaces
- The lower product class can support 120xLFF or 250SFF  and the higher product class can support 240xLFF or 450xSFF
  - 12 x LFF drive shelves
  - 25 x SFF drive shelves

Supported drives, 6Gb SAS
  - SFF
    - 146GB, 15K
    - 300GB, 10K
    - 450GB, 10K
    - 600GB, 10K
    - 500GB, 7.2K midline
    - 200GB SSD*
    - 400GB SSD*
  - LFF
    - 300GB, 15K
    - 450GB, 15K
    - 600GB, 15K
    - 2TB, 7.2K midline

* Will characterize SSD's separate from spinning drives

# Baseline Test Results for Candidate SUT

Starting available config: 370 SAS drives ( no “tuning”)
10 shelves x 25 SFF drives (10K-300GB)
10 shelves x 12 LFF drives (15K-600GB)



# Baseline Test Observations

- Peak workload efficiency metric occurred during RR phase ( ~ 8 IOP/s/W)
- Peak throughput efficiency metric [MB/s/W] occurred during SR phase (~ 0.42 MB/s/W)
- Power consumed for any workload varies only ~12% (4100W at idle to 4600W during RR) → performance can have a bigger influence on the metric
- Note:  Easy to observe that SeqWrite (SW) measurement interval has not reached stability; stability is required in order to have a valid metric measurement

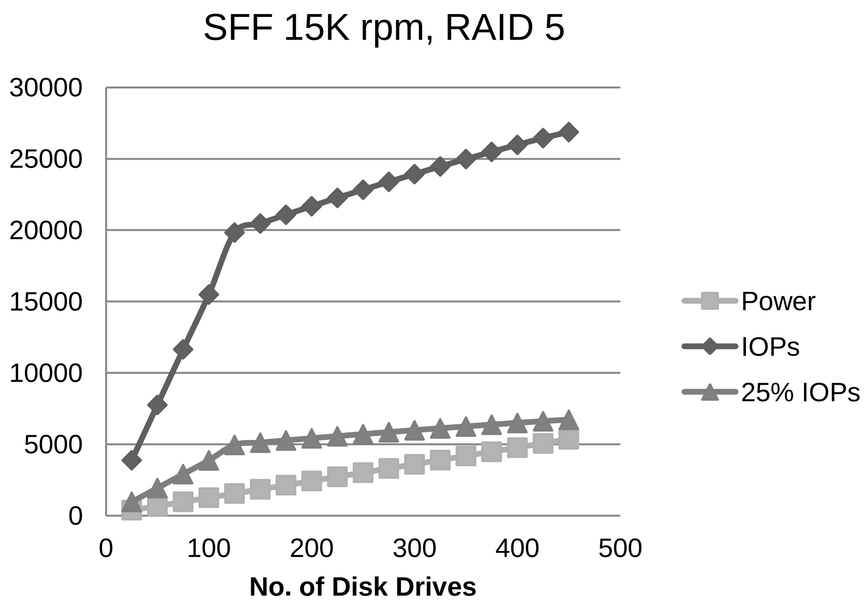
# Finding the Best Foot Forward

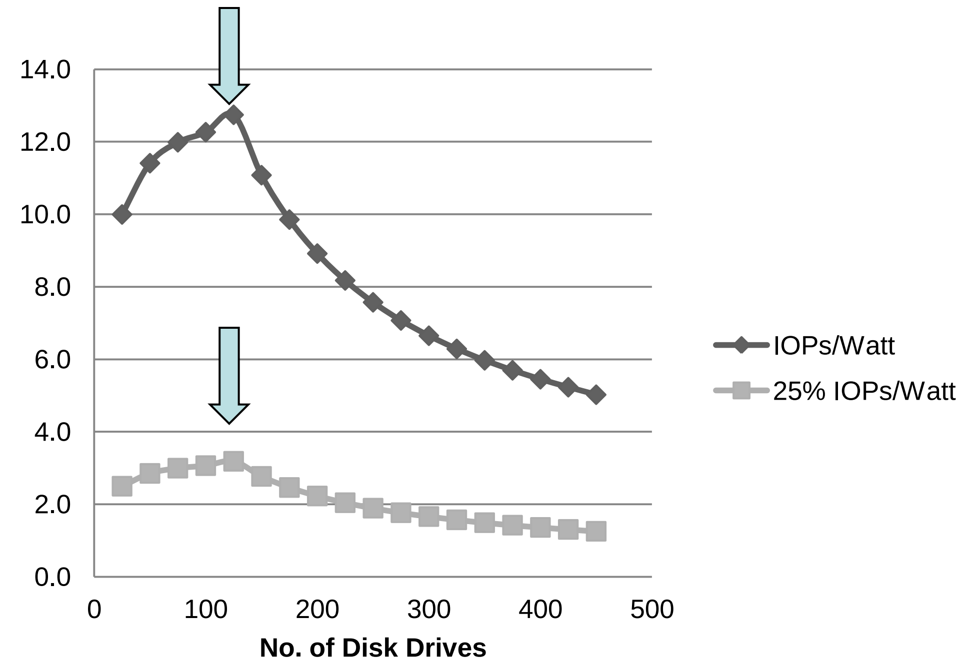
- While there are 7 different Emerald test profiles (online); you may have from 1 to 7 possible different optimized configurations – vendor choice
  - 4 x Random [IOP/s/Watt]
  - 2 x Sequential [MiB/s/Watt]
  - 1 x Ready-Idle [raw capacity, GB/Watt]
- Recommended to use estimator tools that combine power and performance to predict the peak metrics
  - The alternative is educated derivations and potentially a lot of testing that is very labor and resource intensive
  - As long as the simulated results are reasonably accurate, the physical configuration selected for actual test to measure the peak value can be limited in range

# Predicted Peak Metrics for an Online-3 Test Candidate

| Exercise # | Prediction basis |
|---|---|
| 1<br>1.5 | Mixed Workload, Random 70/30 R/W<br>-- Granular level, single drives |
| 2 | Random Read (100/0 R/W) & Random Write (0/100 R/W) |
| 3 | Sequential Read (100/0 R/W) & Sequential Write (0/100 R/W) |
| 4 | RAID level |
| 5 | Ready Idle |

# Exercise 1: Mixed Workload 8K Random 70/30 R/W



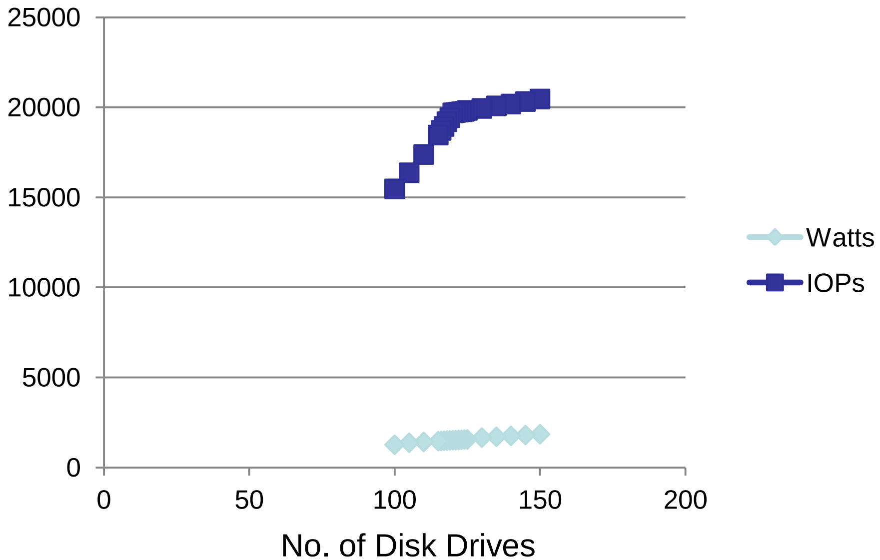



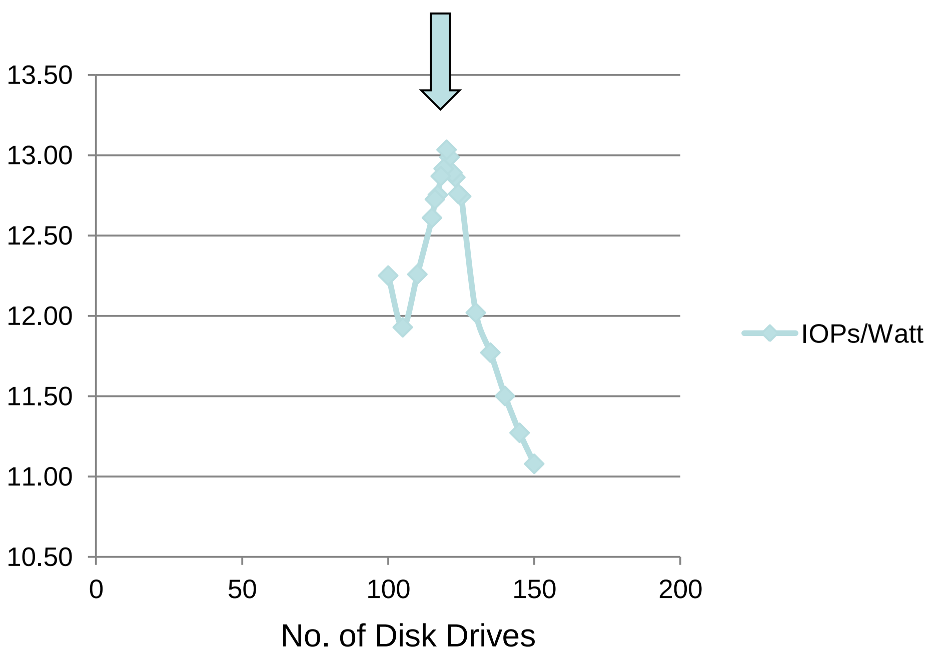

- Peak metric = 12.7 IOP/s/Watt at 125 drives
- Changing the read/write mix changed the metric but not the drive count
  60/40 r/w = 11.5 IOP/s/W; 80/20 r/w = 14.9 IOP/s/W

Note: Incrementing drive count by full JBOD

# Exercise 1.5: Granular Drive Counts (Increment by Single HDDs)

8K Random, 70/30 R/W, SFF 15K rpm, RAID 5



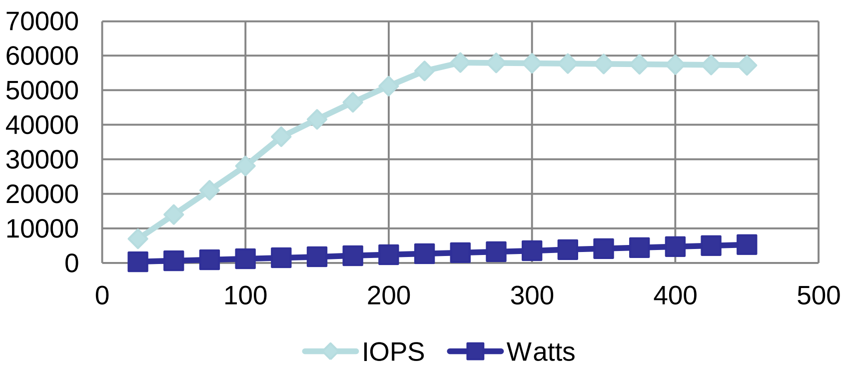

Peak Metric:
13.03 IOP/s/Watt & 120 HDDs

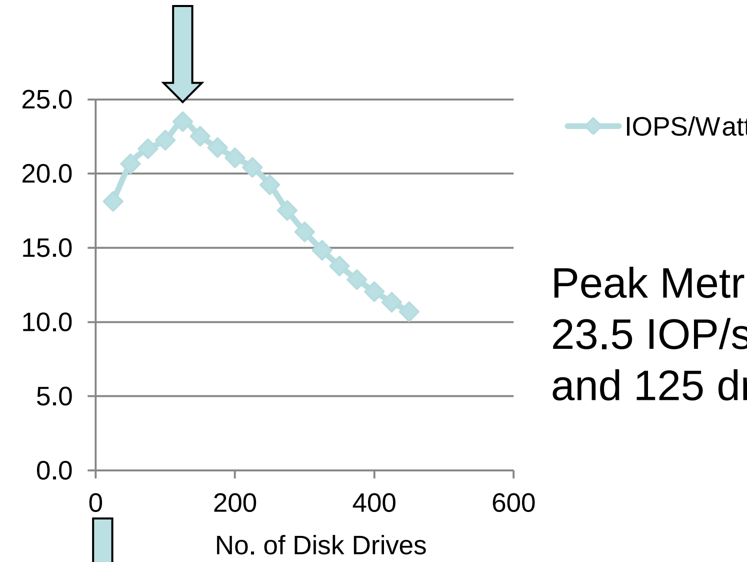
# Exercise 2: 8K Random Read, Write

Education SNIA

SFF 15K rpm, RAID 5

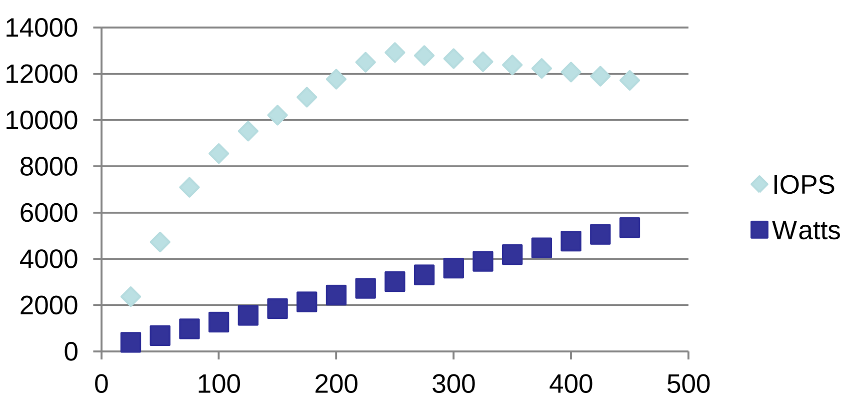
## Random Read

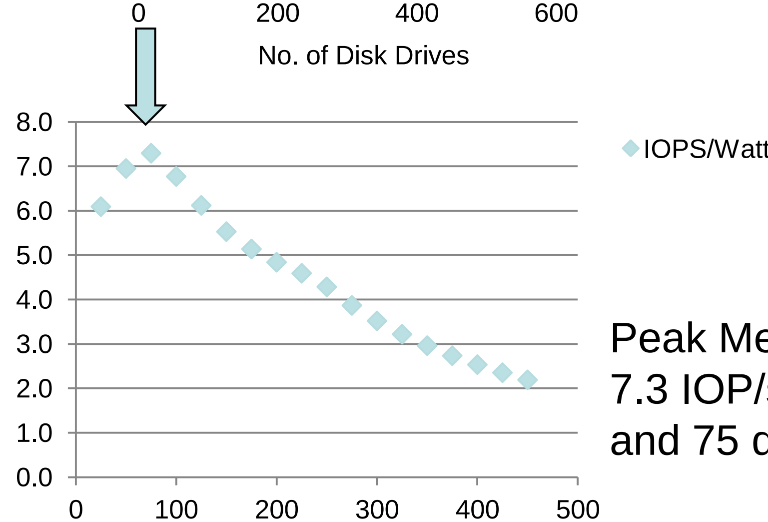



Peak Metric:
23.5 IOP/s/Watt
and 125 drives

## Random Write

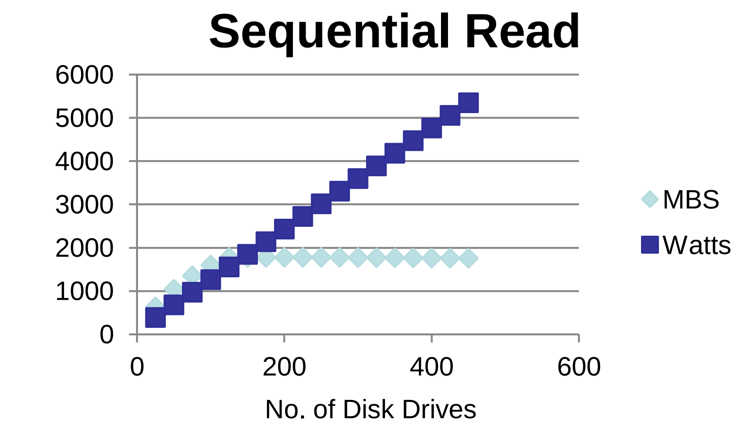


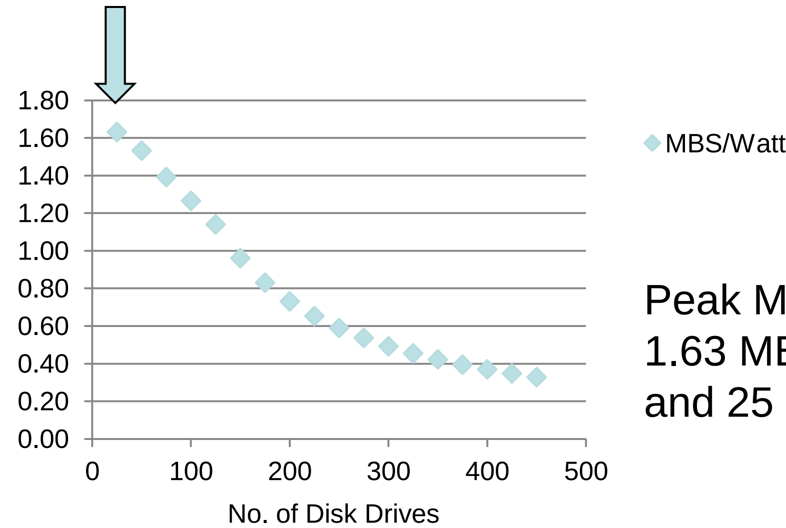




Peak Metric:
7.3 IOP/s/Watt
and 75 drives

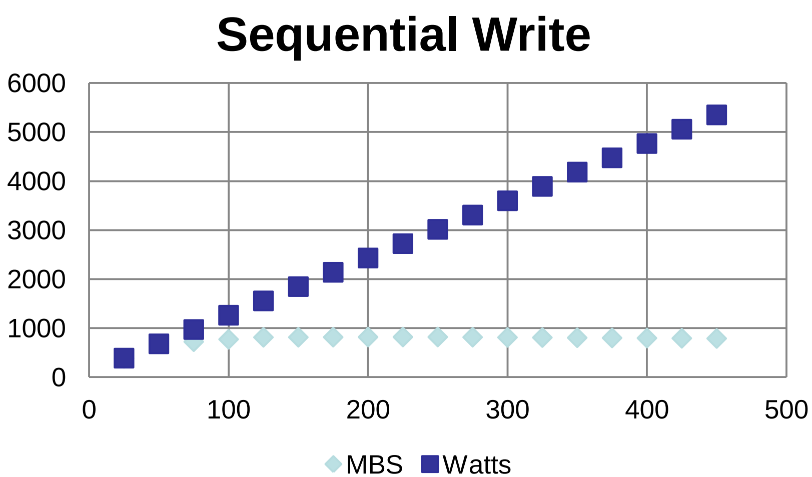
# Exercise 3: 256KB Sequential Read, Write

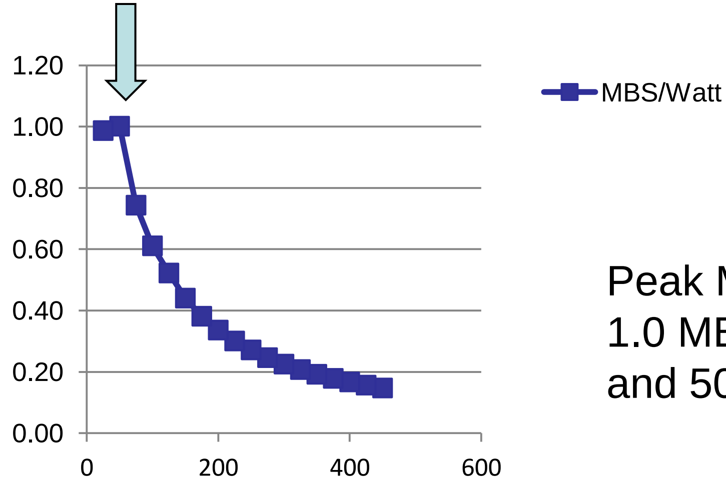
SFF 15K rpm, RAID 5




Peak Metric: 1.63 MB/s/Watt and 25 drives




Peak Metric: 1.0 MB/s/Watt and 50 drives

# Exercise 4: RAID level (SFF 15K rpm)

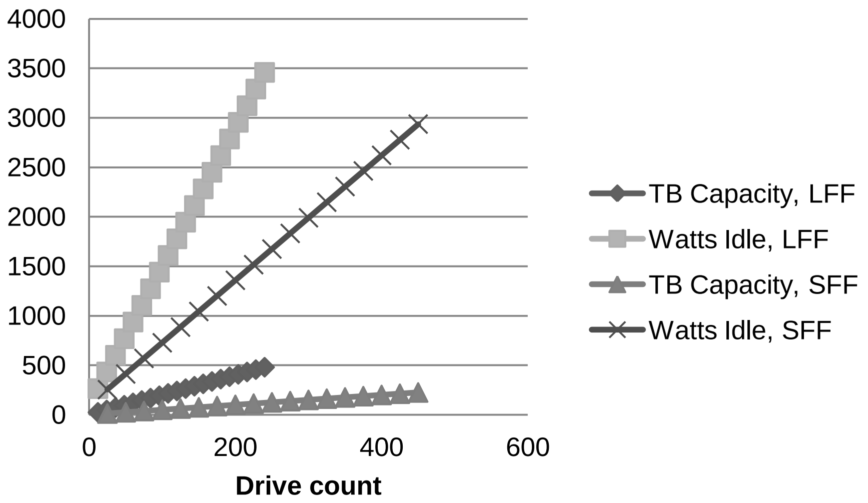
Peak Power Efficiency [IOP/s/Watt] or {MB/s/Watt}, # of HDDs

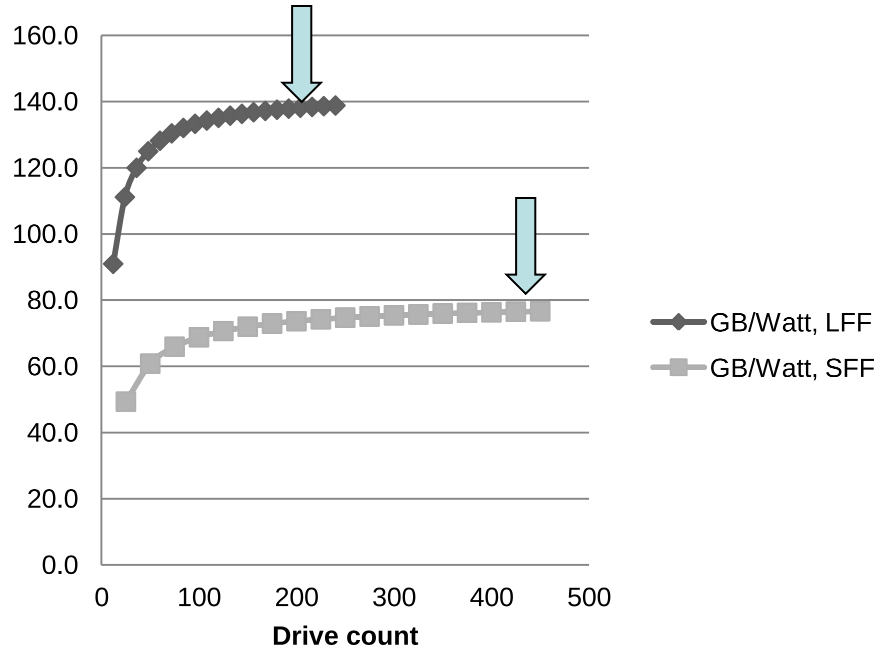
| | 8K Rand Mixed (70/30 R/W) | 8K Rand Read | 8K Rand Write | 128K Seq Read | 128K Seq Write |
|---|---|---|---|---|---|
| RAID 5 Distributed single parity | [12.7], 125 | [23.5], 125 | [7.3], 75 | {1.63}, 25 | {1.00}, 50 |
| RAID 10 Blocks striped and mirrored | [18.6], 125 | [23.2], 125 | [13.3], 75 | {2.0}, 25 | {0.68}, 50 |

Notes:
1) The Online 3 category is required to have RAID protection
2) RAID 6 (double parity) will offer greater redundancy but poorer performance efficiency

# Exercise 5: Ready-Idle

LFF 2TB 7.2K rpm and SFF 500GB 7.2K rpm drives at Ready-Idle




Peak Metrics:
LFF: 138.8 GB/W and 240 drives
SFF: 76.6 GB/W and 450 drives

# General Observations for the Candidate SUT

Education SNIA

- Active cases - the performance* reaches a roll off point relatively early (i.e., smaller drive count); then it either levels out or goes down slightly. The peak [Performance/Power] metric is reached at or before this performance roll off point
  - All peak predictions for Random are reached with the same drive type (15K, SFF) and close in drive count (125 or 75)
  - All peak predictions for Sequential reached with the same drive type (15K, SFF) and close in drive count (50 or 25)
- Ready-idle case - the peak metric levels but continues to slowly rise with drive count (as the controller electronics power is amortized over increasing numbers of drives)
  - Vendor can choose to test and submit a lower drive count configuration, and add to the notes a power calculator based projection for the largest drive count configuration

*Note: very dependent on specific Controller performance and bandwidth behavior

# Sample Data Submission (Online-3 SUT)

Education SNIA

## Operational Power

**Idle power test**

| | | |
|---|---|---|
| Average watts | 592.692 | W |
| Raw capacity tested | 7300 | GB |
| $EP_{RI}$ | 12.317 | GB/W |
| Standard idle metric | GB per Watt | Note: 1 GB = 10^9 byt s<br>a GiB is about 7.4% larger than a GB |

**Active power tests**

| | | | Run length (minutes) | | Average latency | |
|---|---|---|---|---|---|---|
| $EP_{RR}$<br>Small random reads | 17.255<br>I/Os per second per Watt | run length (minutes) | 30 | Average latency | 12 | ms |
| $EP_{RW}$<br>Small random writes | 9.155<br>I/Os per second per Watt | run length (minutes) | 30 | Average latency | 7 | ms |
| $EP_{SR}$<br>Large sequential reads | 3.01<br>MiB per second, per Watt | run length (minutes) | 30 | Average latency | 6 | ms |
| $EP_{SW}$<br>Large sequential writes | 1.22<br>MiB per second, per Watt | run length (minutes) | 30 | Average latency | 9 | ms |
| $EP_{MW1}$<br>Mixed workload 1<br>70% random, 30% sequential, I/O intensity = 100 | 10.501<br>I/Os per second per Watt | run length (minutes) | 30 | Average latency | 13 | ms |
| $EP_{MW2}$<br>Mixed workload 2<br>70% random, 30% sequential, I/O intensity = 25 | 3.486<br>I/Os per second per Watt | run length (minutes) | 30 | Average latency | 4 | ms |

NOTE: power-related numbers are required to be reported to three significant digits

SNIA Emerald™

## The SNIA Emerald Test Data Report

**Disclosure for storage systems and products**

| Capacity Optimizations | On during test? | Available in SUT? |
|---|---|---|
| Deduplication | | NO |
| Compression | | NO |
| Thin provisioning | | YES |
| Parity RAID | YES | |
| Read-only delta snapshots | | YES |
| Writeable delta snapshots | | YES |

**Other mandatory disclosures, per spec**

Test data provided is for a specific configuration that is tuned to achieve the best SeqRead and SeqWrite performance (the "sweet-spot"). The sweet spot data for alternate configurations that are tuned for the best Random and Idle metrics will be added in the near future.

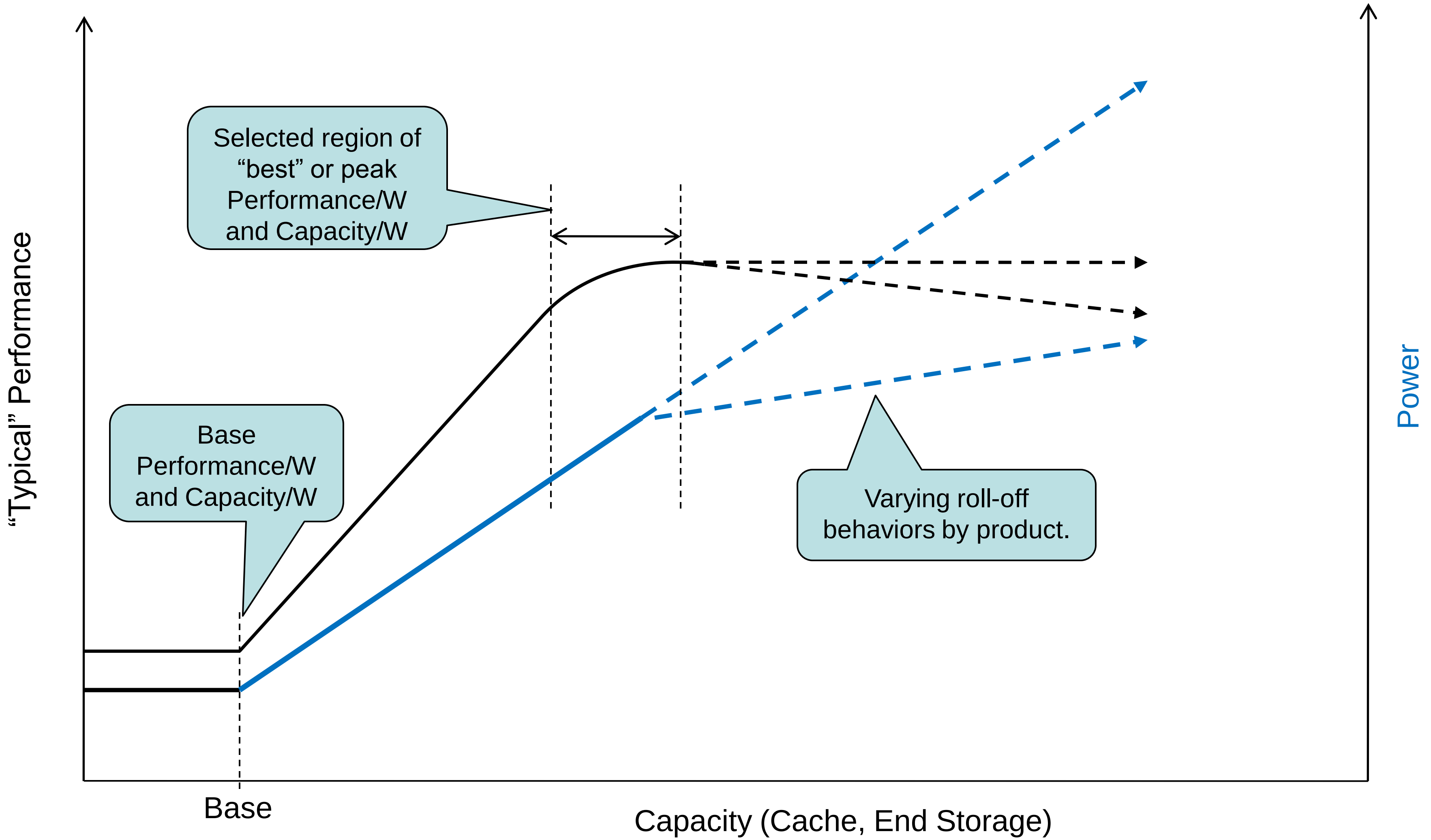
# What dose this Mean to IT Managers

- Provides operational power of a storage system
  - Most labels list peak power requirements (circuit loads/safety)
  - Compare systems based on workload
  - Capacity power efficiency
- Configuration trade offs for workload, performance, RAS features  and power
  - Different configuration if majority of your work is random or sequential
  - Need to understand response time and performance requirements and its impacts on power usage
  - Reliability, Availability, Serviceability
    - Impacts on both active and capacity power efficiency

# What dose this Mean to IT Managers Continued

- When comparing power efficiency numbers understand the system configuration
  - Was the system setup for random or sequential workloads
  - Is the system highly available or not (ie. lots of RAS features)

# Power Performance Peak



# Agenda

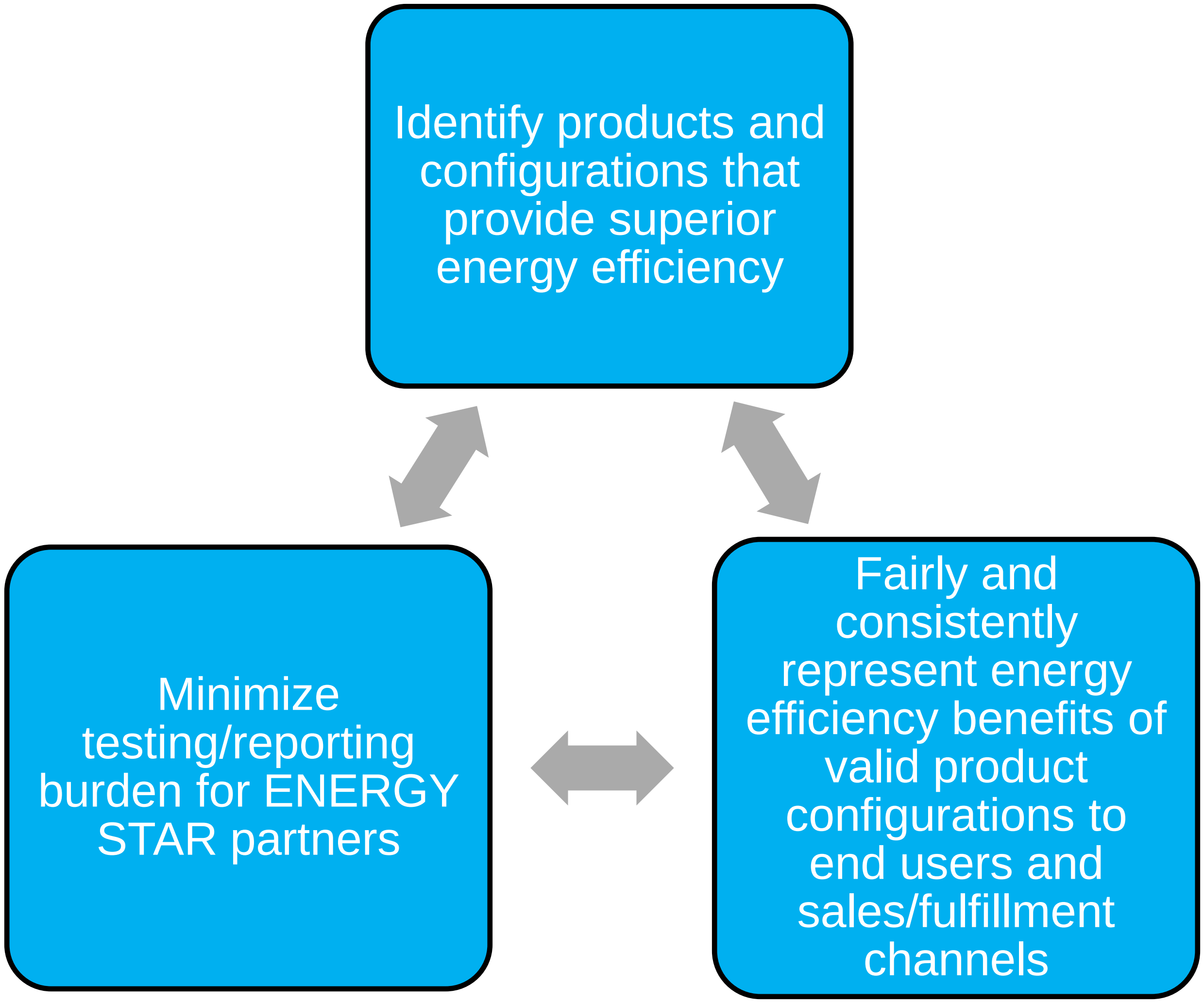
- Overview, Definitions, and what does green storage mean
- SNIA storage green activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald™ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald™ Program
  - SNIA Emerald™ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- **Other Associations' Green Storage Efforts**
- Storage technologies for energy savings
- Typical savings

# ENERGY STAR

## ENERGY STAR for Data Center Storage

- Specification in development
- Draft 2 released October 2011
- EPA is
  - Open and willing to listen
  - Learning the technical and business details
  - Collaboration with industry is a key goal (while meeting the EPA needs and mission)
- Stakeholders participation
  - Wide cross-industry participation – vendors & suppliers
  - SPC,The Green Grid, Wikibon, PG&E, Climate Savers
  - Some end-user participation

# ENERGY STAR Goals




Copy from October 2011 Stakeholders meeting Fall SNW

# ENERGY STAR for Data Center Storage Timeline

| Date | Document | Effective Date (Tier 1) | Pages |
|---|---|---|---|
| April 2009 | Stakeholder Letter | TBD | 2 |
| June 2009 | Framework Document | March 2010 | 12 |
| 20 July 2009 | 1st Stakeholder Meeting | March 2010 | -- |
| 15 October 2009 | Test Procedure Meeting | TBD | -- |
| 2 February 2010 | 2nd Stakeholder Meeting | TBD | -- |
| Jan/Feb/March 2010 | 1st Round Data Collection | | |
| 9 April 2010 | Draft 1 | TBD | 18 |
| 20 July 2010 | 3rd Stakeholder Meeting | TBD | -- |
| April 2011 | Supplemental Data Collection | TBD | |
| July 2011 | Stakeholder Meeting | July 18 | |
| October 2011 | Stakeholder Meeting | | |

“this is one of the most complex areas we have tackled to date” – Andrew Fanara, Team Leader, EPA ENERGY STAR Product Specifications, July 2009 in San Jose

# Other Associations Green Storage Efforts

- SNIA Emerald
  - A complete set of tests intended to measure power use ratios based on all previously mentioned metrics.
- SPC
  - Storage Performance Council mainly oriented to disk subsystems was the first industry association to add power to their benchmark
- The Green Grid
  - Data Center Maturity Model
  - Design guide in progress
  - Working on a usage metrics
    - Measure the IOP/s/W, MiBs/W, GB/W in the datacenter

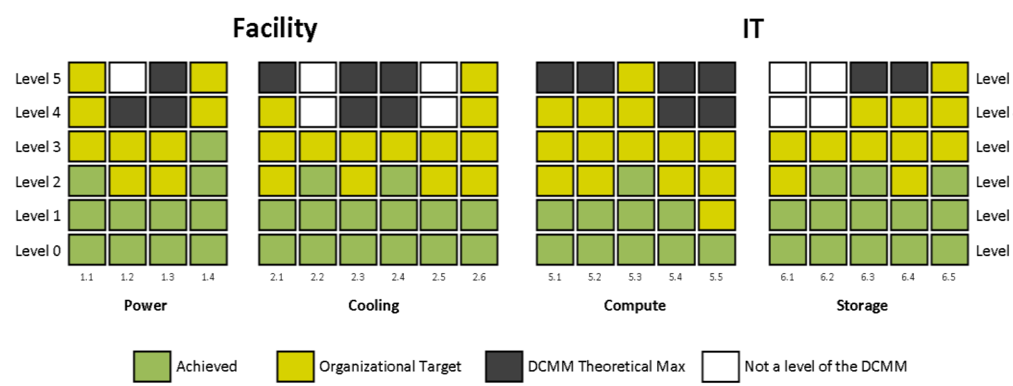
# The Green Grid Data Center Maturity Model

Education SNIA

- “The Green Grid identified the need for a comprehensive model of what could and should be done over time throughout the data center to improve overall energy efficiency and sustainability.”
- Looks at
  - Facilities
    - Power
    - Cooling
    - Other
  - IT
    - Compute
    - Storage
    - Network
    - Other

# The Green Grid
# Data Center Maturity Model Continue

- Has levels of data center maturity for energy efficiency
  - Level 0: Minimal
  - Level 1: Part best practice
  - Level 2: Best practice
  - Level 3 and 4: Reasonable steps between current best practices and visionary
  - Level 5: Visionary – (approximately five years away)
- Usage
  - Map current state
  - Evaluate how to get where you want to be
  - Move to the next level

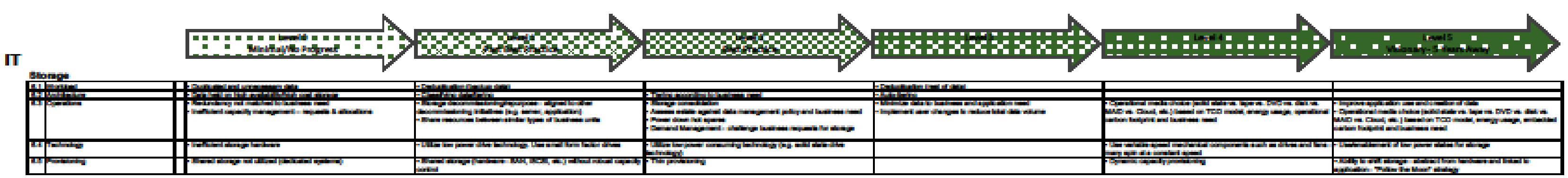


# The Green Grid Data Center Maturity Model Continue Storage

Education SNIA

- Six categories to score
  - Workload
    - Have many duplicated and unnecessary data → Deduplication
  - Architecture
    - Data held on high availability/high cost storage → Auto-tiering
  - Operations
    - Inefficiency capacity management → Operational media choice based TCO, usage
  - Technology
    - Inefficient storage hardware → Use/enablement of low power states of storage
  - Provisioning
    - Dedicated systems → Ability to shift storage (linked to applications)

**Data Center Maturity Model**



# The Green Grid Data Center Storage Efficiency Metrics DCsE

- White paper by the Green Grid
  - Defines operational metrics that IT managers can use to judge how green is their storage usage
  - Metrics
    - Capacity (User Capacity in Use/Data Center Storage Power Consumptions)
    - Workload (Data Center I/O throughput / Data Center Storage Power Consumptions)
    - Throughput (Data Center Transfer Throughput / Data Center Storage Power Consumptions)
  - Provides suggestions on way to generate this metrics

# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA storage green activities
  - Green Storage Initiative
  - Green Storage  Technical Working Group
  - SNIA Emerald$^{TM}$ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald$^{TM}$ Program
  - SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

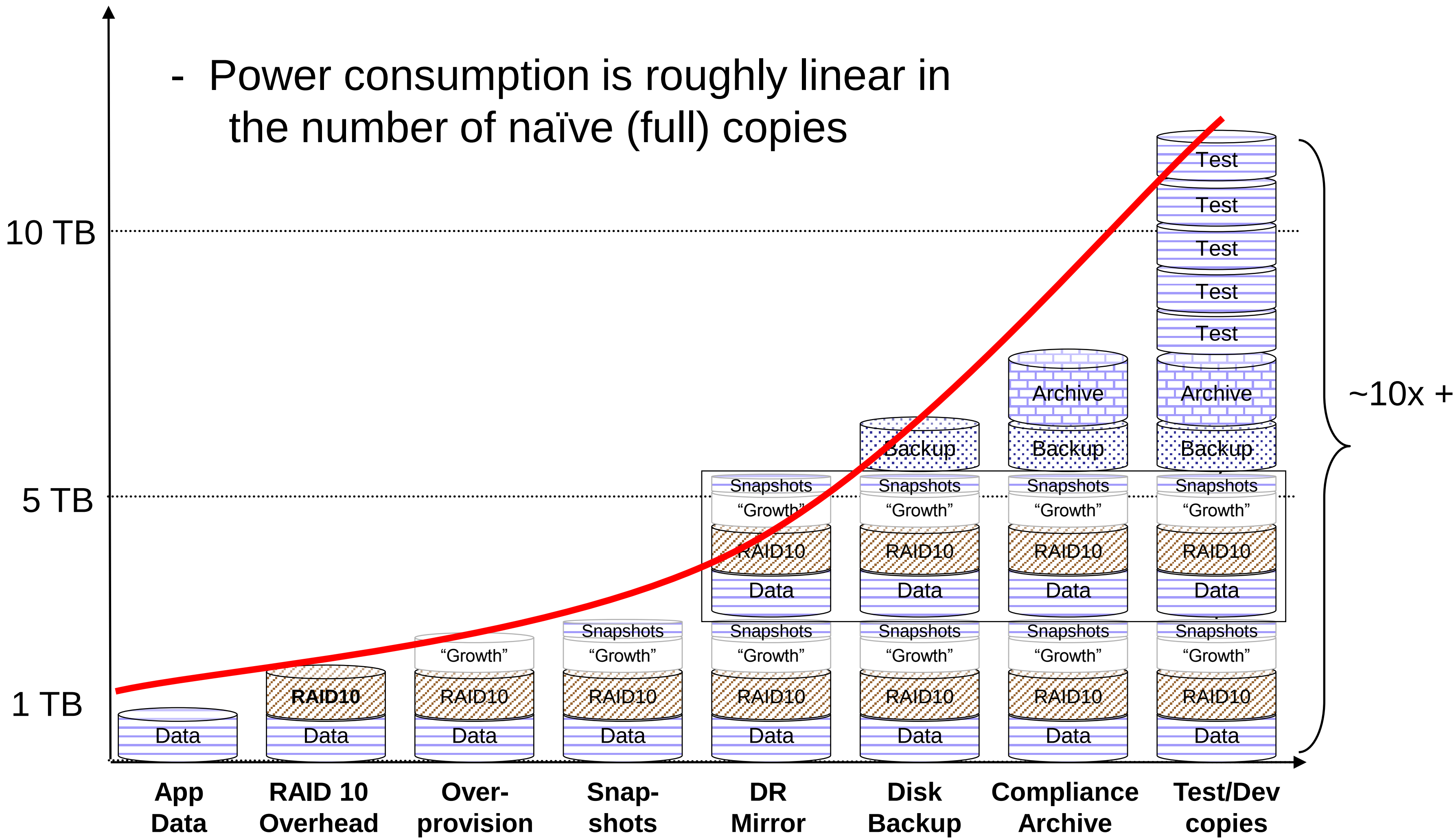
# Objective

- Get more (storage) work done for less money
- Translation: reduce data center footprint
  - in space
    - less storage equipment to buy, power and maintain
  - in energy
    - more energy-efficient equipment
    - less equipment to cool
    - better cooling methodologies
    - better power management
  - in administrative costs
    - less storage equipment to manage

# Problem 1: need for redundancy

- RAID 10 – protect against multiple disk failures
- DR Mirror – protect against whole-site disasters
- Backups – protect against failures and unintentional deletions/changes
- Compliance archive – protect against heavy fines
- Test/dev copies – protect live data from mutilation by unbaked code
- Overprovisioning – protect against volume-out-of-space application crashes
- Snapshots – quicker and more efficient backups and PIT (point in time) copies

# Result of redundancy



# Problem 2: making heat just to cool it

- Servers, storage and switches are HEATERS
  - 100% efficient energy-to-heat conversion
  - Rotating media uses 85% of max power *at idle*!
- A/C is a big "undo" mechanism for overheating
  - But less than 100% efficient (typically 70%)

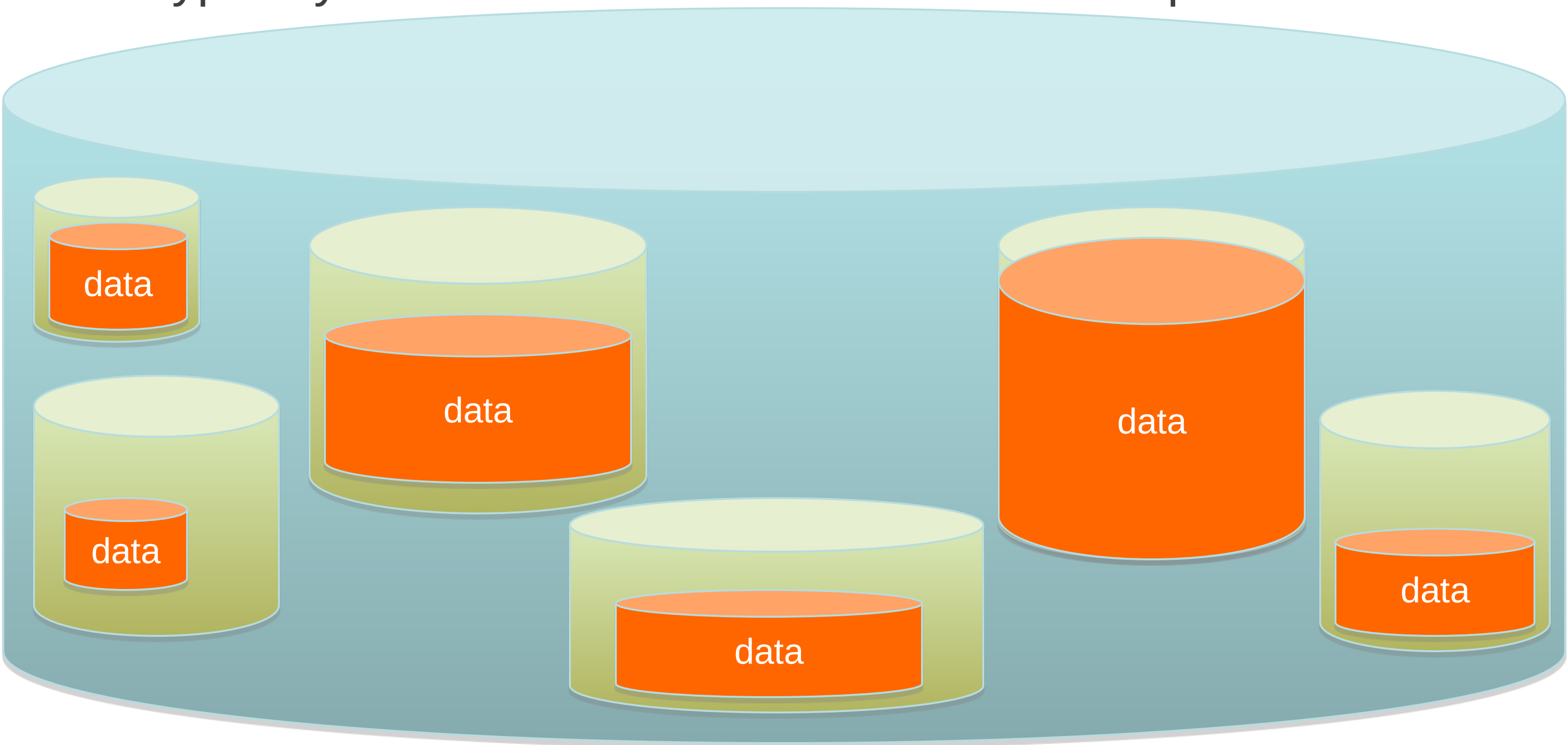
> 60% of the power in a traditional data center does no IT work

(PUE* ~ 2.5)


* PUE defined later

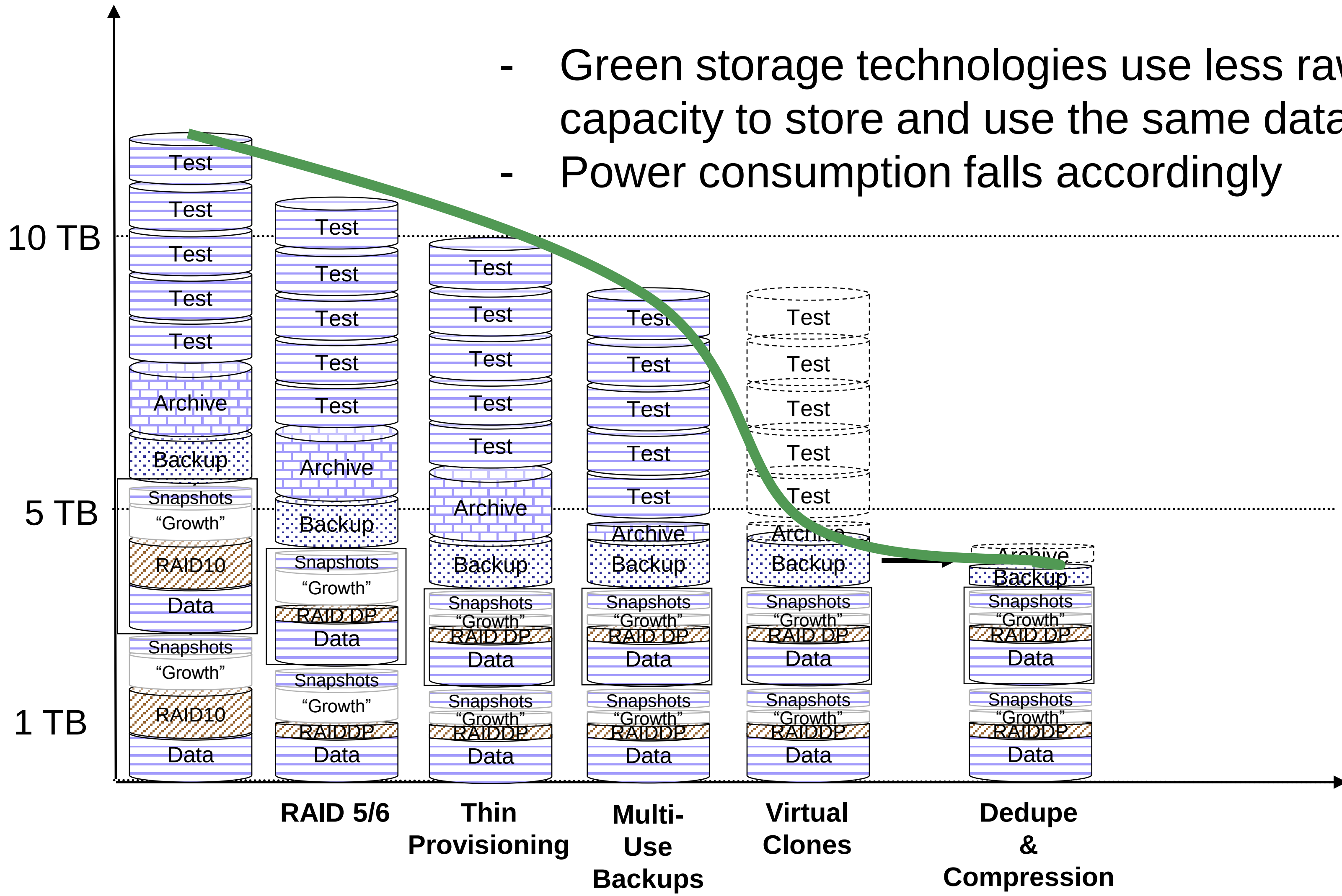
# Problem 3: unused space

- Overprovisioning of systems
- Overprovisioning of containers
  - Typically 30% to 40% utilization of available space



# Solution: green storage technologies

- Green storage technologies use less raw capacity to store and use the same data set
- Power consumption falls accordingly

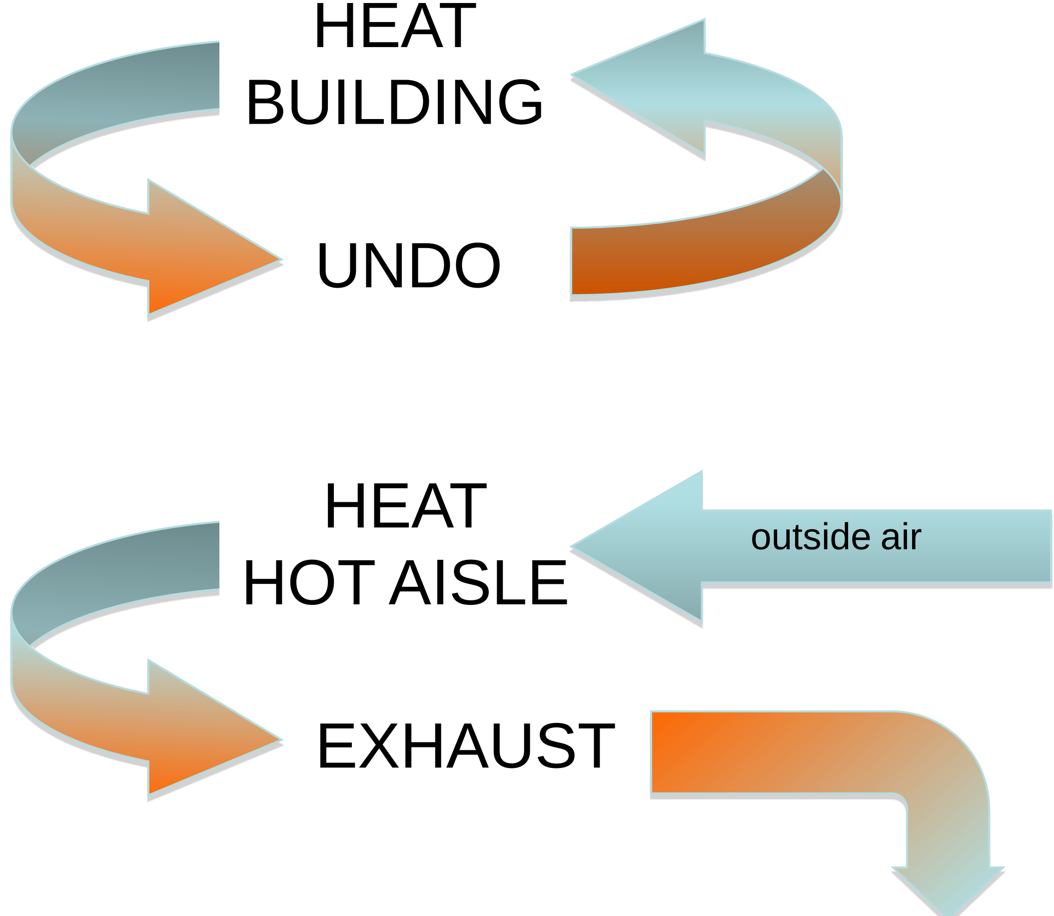


# Solution: green facilities technologies

e.g.

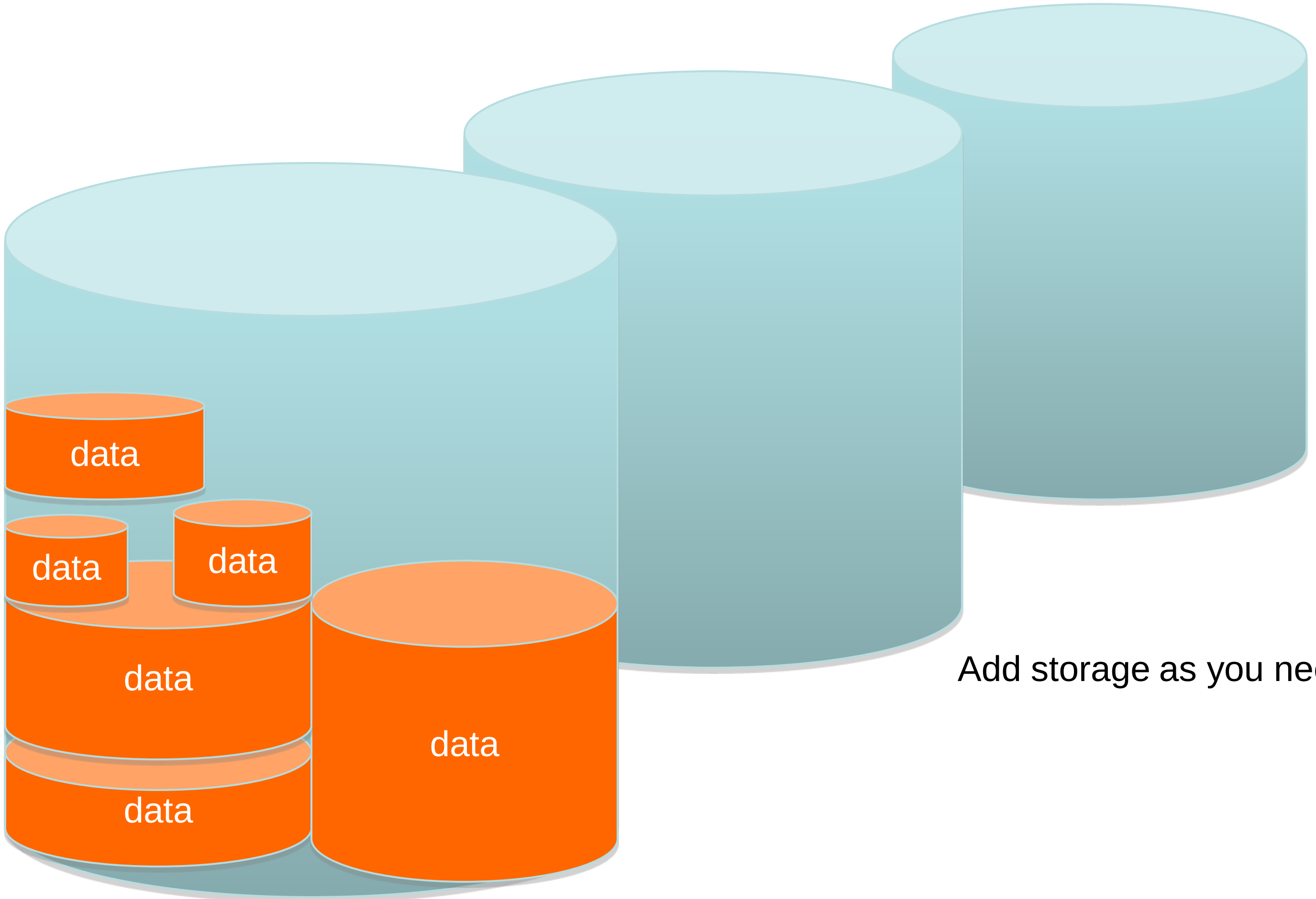
BEFORE

AFTER




**40% SAVINGS!**
**(PUE = 1.5)**

# Solution: getting back free space




Add storage as you need it

# Green Storage Technologies

- Enabling technologies
  - Storage virtualization
  - Storage capacity planning
- Green software
  - Compression
  - Delta snapshots
  - Thin provisioning
  - Non-mirrored RAID
  - Deduplication and SIS

# Green storage technologies (cont.)

- Other storage technologies and power saving techniques
  - Capacity vs. high performance drives
  - ILM / HSM / Tiering
  - MAID
  - SSDs / "Flash and stash"
  - Power supply and fan efficiencies
- Facilities-side technologies
  - Hot aisle/cold aisle
  - Water & natural cooling
  - Flywheel UPSs
  - Spam filtering

# Enabling technologies

- Storage virtualization
- Storage capacity planning

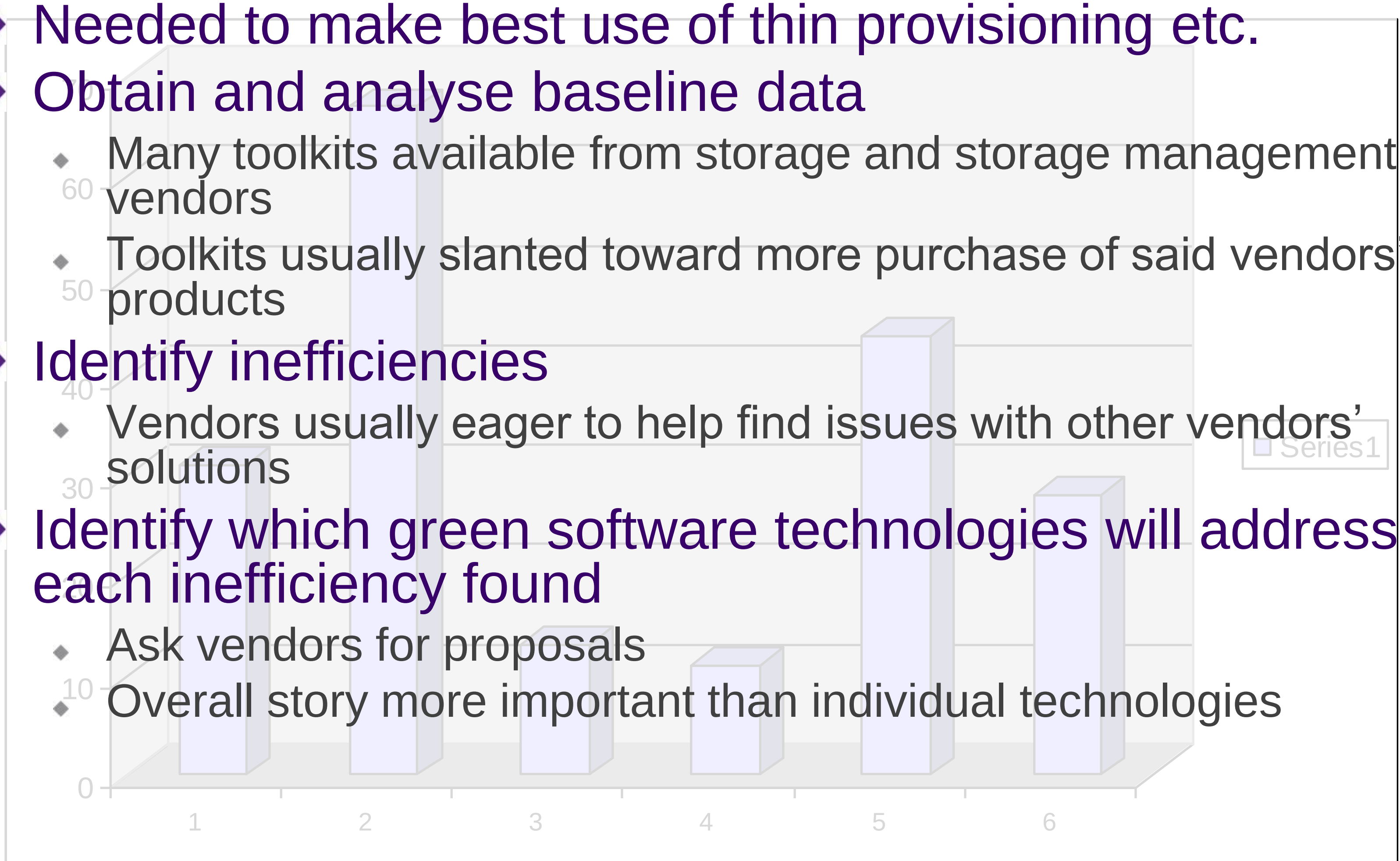
# Storage Virtualization

Education SNIA

- Mapping from physical location to virtual location
  - May exist at multiple layers
- In and of itself, not green wrt storage
  - No reduction in dataset size – it's just pointer remapping
- But foundational for most green storage technologies
  - Thin provisioning
  - Delta snapshots
- Also contributes in other areas
  - Flexibility, manageability, etc.
  - Fundamental for storage clouds

# Storage capacity planning

- Needed to make best use of thin provisioning etc.
- Obtain and analyse baseline data
  - Many toolkits available from storage and storage management vendors
  - Toolkits usually slanted toward more purchase of said vendors' products
- Identify inefficiencies
  - Vendors usually eager to help find issues with other vendors' solutions
- Identify which green software technologies will address each inefficiency found
  - Ask vendors for proposals
  - Overall story more important than individual technologies



# Green software technologies

- Compression
- Delta snapshots
- Thin provisioning
- Non-mirrored RAID
- Deduplication and SIS

# Compression

- Compression
  - Old and venerable
  - Origins in signaling, number and coding theory
    - Motivated by limited bandwidth and lossiness of satellite communications
  - Scattered throughout the data stack
  - Many formats already compressed
    - JPEG, MPEG, MP3, etc.
    - Lossless compression (LZW) necessary for unknown data types
  - Configuration matters
    - Compress before encrypting, decrypt before decompressing
  - Difficult in block-based environments
    - But becoming commonplace—usually in conjunction w/ dedup

# Delta snapshots

- NOT just wholesale copies of the data
  - We call those “snapshots” or “clones”

- Data sharing
  - Form of deduplication
  - Data in snapshot shared with live data until one of them is written
  - Two fundamental techniques
    - Copy Out on Write
    - Write to new live location

**Check out the SNIA Dictionary !**

***www.snia.org/dictionary***

# Delta snapshots (cont.)

- Typical uses – readonly
  - Reducing RPO (recovery point objective)
    - typically from a day to an hour or so
    - key feature is ability to revert live system to a snapshot quickly
  - Increasing backup window length
    - reduced demands on backup hardware, backup window schedules
- Typical uses – read/write
  - Quick bringup of duplicate datasets – minutes per TB
  - What-if scenarios
  - Testing of application changes against up-to-date datasets
  - Testing of new applications with near-online data
  - Booting/running of VM images from a golden master

# Thin provisioning

- Similar in concept to filesystem quotas
  - Volume “size” is merely a promise that that amount of storage will become available on demand
  - Storage is not actually allocated in the system until it is used
  - Admin tracks total cumulative use and makes sure that available storage is kept larger than used by some delta.
- Result: no more overprovisioning to avoid running out of space

# RAID 5

- Allows any (one) drive in a RAID set to fail without data loss
- Requires only one extra drive in a RAID set
  - Much less raw capacity required than for mirroring
    - Typical: 8-disk RAID 5 set: 12.5% overhead vs. 50% for mirroring
- Note: RAID 3 and RAID 4 have the same overhead as RAID 5

Note: these numbers would be 14.3% and 100% respectively if figured as overhead *on top of*, as opposed to *as a percentage of*

# RAID 6

- More dependable than mirroring
  - Mirroring: can survive two failures in a disk group if they're not in the same mirrored pair
  - RAID 6: can survive failure of any two drives in the group
- Requires two extra drives per RAID set
  - However, typically somewhat larger RAID sets
- Necessary as drive sizes increase
  - Probability of a disk failure during RAID 5 parity reconstruct is getting too high
- More green than mirroring
  - 50% overhead in RAID 1 mirroring
  - 14.3% overhead in a 14-disk RAID 6 raidse

16.7% and 100% if figured *on top of*

# Deduplication and SIS

- Find duplicates at some level, substitute pointers to a single shared copy
- Block or sub-file based (dedup)
- Content or name based (SIS *, “file folding”)
- Inline (streaming) and post-process techniques
- Savings increase with number of copies found

* SIS = Single Instance Store

**Check out the SNIA Dictionary !**

***www.snia.org/dictionary***

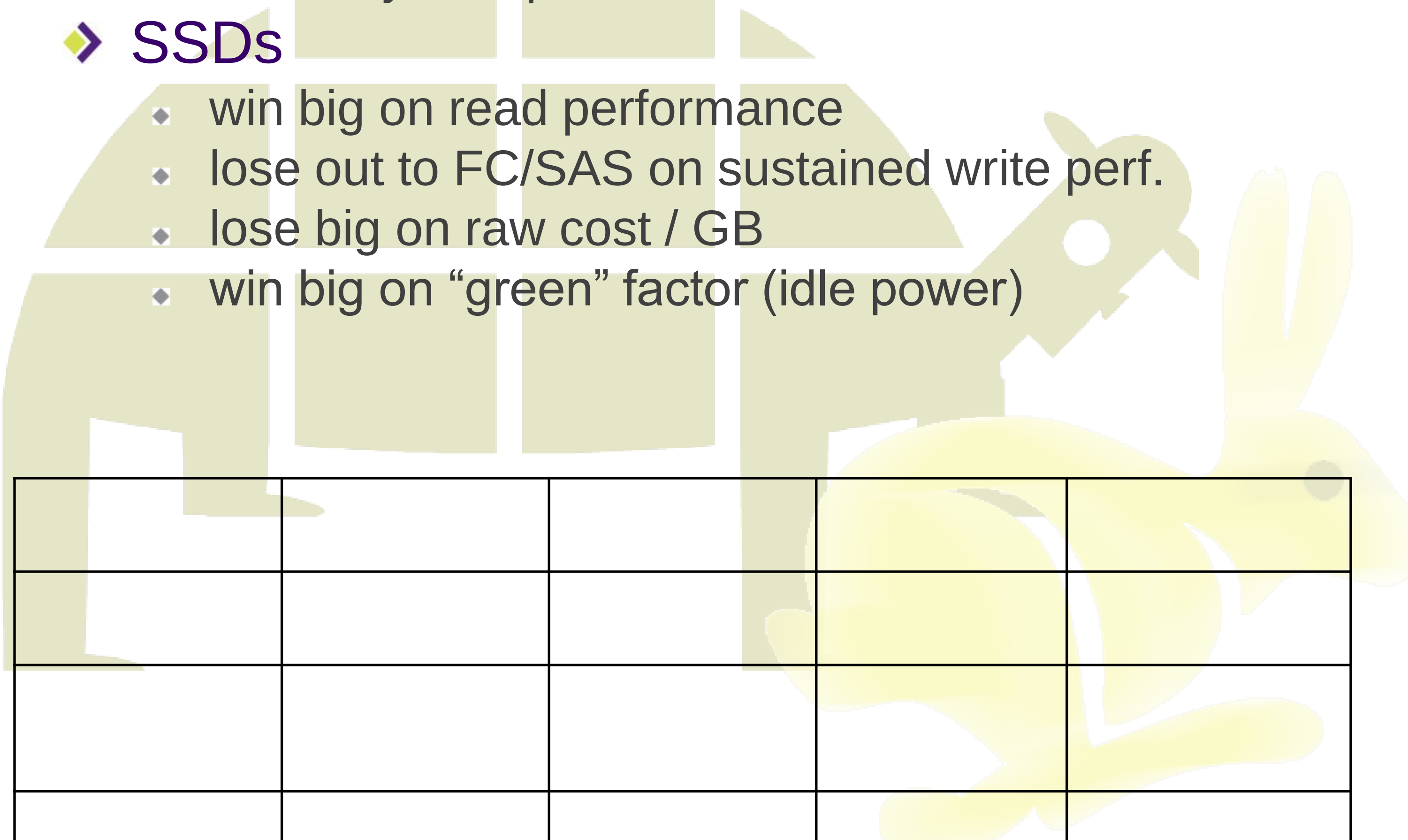
# Other technologies

- Capacity vs. high performance drives
- SSDs
- Flash and stash
- ILM / HSM
- MAID

- Power supply and fan efficiencies

# Capacity vs. High Performance

- Picture rapidly changing due to SSDs
  - Formerly lo/hi perf ←→ SATA/FC
- SSDs
  - win big on read performance
  - lose out to FC/SAS on sustained write perf.
  - lose big on raw cost / GB
  - win big on “green” factor (idle power)



| | | | | |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |

# SSDs (Solid State Disks)

- Usually refers to FLASH-based disks
  - Pros
    - Great READ performance
    - At rest power consumption = 0
    - No access time penalty when idle (cf. MAID)
    - No need to keep some disks spinning (cf. MAID)
  - Cons
    - WRITE performance may be < mechanical disks
    - Cost >> mechanical disks except at very high perf points
    - Wear leveling requires a high space overhead
  - Note: these dynamics changing rapidly with time
    - SSSI – SNIA Solid State Storage Initiative

**Check out the SNIA Tutorials in the Solid State Storage (SSS) Tutorial track**

# “Flash and stash”

- Usually refers to large arrays of SATA-based disks fronted by large flash caches
  - On the order of 1TB flash (or SSDs)
  - Most working sets in flash
  - Reduced power (SATA vs. SAS)
  - Alternative to tiering – well-known caching algorithms

  - Not useful for write-intensive workloads
  - Naive implementations may take a long time to warm the cache after a reboot

**Check out the SNIA Tutorials in the Solid State Storage (SSS) Tutorial track**

# ILM/HSM/Tiered storage

- Exploit cost differences between storage tiers
  - Idea: automatically move data to an appropriate storage platform at each period in its lifetime
- Tier change must have substantial value to make the overhead worth it
  - Cost of system
  - Cost of administration
  - Cost of data movement
- Tiering = “advanced HSM”
  - Colocation of tiers in a single system makes data movement much more efficient
  - May make use of ILM concepts to determine tiering level of data

# MAID (Massive Array of Idle Disks)

- Idea: spin down disks when not in use
  - Pros
    - Disks use no power when spun down
    - > 50% power savings at idle
  - Cons
    - Most data near-online (access times of several seconds)
    - Background disk housekeeping difficult
    - Often the same data center sizing requirements (UPSs, CRAC units, PDUs etc.), but these are used at lower efficiencies
    - Competition from SSDs
    - Competition from....

# Tape

- Oldie but goldie
  - Pros
    - Tapes use no power when inactive
    - > 90% power savings at idle
  - Cons
    - Data is at best near-online (access times of several seconds)
    - Not a random access format
    - Lack of true resilience to format failure
      - – Redundant Array of Independent Tape? (RAIT)

**Check out SNIA Tutorial:**

**Introduction to Data Protection: Backup to Tape, Disk and Beyond**

# Power supply and fan efficiencies

- Efficiency of power supply an up front waste
  - Formerly 60-70%
  - Nowadays 80-95%
    - Climate Savers
    - 80plus group (see http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx)
    - Note: Efficient PSs *are* more expensive
- Variable speed fans
  - Common nowadays
  - Software (OS) control

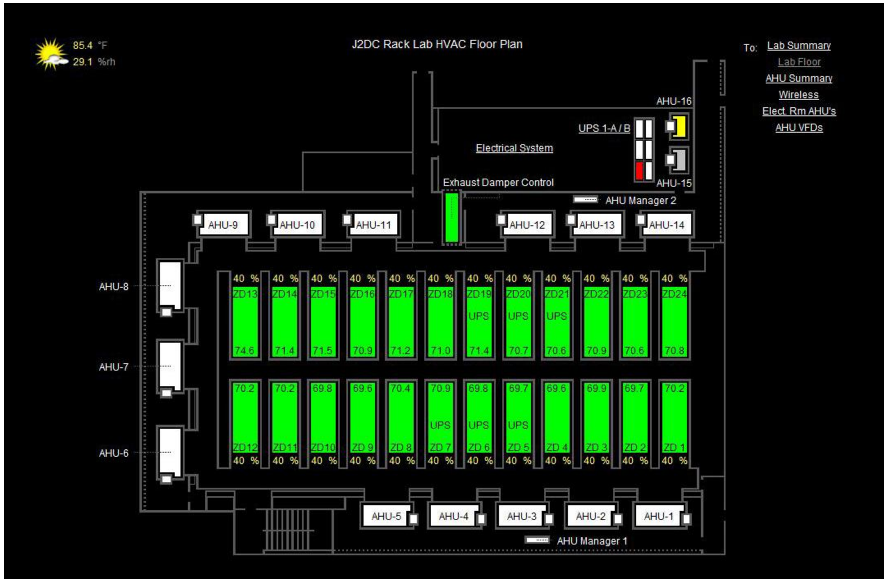
# Facilities-side technologies

- Monitoring
- Hot aisle – cold aisle technologies
- Other
  - Spam filtering
  - Water and natural cooling
  - Flywheel UPSs

    (UPS = Uninterruptible
    Power Suppy)

# Monitoring

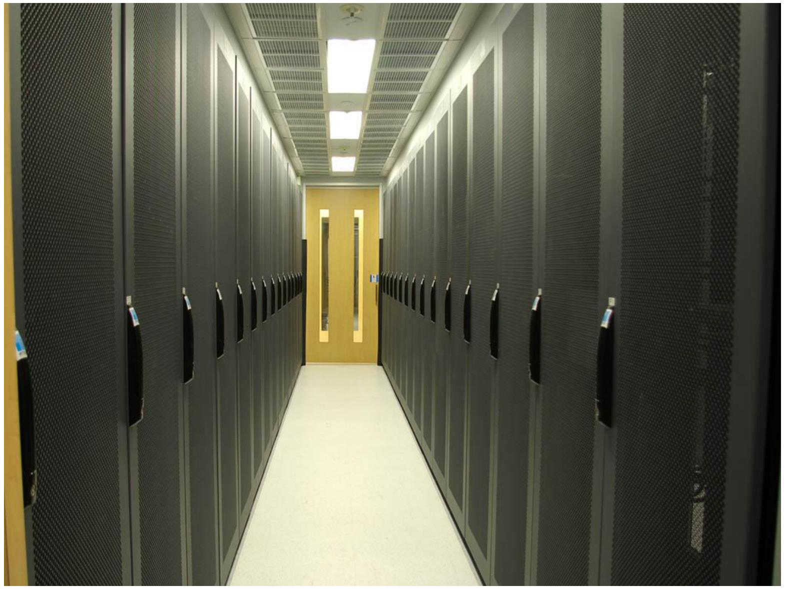
- Critical to increased efficiencies
  - Lights out operation
  - Tightening up of temperature tolerances
  - Better staff utilization
  - Anomaly detection



# Hot aisle / cold aisle technologies

Education SNIA

- Segregate airflows into hot and/or cold aisles (backs and fronts of servers)
  - More precise control
  - Allows higher temperature differentials (more efficient)
  - Several emerging approaches
    - Hot air plenum
    - Complete containment
  - Current trend toward hot aisle containment with cold air plenum
  - Must-have: blanking plates
    - Very important



# Agenda

- Overview, Definitions, and what does green storage mean
- SNIA storage green activities
  - Green Storage Initiative
  - Green Storage Technical Working Group
  - SNIA Emerald$^{TM}$ Program
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Introduce metrics
- Storage taxonomy
- How storage vendors use the Emerald$^{TM}$ Program
  - SNIA Emerald$^{TM}$ Power Efficiency Measurement Specification
  - Best foot forward (sweet spot)
  - Exercise showing best foot forward
- Other Associations Green Storage Efforts
- Storage technologies for energy savings
- Typical savings

# Savings calculations

- Space savings
- Equipment power savings
- Facilities power savings

# Savings calculations (storage)

- Calculations herein are for space savings
- Relationship of space to $$ is loose
- But every TB of disks you don’t buy saves you
  - CapEx (Capital Expenditure)
    - for the equipment
    - for the footprint
    - for plant, power conditioning and cooling
  - OpEx (Operational Expenditure)
    - for equipment power
    - for power conditioning and cooling
    - for storage management
    - for service contract fees

# Typical space savings

- Compression
  - 15 – 40% savings
  - Remember, no savings from already compressed file formats
- Readonly delta snapshots
  - 90 – 99% savings per snapshot, compared to full PIT copies
- Writeable delta snapshots
  - 80 – 99% savings per snapshot, compared to full PIT copies

# Typical space savings (cont.)

- Thin provisioning
  - 20 - 50% savings
  - Average 30% utilization ➔ over 80% utilization
- RAID
  - 35% savings for e.g. 14-disk RAID 6 set, compared to RAID 1/10
  - Note: use the right RAID level for your application
- Deduplication
  - primary and archive storage: ~35% savings
  - backup: 35% - 95% savings, depending on dataset, time interval, and ratio of incremental to full backups

# Caveats

- Savings estimates are real, but best taken as anecdotal
  - YMMV – your mileage may vary
  - Make your vendors prove their claims in your application environment
- Green options in some implementations may degrade RAS and/or performance
  - E.g.
    - RAID 5 vs RAID 10 – reduced RAS
    - 7.5K SATA vs 15K FC – reduced performance
    - Tape vs disk during legal discovery – large fine if too slow
  - Make your vendor tell you the cons that go with the pros

# Caveats (cont.)

- Savings estimations are always imprecise
- Obviously can’t achieve infinite savings by applying all technologies
  - Diminishing returns

# Overall savings

- State of the art data centers
  - PUE* drops from 2.25 to 1.25 = 45% savings
    - 10MW → 5.5MW
    - $6.0M → $3.3M annually
  - Rebates in the $M from utilities on top of savings
- Reduced equipment footprint per TB data
  - Vendors claiming up to 50% reductions
  - *Require proof or guarantees*

* Power Utilization Efficiency – see [www.thegreengrid.org/gg_content](www.thegreengrid.org/gg_content)

# Conclusions

- It's a brave new world
- Most vendors shipping most of these technologies
- Ask your vendor to justify use of the technology against your business problems
- Use power vendor power calculators or SNIA Emerald$^{TM}$ test data
- The advantage is yours: seize the day!

# SNIA and Storage Industry Continued Green Efforts

- Operational metrics (show storage power performance in the datacenter)
- In-depth training
  - Tester and Auditor
- SNIA Emerald™ V1.X measurement specification
  - File systems
  - New workload for large cache systems
- Ongoing collaboration with EPA
- Ongoing collaboration with Green Grid and other green organizations
- More testing

# References

- SNIA Green Storage Initiative – www.snia.org
- SNIA Emerald - www.sniaemerald.com
- The Green Grid – www.thegreengrid.org
- EPA ENERGY STAR™ (Data Center Storage – www.energystar.gov/index.cfm?c=new_specs.enterprise_storage
- SPC – www.storageperformance.org
- SPEC – www.spec.org (SPECpower_ssj2008)
- 80+ - www.plugloadsolutions.com/80PlusPowerSupplies.aspx
- Power calculators at various vendor sites

# Back Up Slides

# Specification Requirements

- Test setup requirements
  - Input voltage – standard world voltages
  - Environmental – standard datacenter conditions
  - Benchmark driver – VdBench, IOMeter, required conditions
  - Meters
    - Power – accuracy level
    - Temperature- accuracy level
- Equations
  - Average response time and power
  - Periodic power efficiency
  - Metric stability – 10 point rolling average
  - Time interval of 1 minute or a specified measurement interval