Enterprise Storage – Storage Efficiency

Leah Schoeb, Vice Chair, SNIA Technical Council

SNIA Emerald™ Training

SNIA Emerald Power Efficiency Measurement Specification, for use in EPA ENERGY STAR®

June 24-27, 2013

Topics

- Optimization
  - Challenges
  - Drivers and Initiatives
  - Economics
  - Optimization Initiatives to Meet Demands
  - Storage Perspective – Trends and Dynamics
- Increasing Storage Efficiency
  - Capacity efficiency
  - Solid State Technology

Topics Not Covered

- Data Protection and Management
  - http://www.snia.org/education/tutorials/2012/spring/data
- Cloud Archive
  - http://www.snia.org/cloud/archive

Challenges in IT

- **Meeting capacity demand**
  - Capital expenditures for systems
  - Space, power, and cooling issues
  - Infrastructure expansion
  - Data protection – expanding on current practices
  - Administration
- **Budgets**
  - Increases not parallel with demand
  - Expected services greater than budget allowance

Unstructured Data – 90%
Structured Data - 10%

IT Challenges

SNIA™ Green Storage Initiative

23% do no capacity planning

85% have no show back capabilities

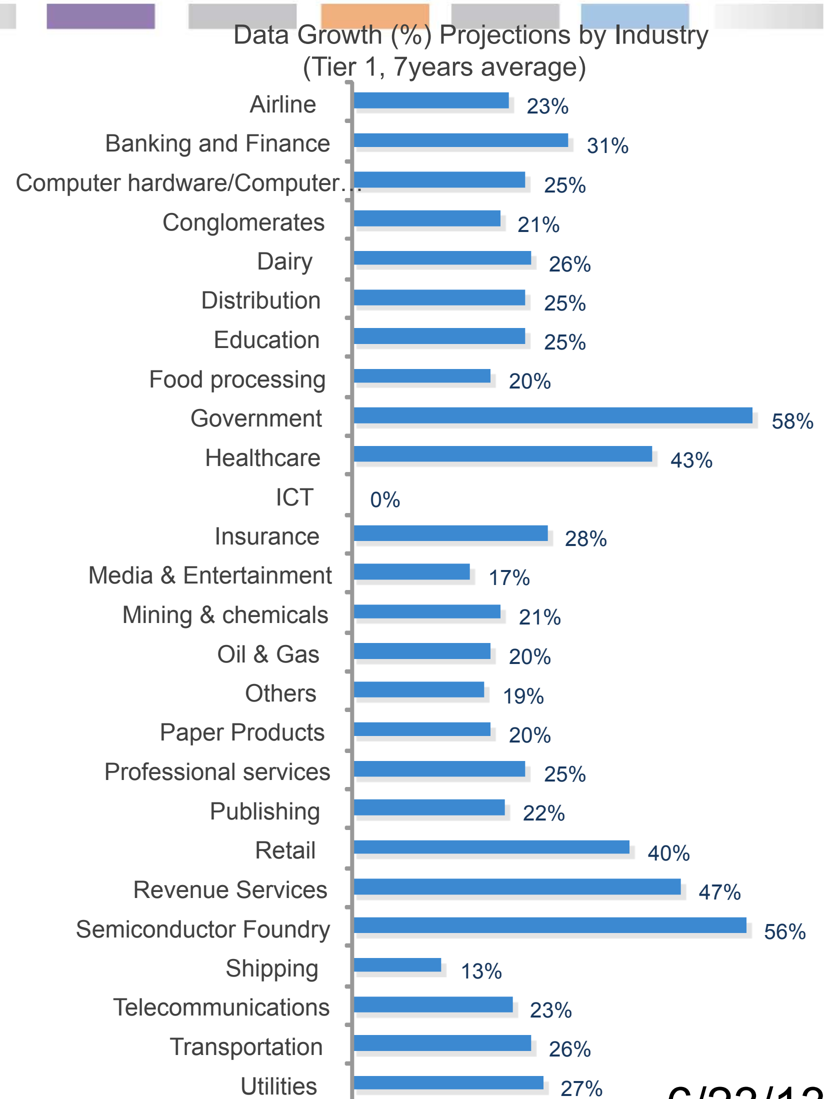
70% do not use a storage catalog

90% have little to no process documentation

70% do not use storage virtualization

70% do not have archive processes

68% are not using de-duplication



Source:

- IBM Storage Infrastructure Optimization (SIO) study

Challenges in IT

- External inputs: "Why aren't you doing this?"
  - Cloud implementations
    - Public clouds – buzzwords and confusion
    - Private cloud – data center transformation to IT As A Service
    - Software As A Service
    - Big data analytics
- Mobility
  - Continued use expansion of tablets and notebooks
    - BYOD (A change in the "B")
  - Security concerns for information
  - Administrative efforts – where is the line

Top CIO Priorities for 2012

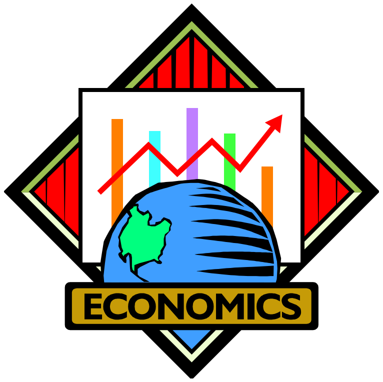
SNIA Green Storage Initiative

1. **Consolidation / Optimization**
   - Server, Desktop, Storage, Services
2. **Cloud Services for IT – IT as a Service**
3. **Security**
4. Mobile services / mobility
5. **Budget and cost control**
6. Shared services – charge back and utility-based services
7. Health care - insurance issues
8. Legacy modernization
9. Interoperable networks
10. **Disaster recovery / Business continuity**

US government NASCIO survey 11/13/12

Drivers in Optimization and Efficiency

- Expansion of business / organization
  - Greater needs for capacity, availability
  - Rapid deployment requirements
  - New demands in server virtualization and VDI
- Return on investment into IT – compared to alternatives
  - Optimize the IT investment
  - IT as value - not a tax
- Continued advancement…
  - Remain competitive
  - Adapting to changing demands
  - ***"Technology as a competitive advantage"***



Optimization - Economics

SNIA™
Green Storage Initiative

Practical Usage

- Costs – assessed, profit center, fixe
  - Facilities
  - Administrative staff
  - Equipment
  - External / contract / consultant
- Use of metrics
  - Measure for improvement
  - Comparison against industry – defend costs
  - Requires a discipline to collect and calculate

Optimization – Efficiency

- Utilization of resources
  - Effectively using what was paid for
  - Examples:
    - Using processor cycles in a server
      - Limitations preventing effective usage
      - Effort to improve
    - Storage space
      - What is the utilization
      - Limitations – inherent architectural considerations
      - Technology changes that have an impact
- Green storage technologies use less raw capacity to store and use the same data set
  - Power consumption falls accordingly

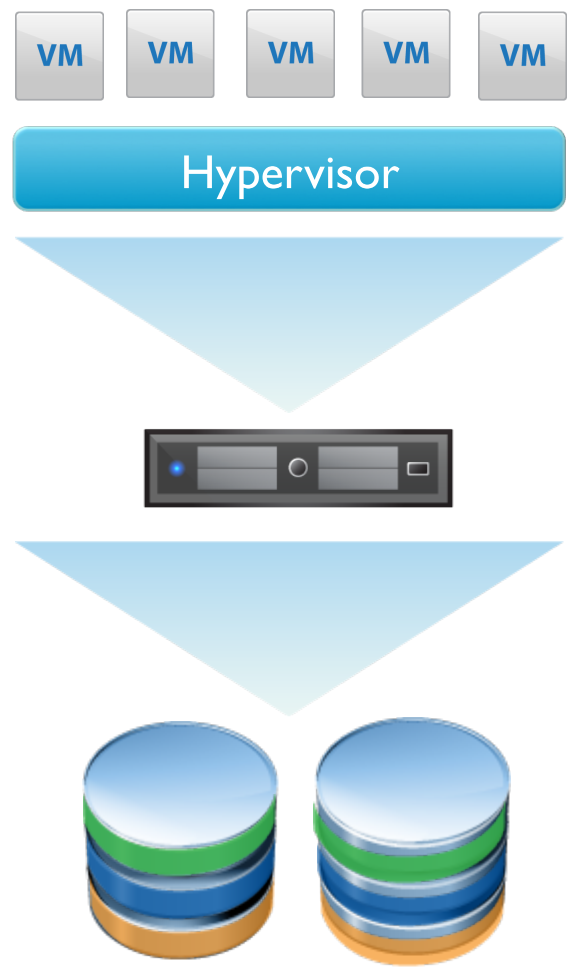
Optimization – Efficiency

- Delivery of services
  - Measurement is time: how long
  - Examples:
    - Time for application deployment: software, server, storage, physical space, power, cooling, infrastructure
    - Time for connectivity: network, userids,
    - Time for storage – expansion: how long, movement required, protection
    - Time for more compute cycles

Storage Perspective – Trends & Dynamics

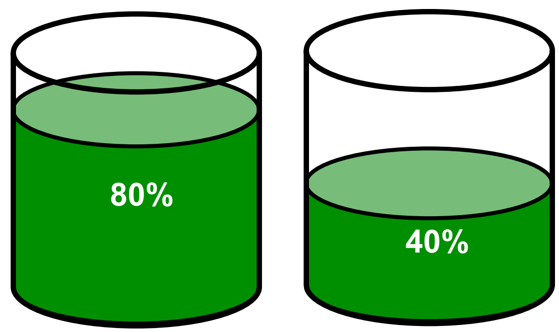
SNIA Green Storage Initiative

- **Virtualization (Server & VDI)**
  - Reduce number of physical servers
  - Greater use of resources
  - Security driven
- **Storage efficiency**
  - Consolidation
  - New features / technologies
    - Tiering
    - SSD's
    - Thin provisioning & reclamation
    - Data reduction
- **Future - Software Defined Storage**
  - Streamline storage consumption
  - Monitoring, management, and metering
  - Protection framework for highly scalable storage services



Storage Efficiency

Optimized Capacity for a Greener Datacener



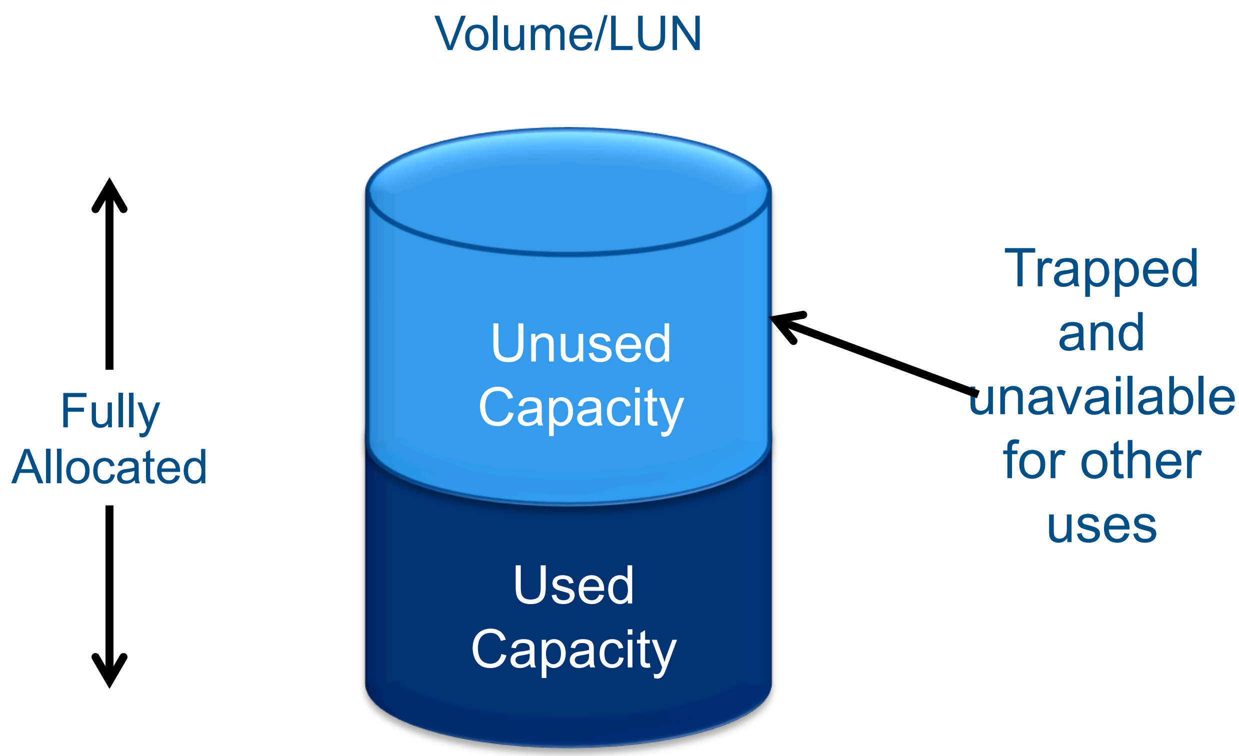
Enterprise Storage Optimizing Features

SNIA Green Storage Initiative

- Thin Provisioning
  - Use of disk virtualization
- Data Reduction
  - Deduplication
  - Compression
- Using Solid State Storage
- Auto Tiering using solid state storage

Thin Provisioning

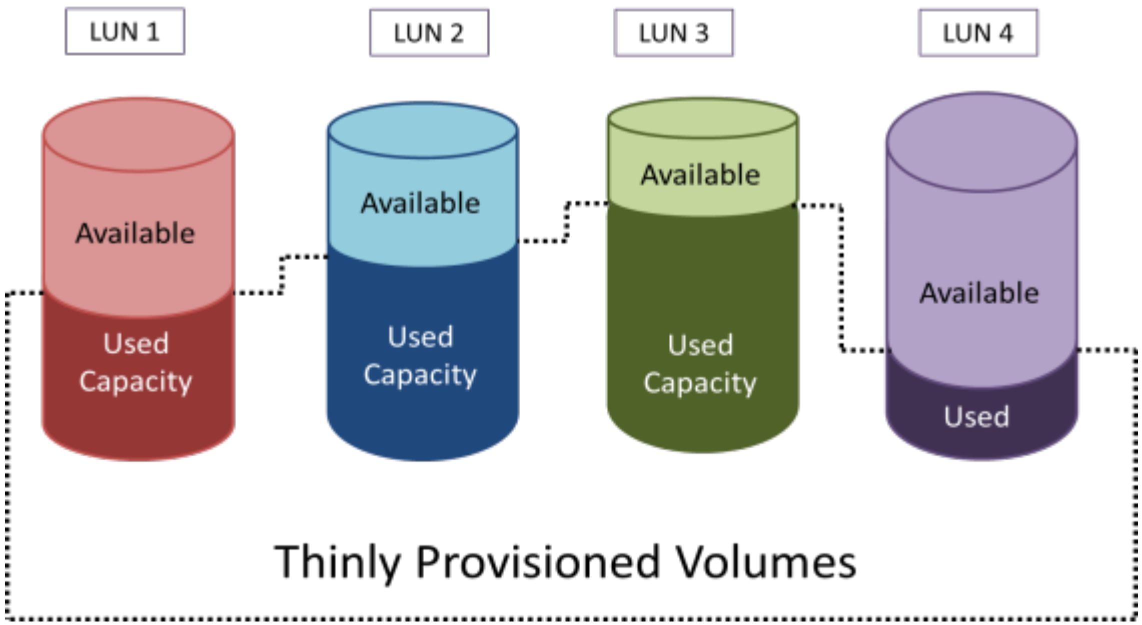
Before Thin Provisioning



Traditional Allocation

Thin Provisioning

- Allocate space when data is written
  - PG&E adds thin provisioning to their energy-saving incentive program
  - More effective use of capacity and space efficiency
  - Cuts down or eliminates over provisioning
  - Notifications when space consumption reaches thresholds
  - Space Reclamation



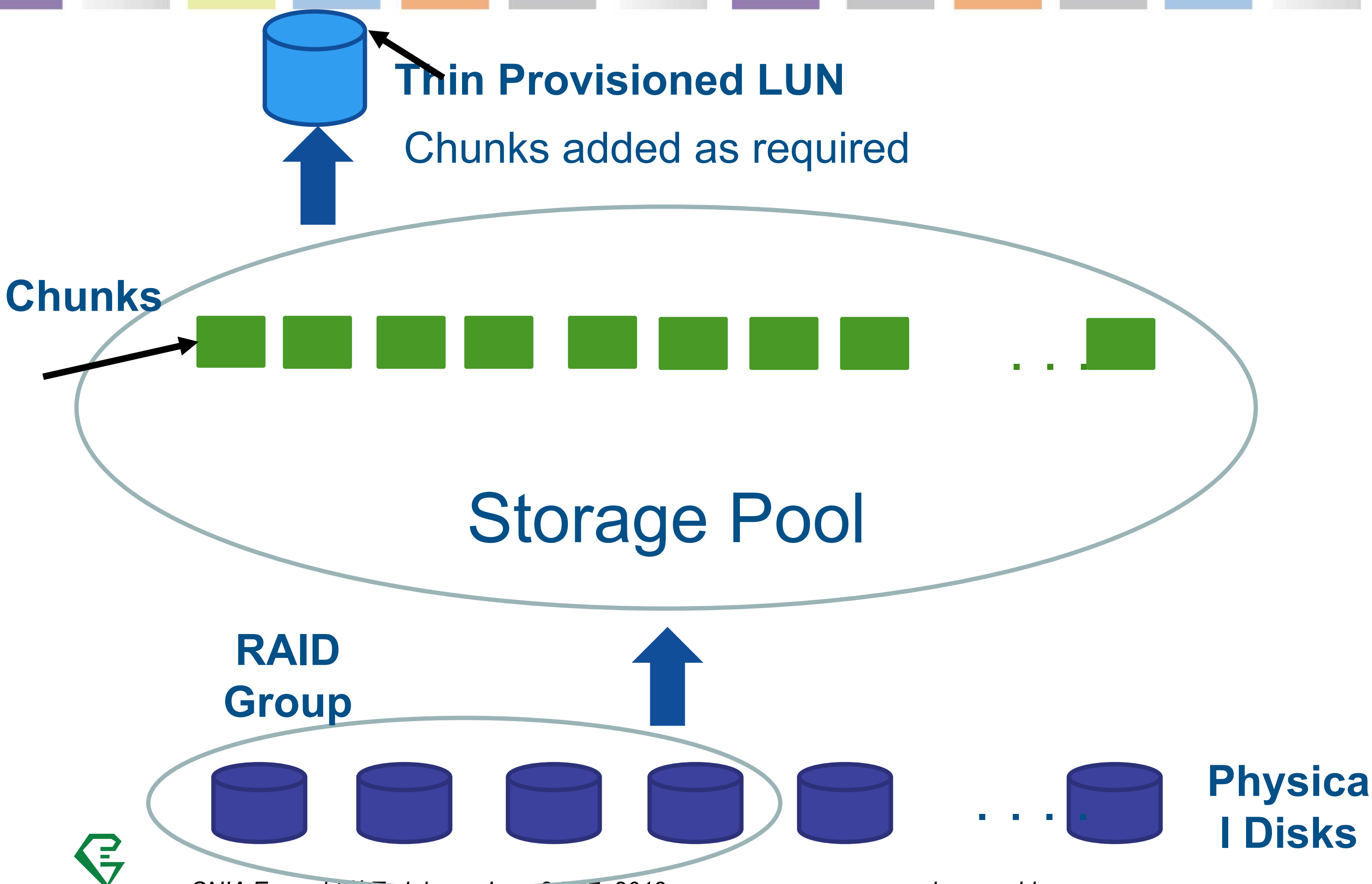
- Can save up to 75% in energy costs
- Can eliminate up to 70% of legacy capacity

Disk Virtualization for Thin Provisioning

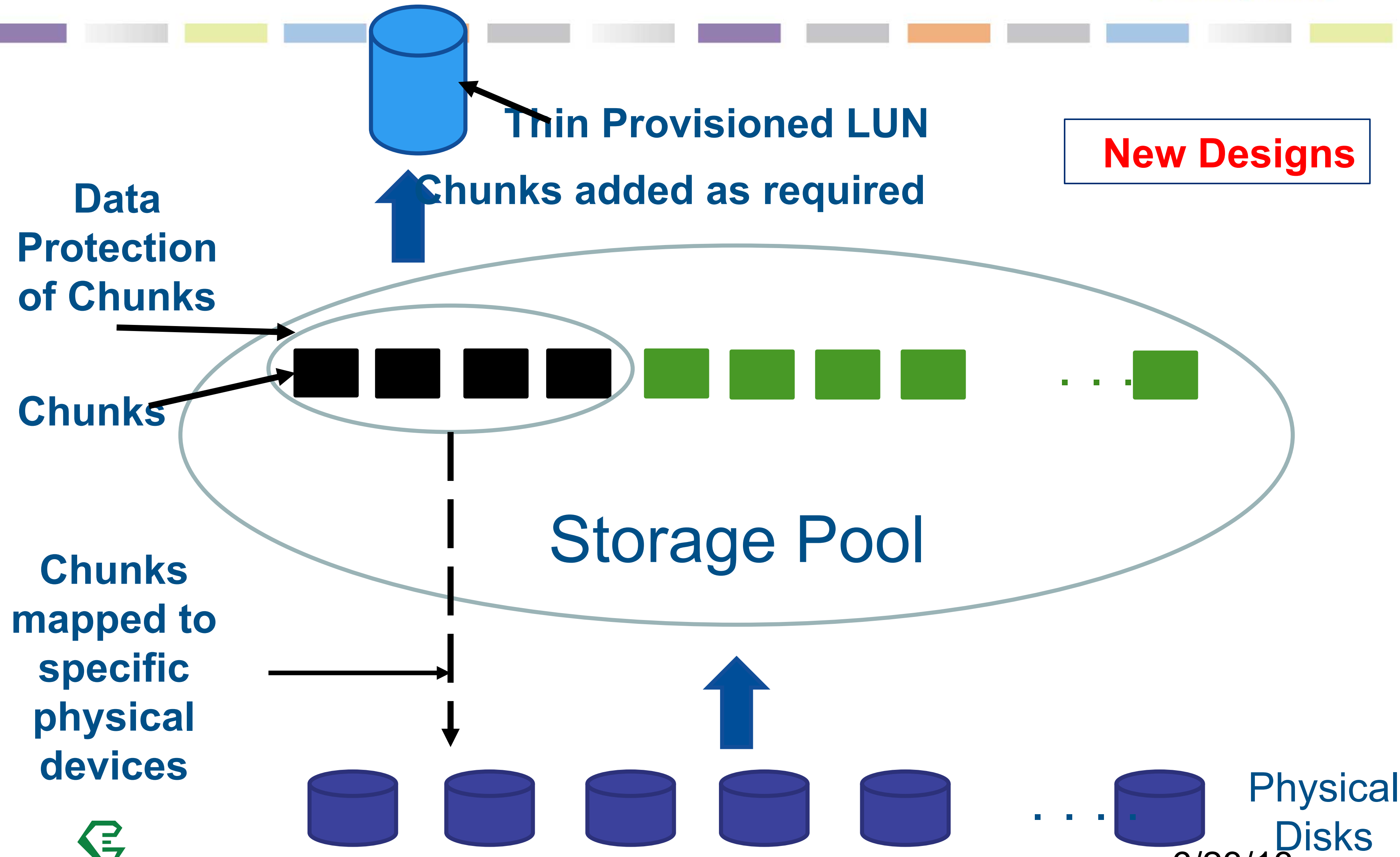
SNIA
Green Storage Initiative

- New storage system architecture to only assign capacity as it was written
- Created storage pools
  - Set of chunks of data – size of chunk could be specified at time of configuration
  - Volumes were created without assigning real capacity
  - As data written, chunks were assigned
- For newer architecture, chunks were distributed across physical disks to match data protection choices
  - Different RAID levels
  - Advanced data protection techniques

Thin Provisioning from Storage Pools



Thin Provisioning from Storage Pools with Data Protection

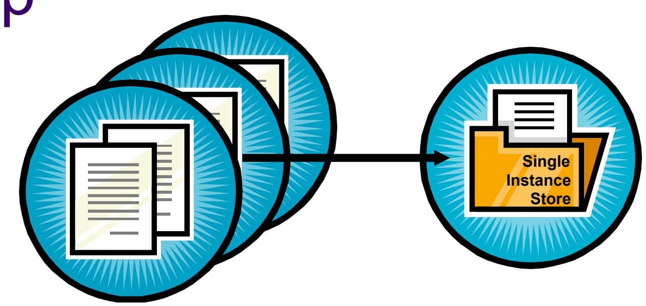


Data Reduction Technology

Deduplication and Compression

What is Deduplication?

- Top-level definition: storing only a single instance of data
  - **At block level**: look for redundant blocks
  - **At byte offset level** (also called bit-level sometimes): look for stream of bytes that are duplicate
  - **At a file level**: look for duplicate files
- Redundant data is replaced with a pointer (index) to the unique copy
- Streaming (in-band) and offline techniques
- Makes a greener disk-to-disk backup
- Up 90% of data capacity saved
- Different than compression
- Store more data in less space



What is Compression?

SNIA
Green Storage Initiative

- Involves encoding information using fewer bits that the original representation.
- **Lossless Compression** – reduces bits by identifying and eliminating redundancy.
  - represent data more concisely without losing data
- **Lossy Compression** - reduces bits by identifying unnecessary information and removing it
- Decompression comes at a performance **cost** which varies
  - Degree of compression
  - Amount of distortion introduced when using lossy compression
  - Computational resources required to compress and uncompress data

Data Reduction Technologies

- Reduce amount of data
  - Data compression
  - Data deduplication
- Backup tapes – compression
- Backup to disk – typically deduplication
- Backup to cloud – compression
- Primary or secondary storage – deduplication / compression
- Archive – compression

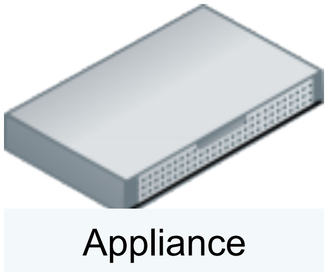
Green Storage technology savings calculation

- Every TB of disks you don't buy saves you
  - CAPEX for equipment and footprint

Space Reduction Ratio	Space Reduction Percentage = 1 – (1 / Space Reduction Ratio)
2:1	1/2 = 50%
5:1	4/5 = 80%
10:1	9/10 = 90%
20:1	19/20 = 95%
100:1	99/100 = 99%
500:1	499/500 = 99.8%

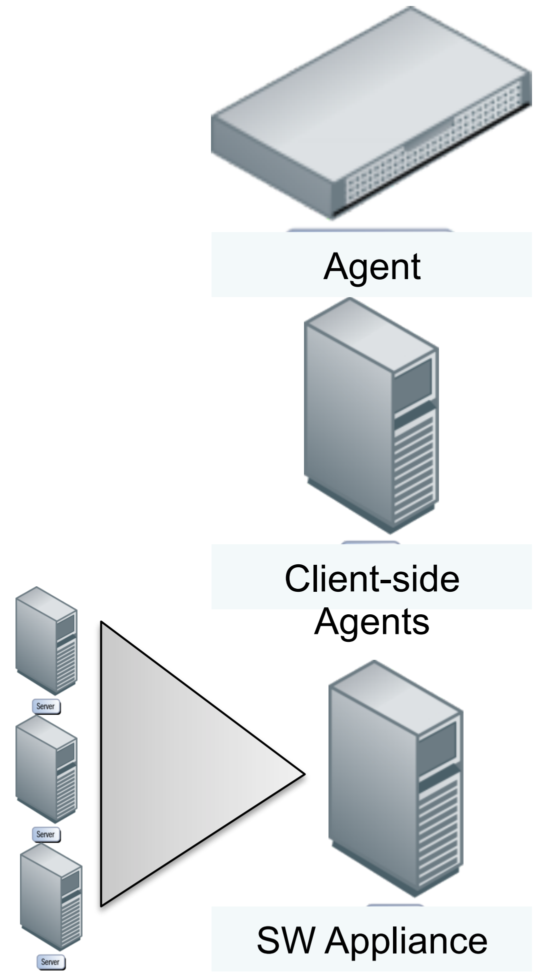
3 Major HW Design Approaches for Datacenter

- Storage System
  - General purpose storage system that can dedup data
- Gateway
  - Hardware capable of dedup data on external storage hardware
- Appliance
  - Hardware capable of dedup and storing data in an appliance for this specific purpose



3 Major SW Design Approaches

- **Appliance**
  - HW for dedup and storing data in an appliance for this specific purpose
- **Agents**
  - Client-side SW for a specific application
- **Software Appliance**
  - Host-side SW for storing dedup data from application agents



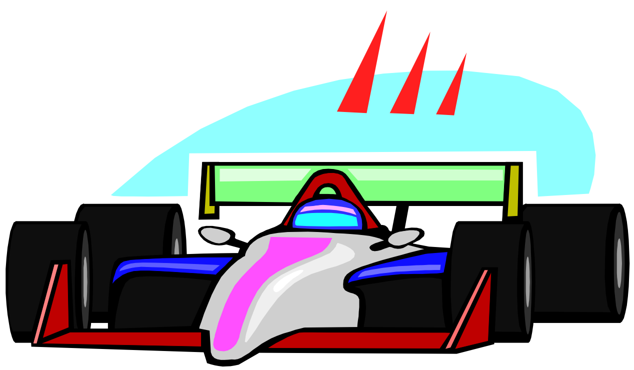
Benefits – Deduplication and Compression

- Satisfy ROI/TCO requirements
- Manage data growth
- Increase efficiency of storage and backup
- Reduce overall cost of storage
- Reduce network bandwidth
- Reduce operational cost including
  - Infrastructure costs required space, power and cooling
  - Movement toward a greener datacenter – carbon footprint friendly
- Reduce administrative costs

Deduplication Summary

- Reduces physical data storage requirements by removing duplicate data
- Allows data to be retained on disk for longer periods, using less physical space
- Involves tradeoffs relating to degree of physical space reduction and performance considerations (some designs are better than others)
- Can be used with primary storage or as part of a backup solution.
- Thoughtful deployment allows users to benefit from a greener datacenter.

Solid State Technology

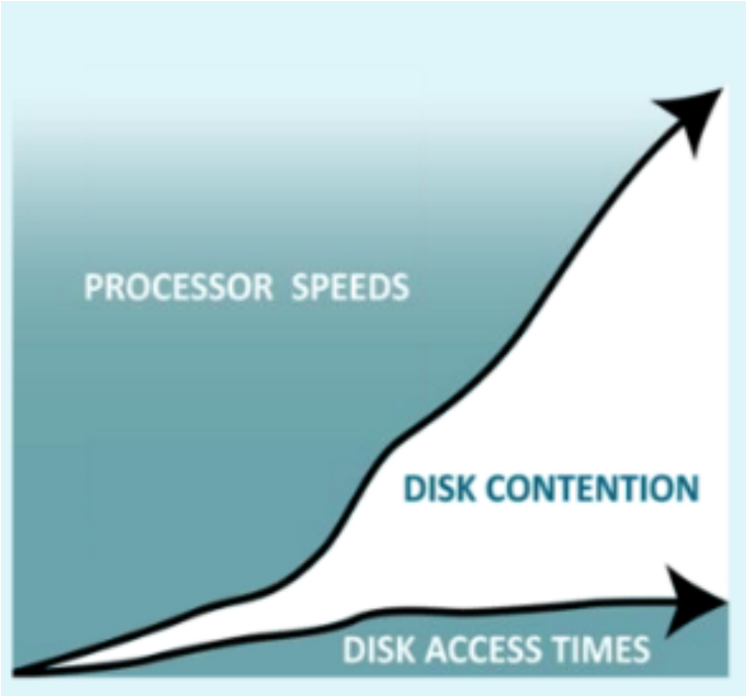


Solid State Device Usage

SNIA Green Storage Initiative

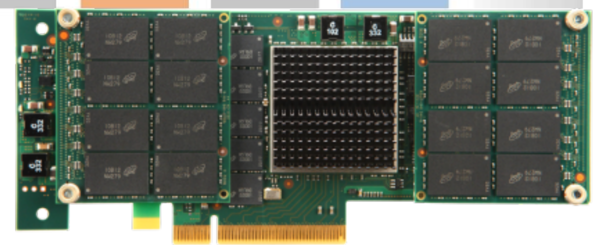
Justifications for SSDs

- Rugged – no moving parts
- Performance – no rotational latency or actuator movement as with HDDs
- Reliability: MTBF – typically 6x over HDD
- Operational costs – power requirement is typically 1/4th of HDD



Solid State Technology

- Performance is primary gain with solid state
- Focus on NAND Flash today
- General Usages
  - SSD: Solid-state device,   Solid-state disk or drive
  - Flash memory cards for servers
  - Tiering in storage systems
  - Caching in storage systems
  - All solid state storage systems
  - Consumer

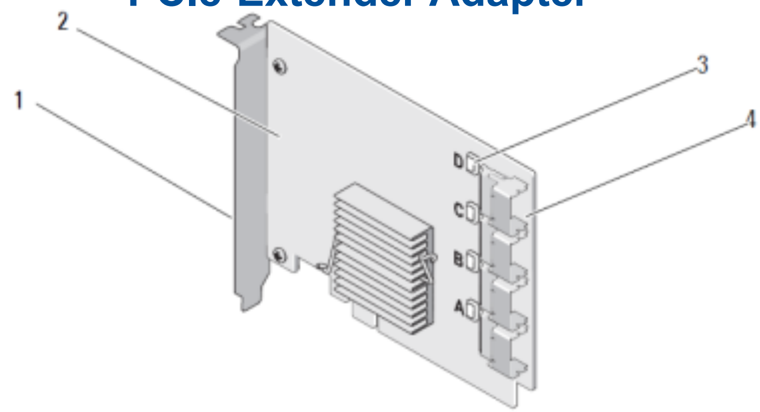


PCIe Flash Cards

NAND Flash Drives

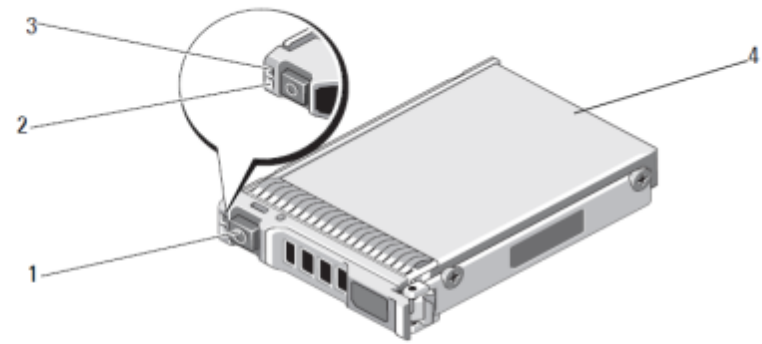
Solid State Technology

PCIe-Extender Adapter



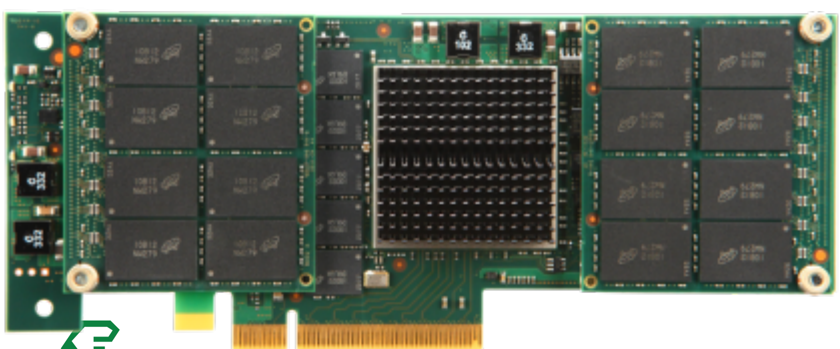
1	filler bracket	2	PCIe extender adapter card
3	port link status LED(4)	4	adapter cable connectors (4)

PCIe-SSD



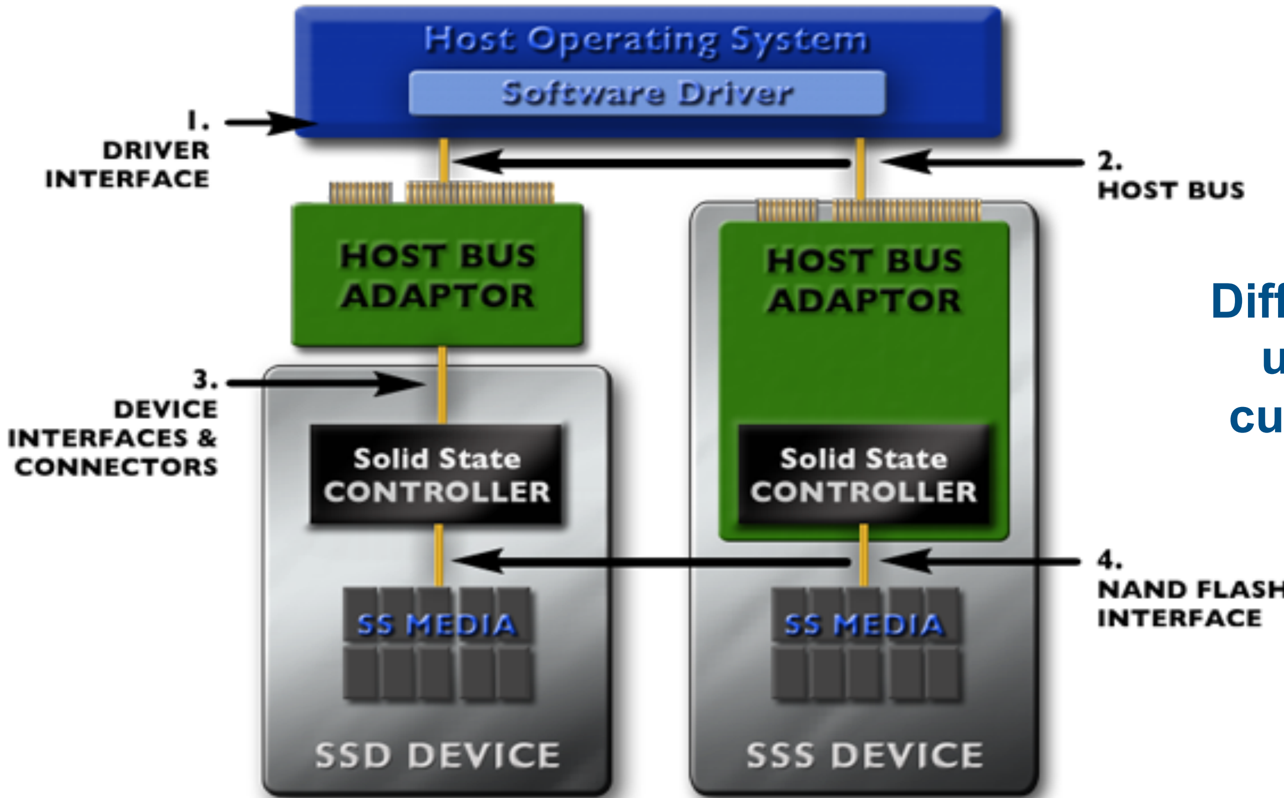
1	release button	2	activity indicator
3	status indicator	4	PCIe SSD device

Flash Card

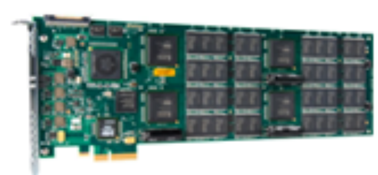


SAS, SATA-SSD

Solid State Technology Host Side Storage



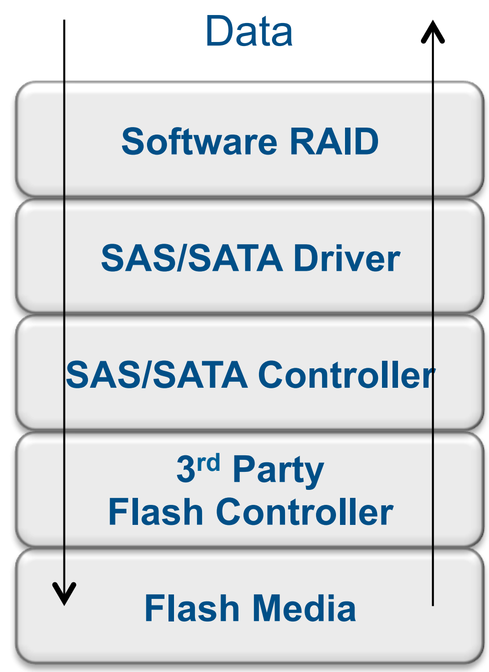
Difference between use of SSD's and custom solid state designs



Solid State Technology Storage Arrays

- Solid state in existing design storage system
  - SSDs as a storage tier
    - System manages data and accelerates operations
    - Embedded software to migrate data based on access patterns
  - SSDs as cache tier
    - Highly active data keep in SSDs
    - May be separate read and write
    - Value over DRAM is cost – non-volatility simplicity
  - All-SSD array (existing design with only SSDs)
    - Fundamental architecture still based on spinning disks
    - Configuration difference only
  - Storage system use of internal Flash cards
    - Typically as cache

SAS/SATA based SSD



Solid State Technology Storage – New All Flash Arrays

SNIA Green Storage Initiative

- New storage system architectures for all solid state
  - Internal design to handle memory speed
  - Manage technology characteristics – writes, cell sparing
    - Done with custom Flash controllers
  - Different types – SSD, Flash card, next generation
  - Key characteristics to look for:
    - ***Data reduction – reduces number of writes required***
    - ***Snapshot as a mapped technology***
    - ***Remote replication***
    - ***Self-healing techniques***
    - ***Lifespan***
    - ***Scale-out capability – usually clustering***

Efficiencies w/ Solid State Storage

SNIA™
Green Storage Initiative

- SSDs do not require special cooling and can tolerate higher temperatures than HDDs.
- Resistant to shock and vibration
- Small and light weight leads to higher capacity density
- Trend – data reduction with SSDs
  - Compression & deduplication to ***multiply capacity and reduce number of writes required***
  - New techniques will drive towards price parity with HDDs

Factors to Consider in SSS

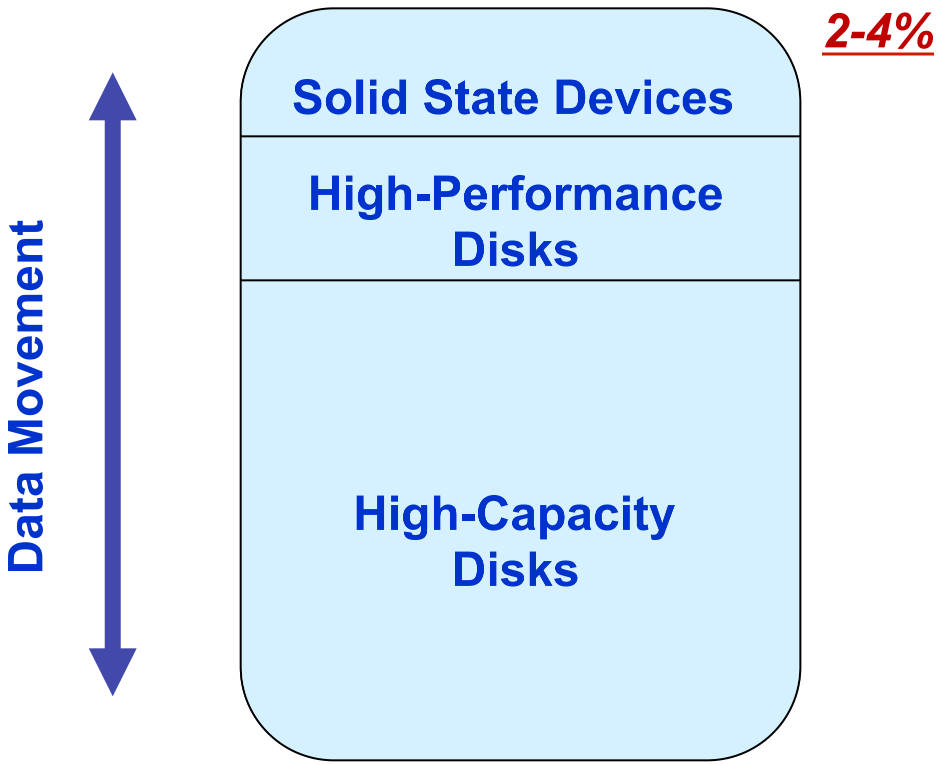
SNIA™ Green Storage Initiative

- Unique Performance behavior
  - No moving parts (no rotational latency or seek times)
  - Access time in the microseconds
- Wear Leveling Algorithms
  - Not all algorithms are created equal
- Misalignment
  - Performance degradations as high as 30% in some cases
  - Shortened SSS lifespan
- SLC vs. MLC Flash
  - Single-Level Cell Flash stores one unit of information in a single cell.
  - Multi-Level Cell Flash stores multiple units of information in a single cell.
  - SLC access times are faster than MLC
  - MLC is less expensive
  - TLC is coming on the market – Good for archive solutions

Tiered Storage

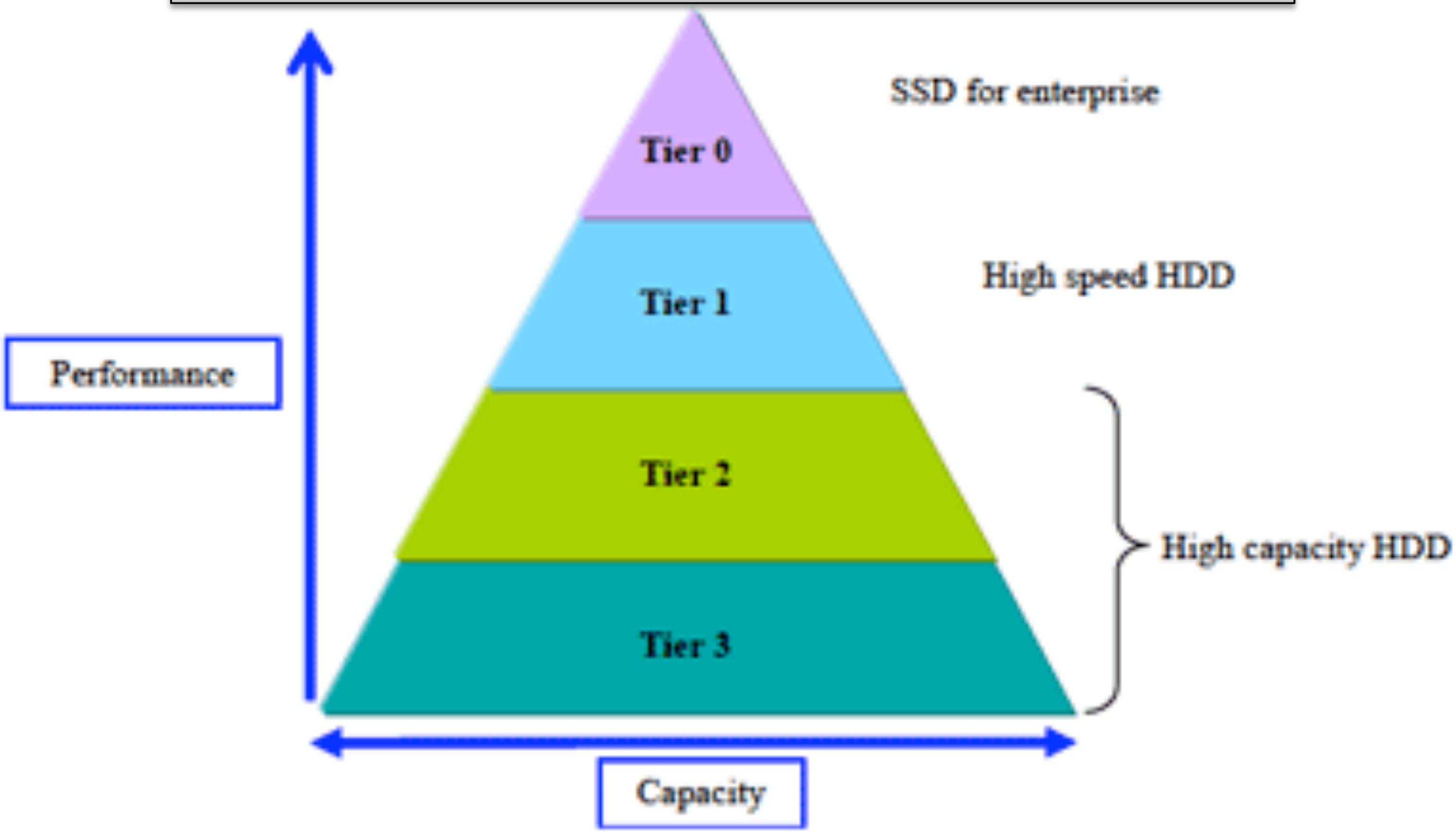
Storage System Tiering

- In-the-box tiering
  - Major performance boost
  - Limit economic impact of costs of higher performance drives
  - Utilizes resources better
  - Economic advantage



Tiered Storage & Auto Tiering

Solid State Technology can work well in Tiered Storage Designs



Tiering vs. Caching

- Storage tiers
  - Persistence
  - Implies a particular price and performance metric
  - Provides actual capacity
  - All content resides on media
  - Performance is limited only to media speed
  - May be improved with caching
  - Major gain: storage consolidation

Actual location for data

- Caching
  - Not considered actual storage
  - Not persistent
  - All capacity must be backed by non-volatile media
  - Performance limited to size of cache
  - Limited use for random workloads

Data is transient

Gains from Optimization and Efficiency

- Greater storage efficiency - $
  - Space, power/cooling, utilization
  - Direct economic gains
- Consolidation
  - Lower datacenter footprint
  - Higher density
  - More effective use of current storage
- Provable optimization savings
  - Results – TCO and ROI
  - Initiatives – show improvements in incremental steps

SNIA
Green Storage Initiative

Thank You

Leah Schoeb

leah@evaluatorgroup.com

Twitter:@vLeahSchoeb