Education

Green Storage Technologies, CAPEX and OPEX

Alan G. Yoder, Ph.D.
NetApp

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 Technologies, CAPEX and OPEX

- The best green storage technologies don't just affect power consumption--they can also significantly reduce your acquisition costs. This tutorial takes a dollar-and-cents view of the green technology and process areas, helping you prioritize your acquisition planning and attention. We cover green storage technologies, the importance of efficient facilities, and recent developments in ENERGY STAR regulation of the storage sector..

Learning objectives

- Learn how capacity optimization and storage efficiency can help you conserve IT budget and reduce your power footprint.
- Learn how efficient facilities can save your company significant amounts of money
- Find out what the EPA is up to. Is there an ENERGY STAR program for storage?

Outline

- What does "green" mean to data center operators?
- Where does all the energy go?
- Storage and energy
- The big rocks
- Other topics

What does "green storage" mean to data center operators?

- It means reducing the data center footprint of storage
  - in space
    - less equipment to put in place
  - in energy
    - more energy-efficient equipment
    - less equipment to power and cool
    - more efficient cooling and power management
  - in administrative costs
    - less equipment to buy and manage
- In other words, MONEY
  - Less CAPEX for IT (and maybe facilities too)
  - Less OPEX for IT (administrative costs go down)
  - Less OPEX for Facilities (energy cost savings, but more admin)

How much is “green” worth?

- More than you think!
- Facilities: over ½ total energy cost, sometimes
  - MW* data center ~= $100/hr = $876K/yr
  - Taking PUE** from 2.5 to 1.25 saves ½ of energy bill
- Servers: over ½ total energy cost, sometimes
  - Anecdotal increases in efficiency of 3:1
- Storage: over ½ total energy cost, sometimes
  - Storage software optimizations are key

* MW megaWatt
** PUE defined later

Where does all the energy go?

Power to Data Center

UPS / Power Conditioning

Power losses

Cooling

"Useful Work"

IT Equipment

(Only major power loss sources are shown)

What do we mean by IT equipment?

Power to
IT Equipment

IT Equipment

Servers (about 60%)

Networking (about 10--20%)

Storage (about 20--30%)

YMMV!

PUE

- "Power Usage Effectiveness"

$$\frac{\text{Power to Data Center}}{\text{Power to IT Equipment}}$$

- Historically 2.25 to 3.0 and even higher
  - Because IT work transforms power into heat, which must be eliminated, and air conditioners are not 100% efficient...
  - 2.0 means 50% of the power is being lost or used for cooling
- Modern best practice asymptote is around 1.25
  - 80% of the power gets delivered to IT equipment

"Gaming" PUE

- Results below 1.2 are suspect
  - 1.0 implies no distribution or cooling cost – perpetual motion!
- PUE *sounds* intuitive
  - All power losses upstream of the equipment wall plug drive the number up from 1.0
  - Easier to violate than one would first think
- Typical problem – using IT equipment for cooling
  - e.g. equipment fans used to exhaust an enclosed hot aisle

Putting numbers on the diagrams

Power to Data Center

2792kW

UPS / Power Conditioning (85% efficiency) 364kW

Typical Historical Figures

1000kW

"Useful Work"

Cooling (at 70% eff.) 1428kW

Servers – 600kW
Networking – 100kW
Storage – 300kW

PUE = 2.8

Putting numbers on the diagrams

Power to
Data Center
1250kW

Flywheel UPS
(96% efficiency)
50kW

Adaptive Cooling
200kW

1000kW
"Useful Work"

Servers – 600kW
Networking – 100kW
Storage – 300kW

Best Practice
2010
Figures

PUE = 1.25

What do those numbers mean?

- 1 MW IT power = $876,000 / year
  - at $0.10 per kWh, and ignoring demand charges
  - (ignoring power factor too, because the utility eats that)
- With a PUE = 2.8 → $2,452,800
- With a PUE = 1.25 → $1,095,000

- Power delivery cost ("tax") to deliver that $876K of energy per year
  - PUE = 2.8 : $2453K - $876K = $1577K tax / yr
  - PUE = 1.25 : $1095K - $876K = $219K tax / yr

Facilities Technologies

- Covered in more detail in past tutorials
  - See "Technologies for Green Storage" at www.snia.org/education/tutorials/2009/fall#green
- Flywheel UPS
- Air handling systems (use of outside air)
- Variable speed fans
- Hot aisle / Cold Aisle technologies
- Monitoring and feedback control systems
- Spam filtering (!)

Why am I talking about facilities?

- Traditional facilities use most of the power just to deliver it
  - At a PUE of 2.8, storage probably only uses about 11% of your total power
  - Optimize the big stuff first (Amdahl's Law)
  - Understand the pressures on the facilities side
- Working with your facilities people, you can be a leader in your company in the move toward Energy Efficient Data Centers (EEDC)

On to storage

- Okay, let's talk about storage and your IT budget
- Four basic strategies
  - Make the equipment more power-efficient
    - Power supplies, fans, drive speeds, etc.
  - Less physical redundancy
    - Delta snapshots, parity RAID
  - Commit less space
    - Thin provisioning
  - Squeeze more data into available space
    - Data deduplication, compression

Power efficiency

- Capacity vs. high performance drives
- ILM / HSM
- MAID
- Tape
- SSDs
- Power supply and fan efficiency
- Disk drives

Capacity vs. high performance drives

- Capacity
  - focused on GB/watt at idle
    - 1 TB SATA: 12W
    - 4 x 250 GB FC: 64W
  - also tend to have better $/GB
  - NOTE: power use is theoretically quadratic with respect to rotational speed
    - Use the slowest drives that will fit your needs
- Performance
  - focused on seek time
    - 1 TB SATA: 12 – 15 ms
    - 300 GB SAS: 3 – 4 ms
  - also designed for higher RAS * environments
  - significant CAPEX disadvantage relative to SATA
  - significant CAPEX advantage over SSD

* RAS = Reliability, Availability, Security

ILM / HSM * / storage tiering

- 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 ILM/HSM system
    - Cost of administration
    - Cost of data movement
  - Significant renewed investment in tiering with advent of SSD
- Practice
  - Storage declines in value as it ages (like the data it holds)
  - Manual movement of data sets
- Movement toward "flash and stash"
  - Current marketing war is between systems using flash/SSD as a cache in front of slow disk, and systems using it as a separately addressed tier

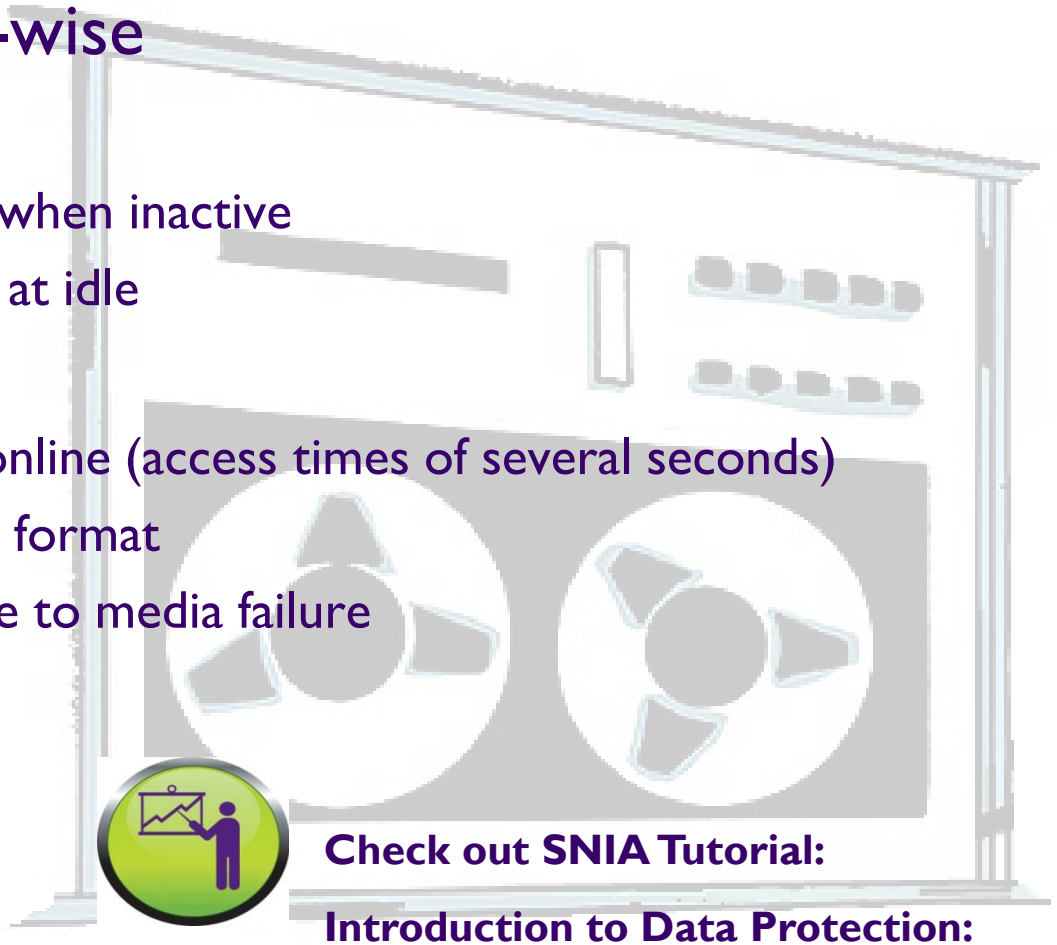
* ILM = Information Lifecycle Management
HSM = Hierarchical Storage Management

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
    - Impending competition from SSDs
    - Competition from tape
  - Best practice: spin down hot spares
    - Savings = the percentage of spares you're running

Power efficiency: Tape

- Still the best, energy-wise
  - 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 media failure



Check out SNIA Tutorial:

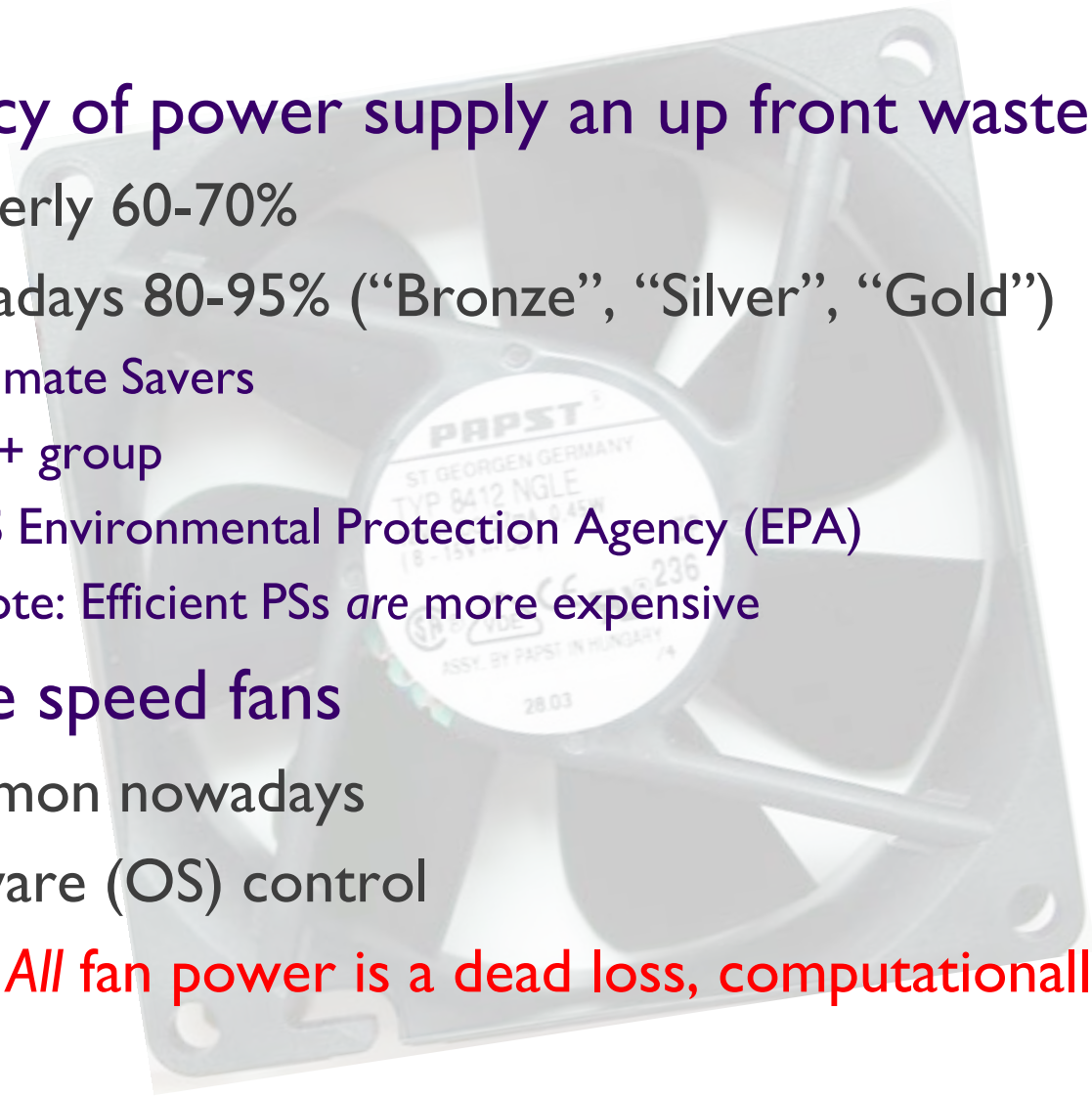
Introduction to Data Protection: Backup to Tape, Disk and Beyond

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

Power supply and fan efficiencies



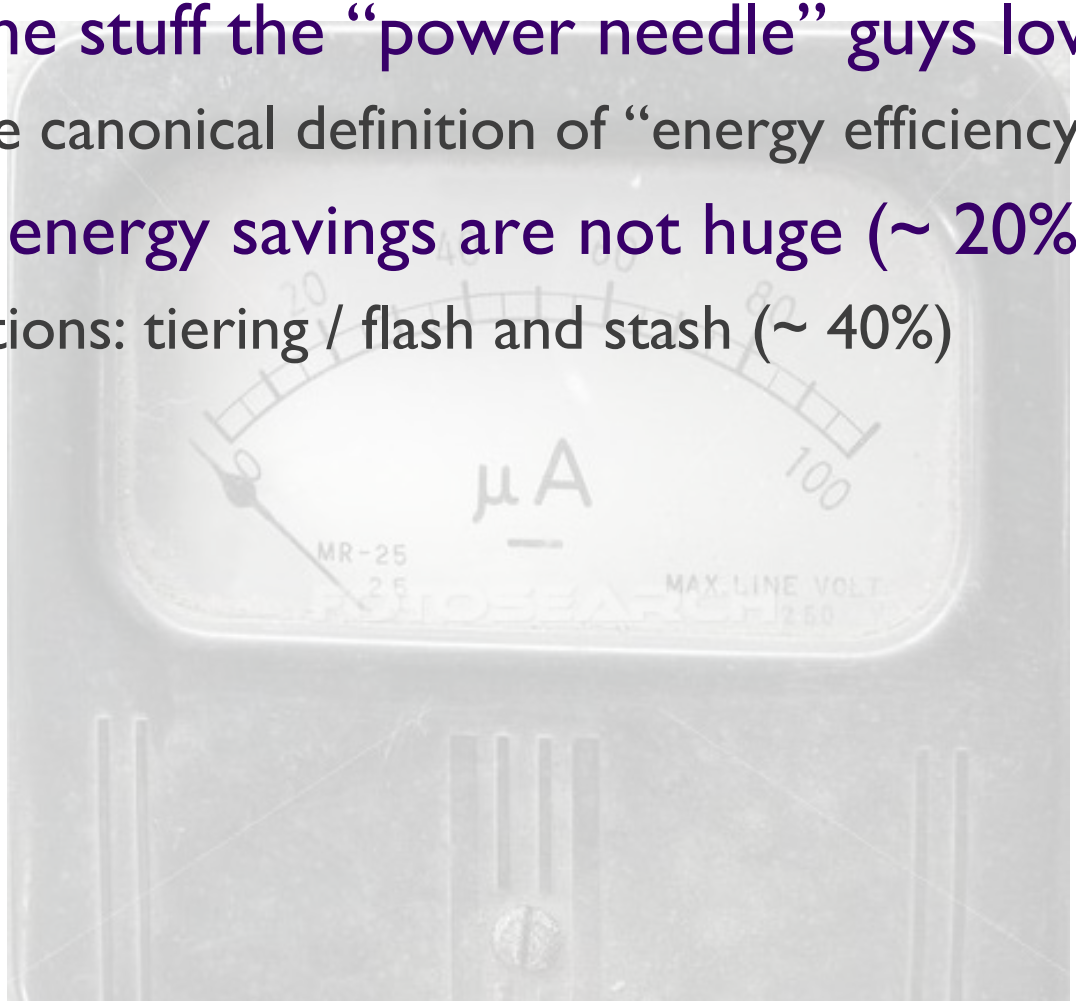
- Efficiency of power supply an up front waste
  - Formerly 60-70%
  - Nowadays 80-95% ("Bronze", "Silver", "Gold")
    - Climate Savers
    - 80+ group
    - US Environmental Protection Agency (EPA)
    - Note: Efficient PSs *are* more expensive
- Variable speed fans
  - Common nowadays
  - Software (OS) control
  - N.B. *All* fan power is a dead loss, computationally

Spinning disk drives

- Test results – SNIA Green Storage TWG (Technical Working Group)
  - Drives at "idle" – meaning no I/O to the array – use 85% or more of the power they use under load
  - Larger arrays may even use more power at idle than when under load
    - Many workloads are less intensive than housekeeping and deduplication operations that get kicked off when the array is idle
  - Slower drives (SATA) use much less power
- Best practice – Storage Tiering or SATA drives fronted by large caches
  - Except in ultra-high-performance and high write-bandwidth scenarios
  - Fairly significant energy savings – up to about 40%
    - Solid State people would like to help you get rid of the other 60% too, but the CAPEX economics aren't there at this point

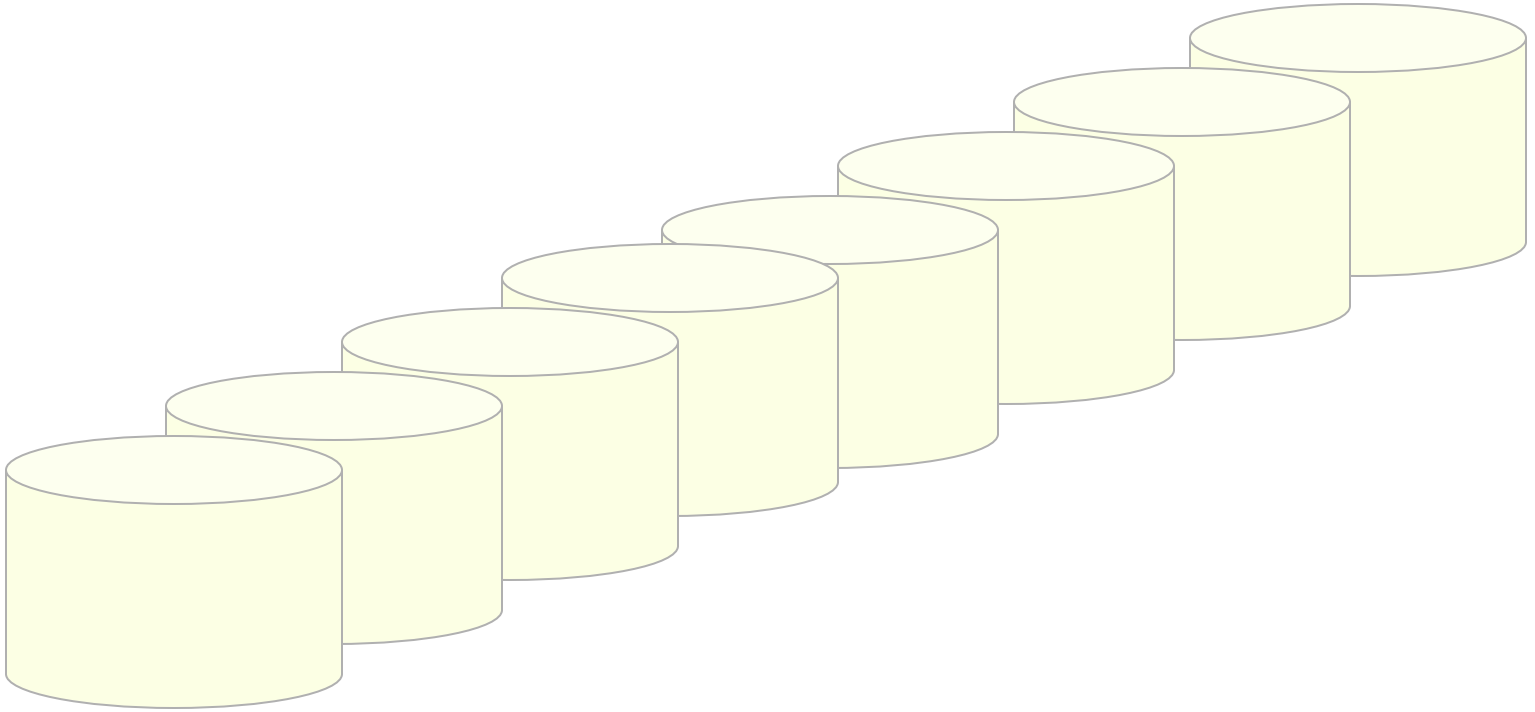
Power efficiency: Summary

- This is the stuff the "power needle" guys love
  - Fits the canonical definition of "energy efficiency"
- Overall, energy savings are not huge (~ 20%)
  - Exceptions: tiering / flash and stash (~ 40%)

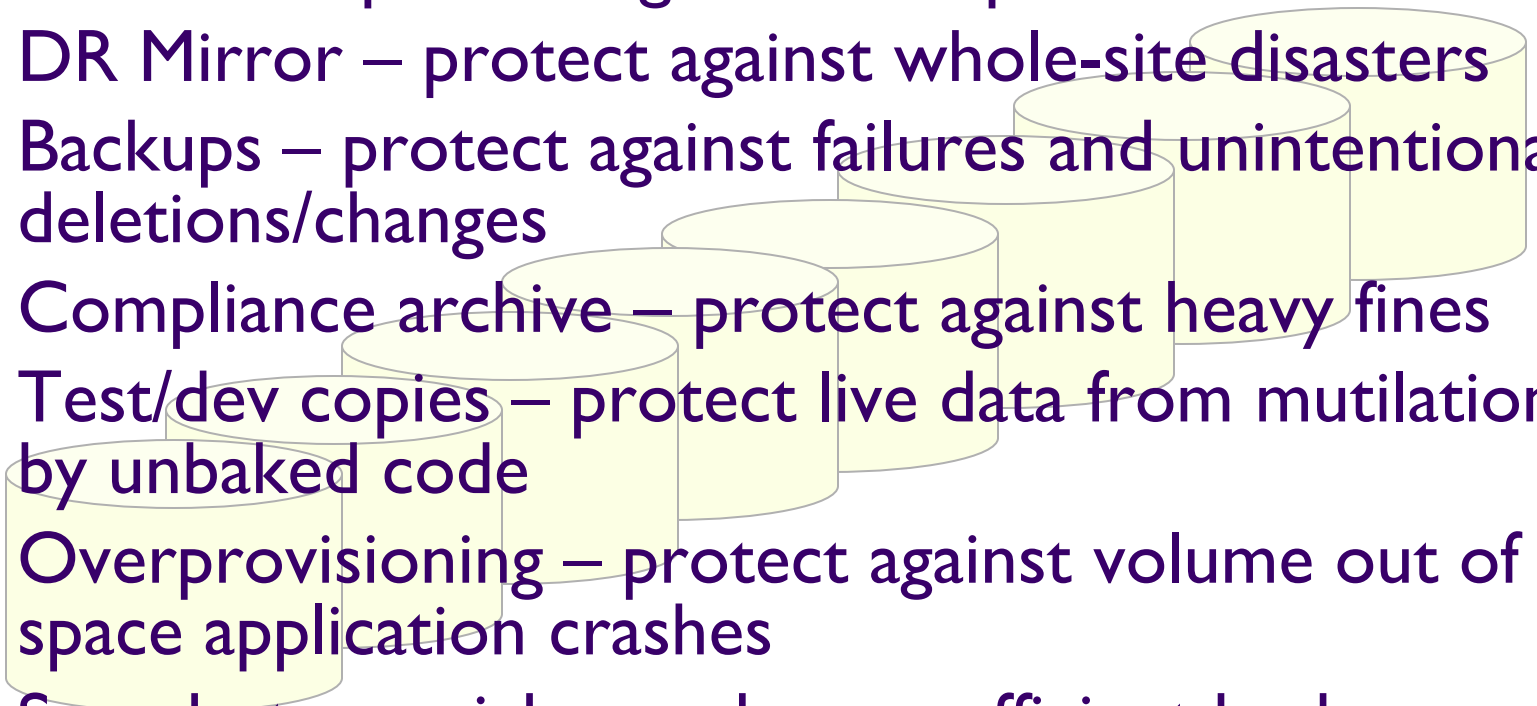


Less physical redundancy

- High value storage uses redundancy heavily
- Many enterprise applications also use redundancy



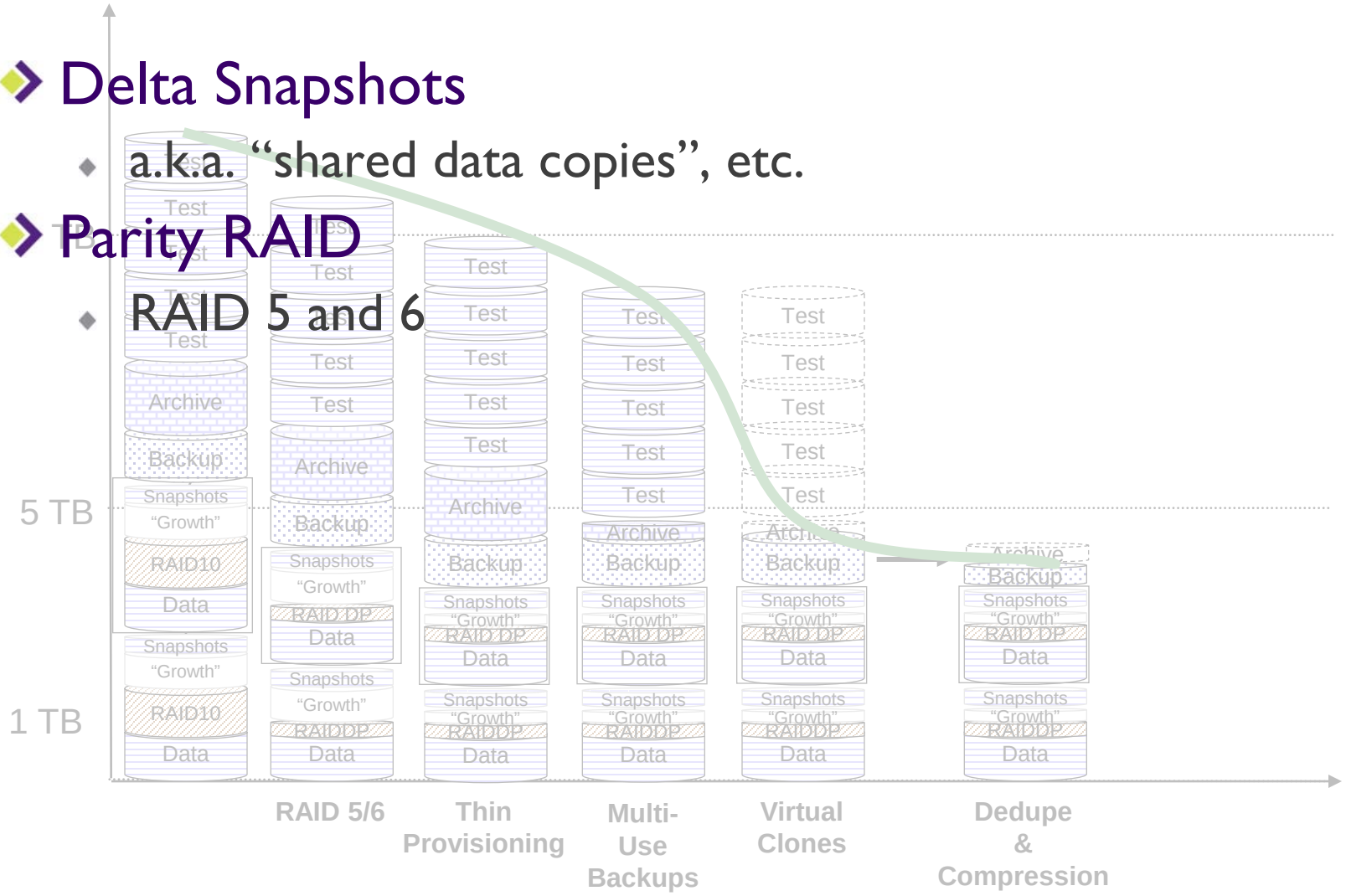
Need for redundancy



- 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

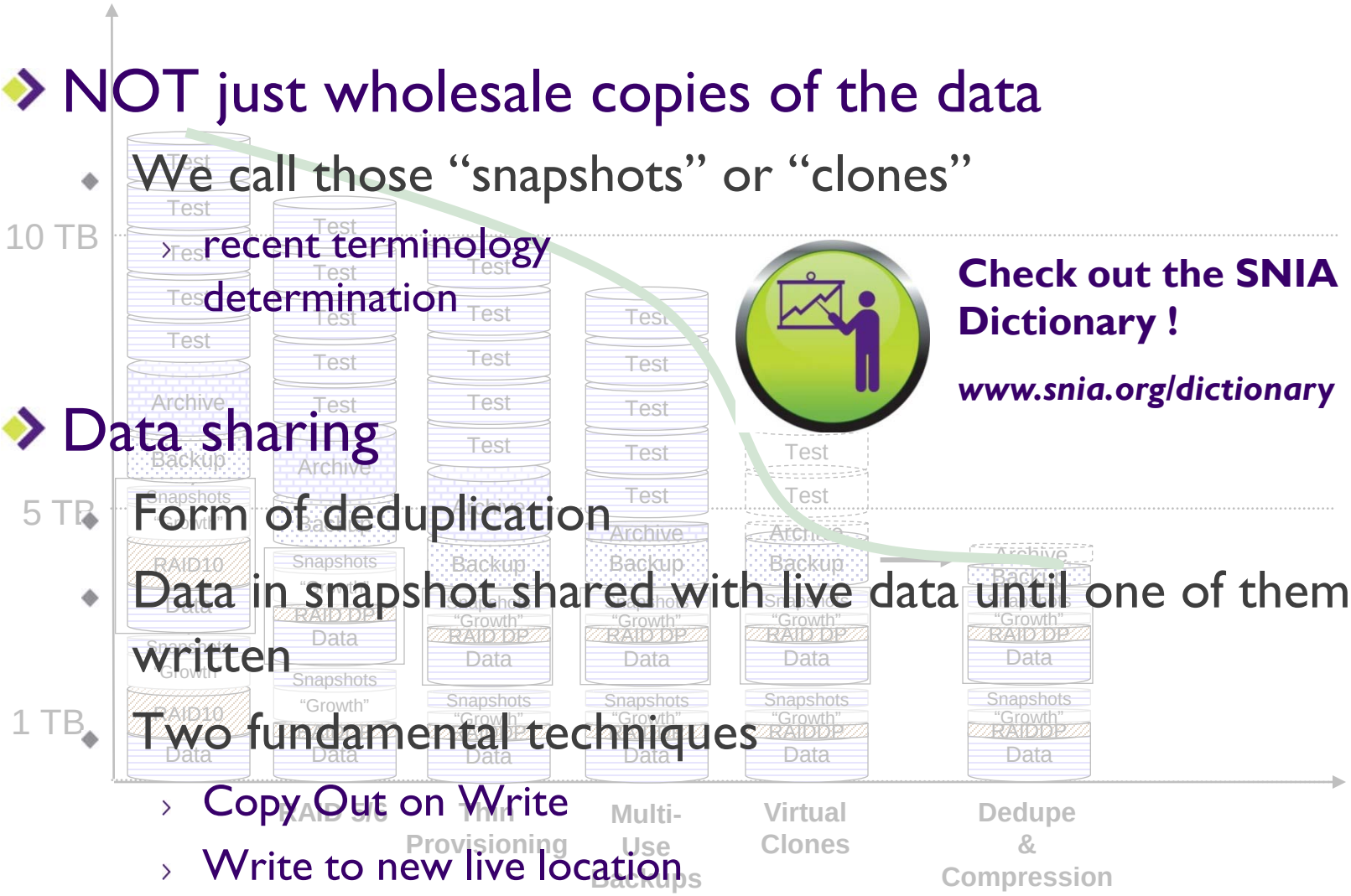
Less physical redundancy: techniques

- Delta Snapshots
  - a.k.a. "shared data copies", etc.
- Parity RAID
  - RAID 5 and 6



Delta snapshots

- NOT just wholesale copies of the data
  - We call those "snapshots" or "clones"
    - recent terminology determination
- 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

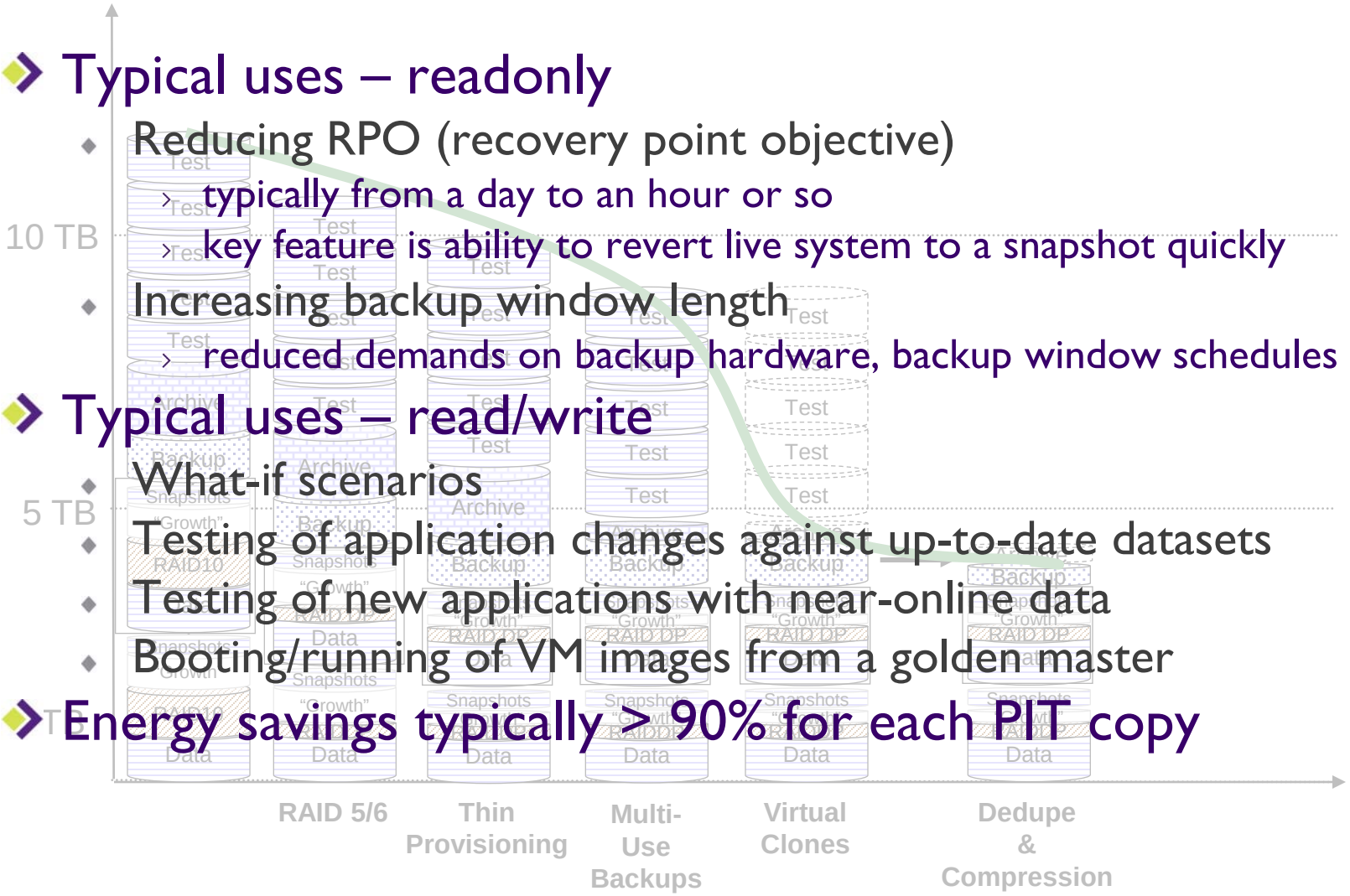


Check out the SNIA Dictionary !

www.snia.org/dictionary

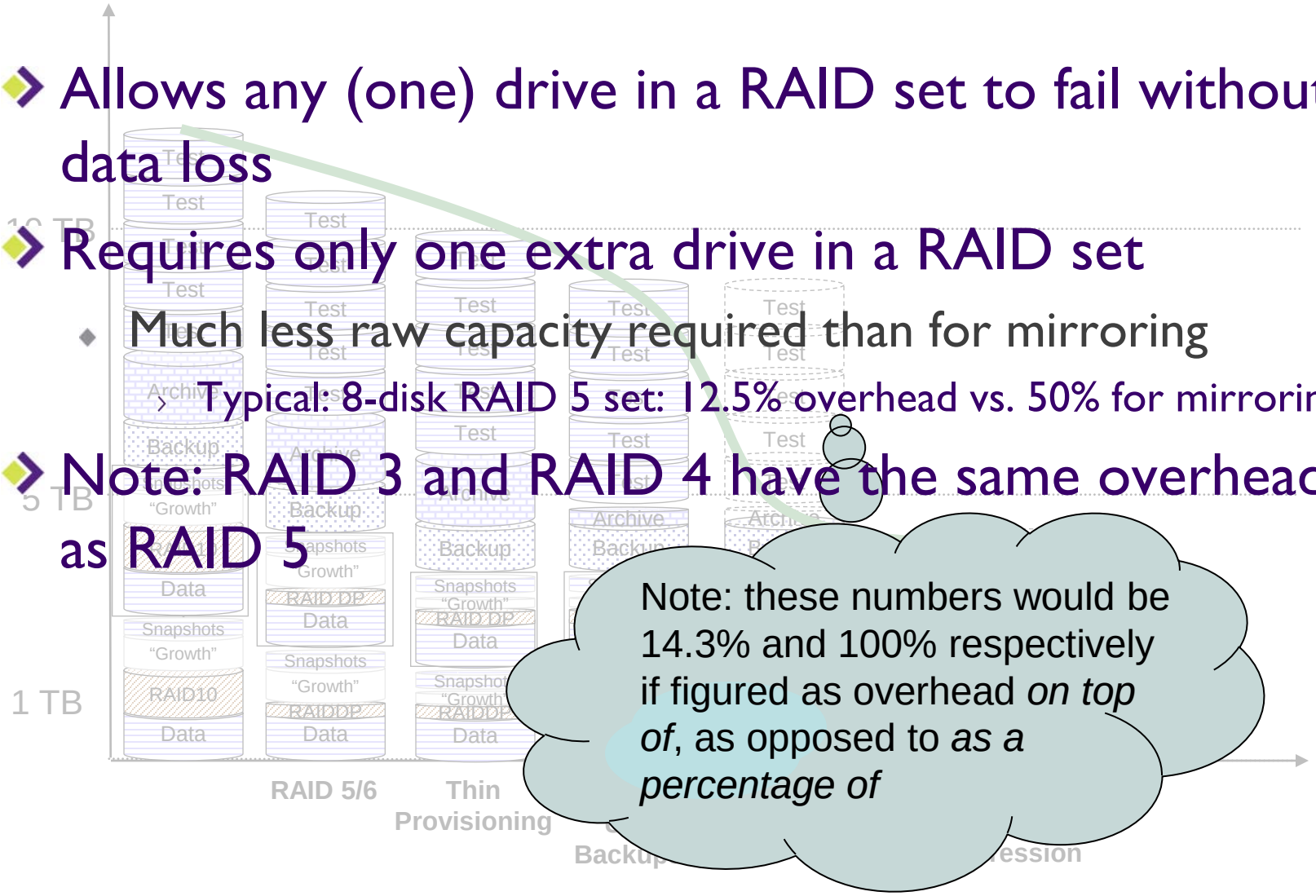
Delta snapshots

- 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
  - 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
- Energy savings typically > 90% for each PIT copy



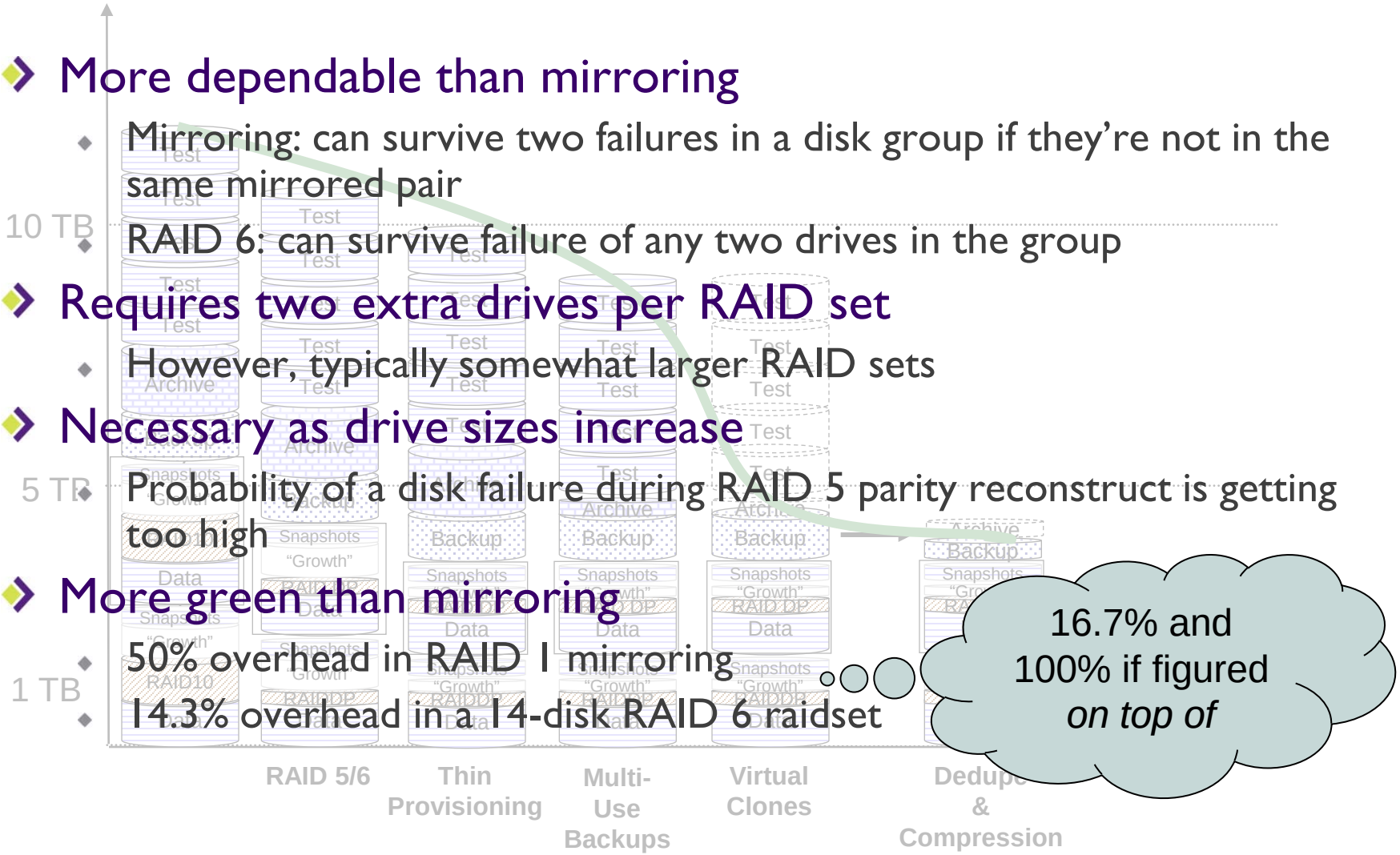
RAID 5

- 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



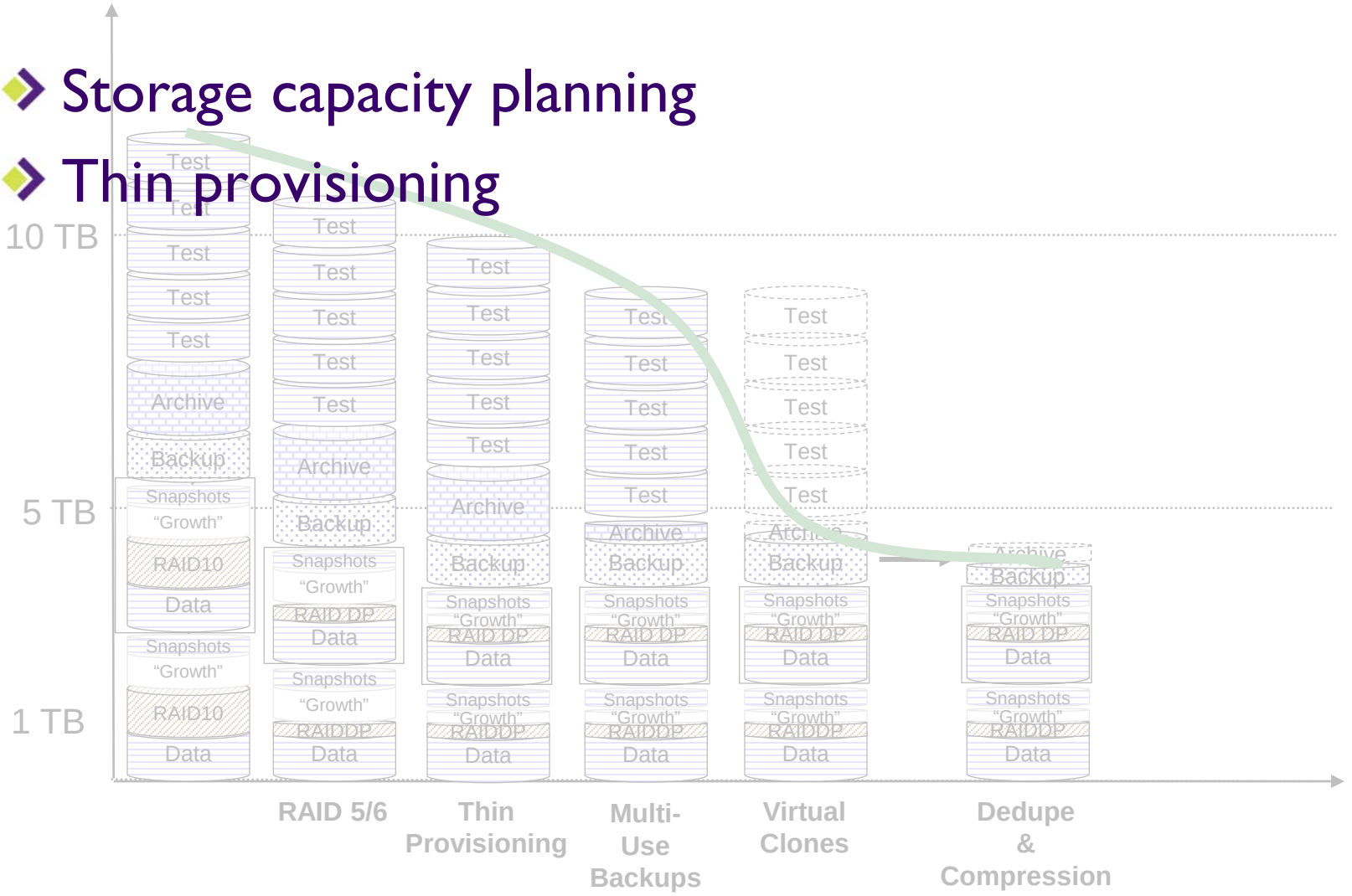
RAID 6

- 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 raidset



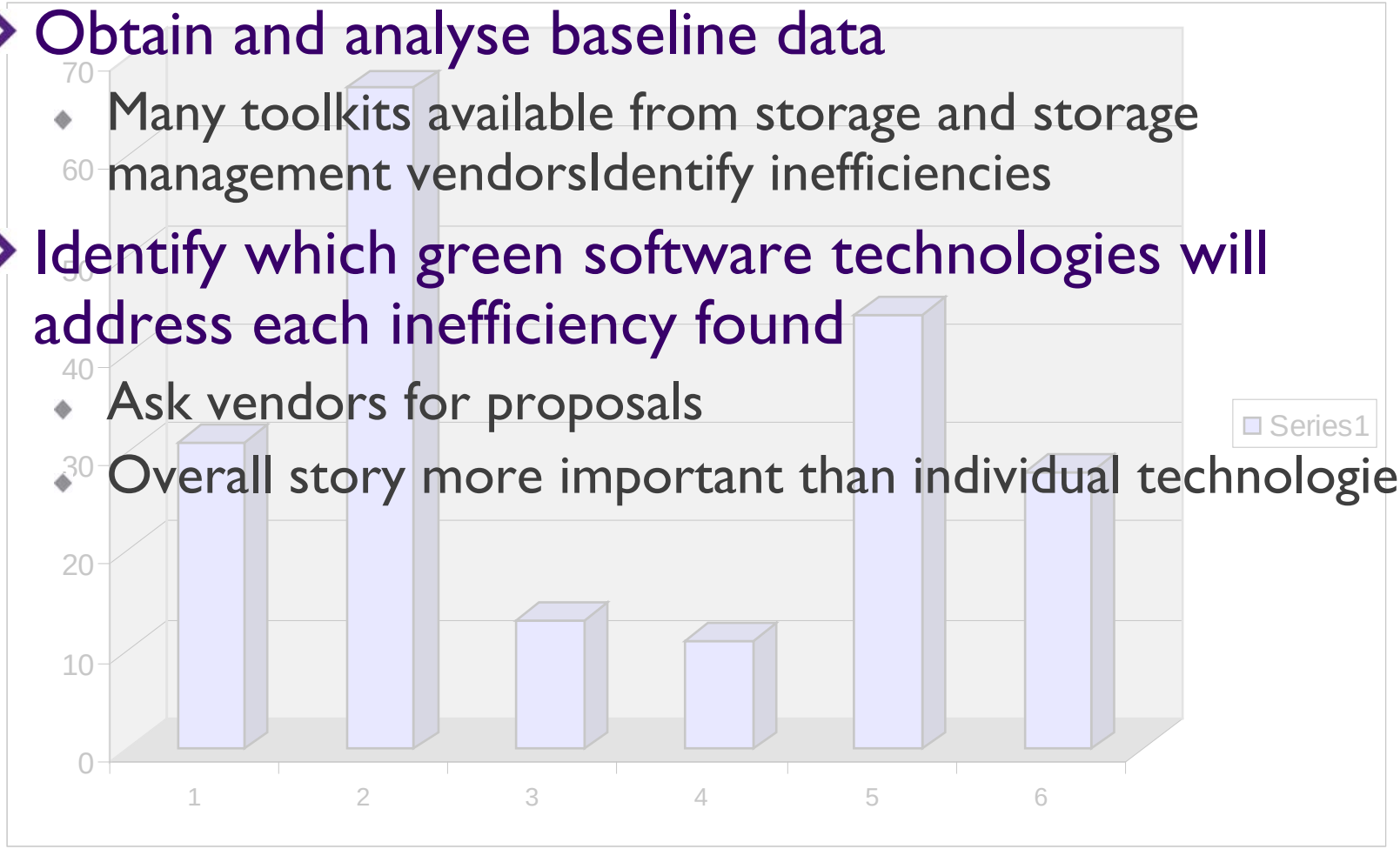
Committing less space

- Storage capacity planning
- Thin provisioning



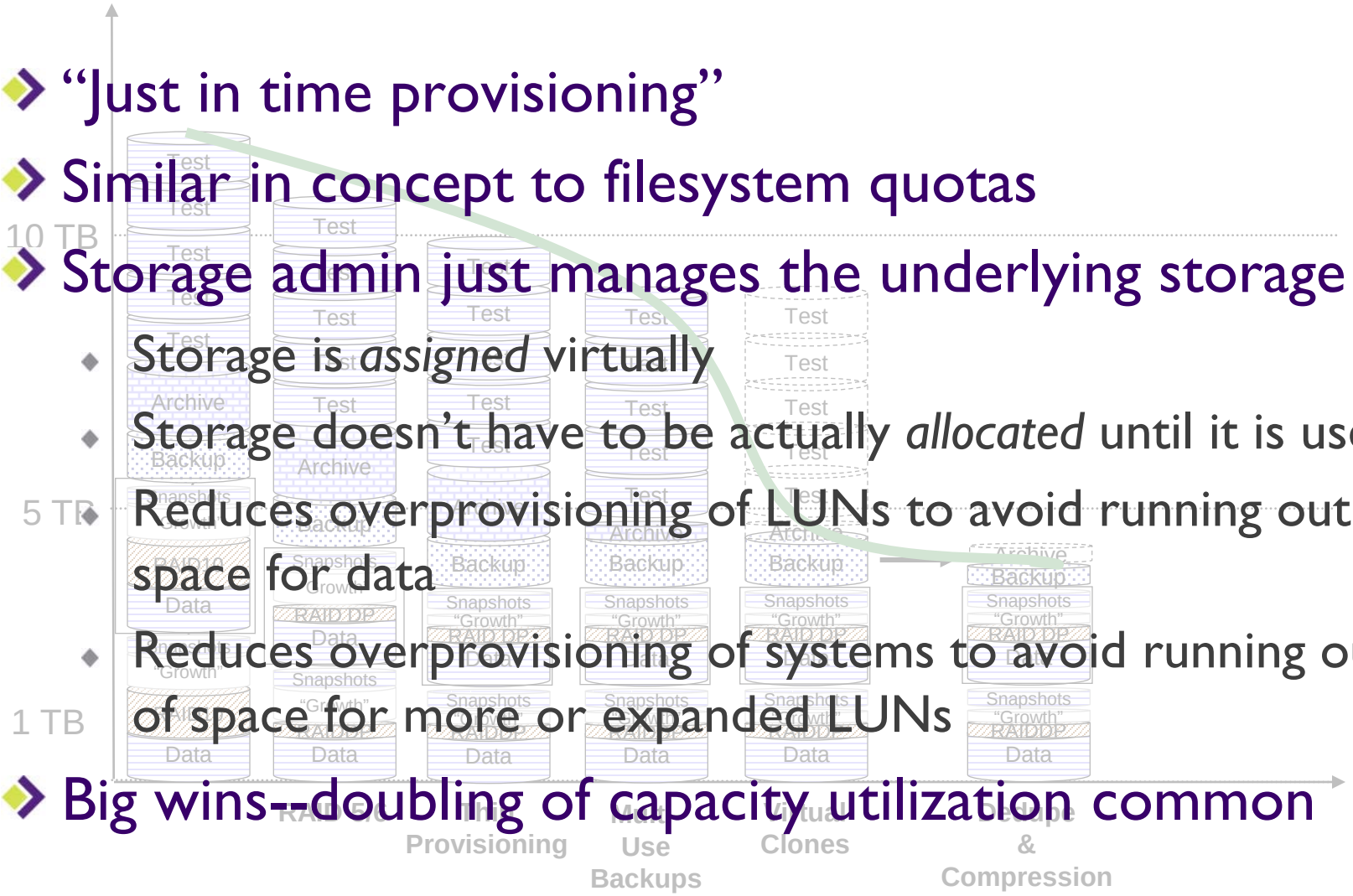
Storage capacity planning

- Obtain and analyse baseline data
  - Many toolkits available from storage and storage management vendorsIdentify inefficiencies
- Identify which green software technologies will address each inefficiency found
  - Ask vendors for proposals
  - Overall story more important than individual technologies



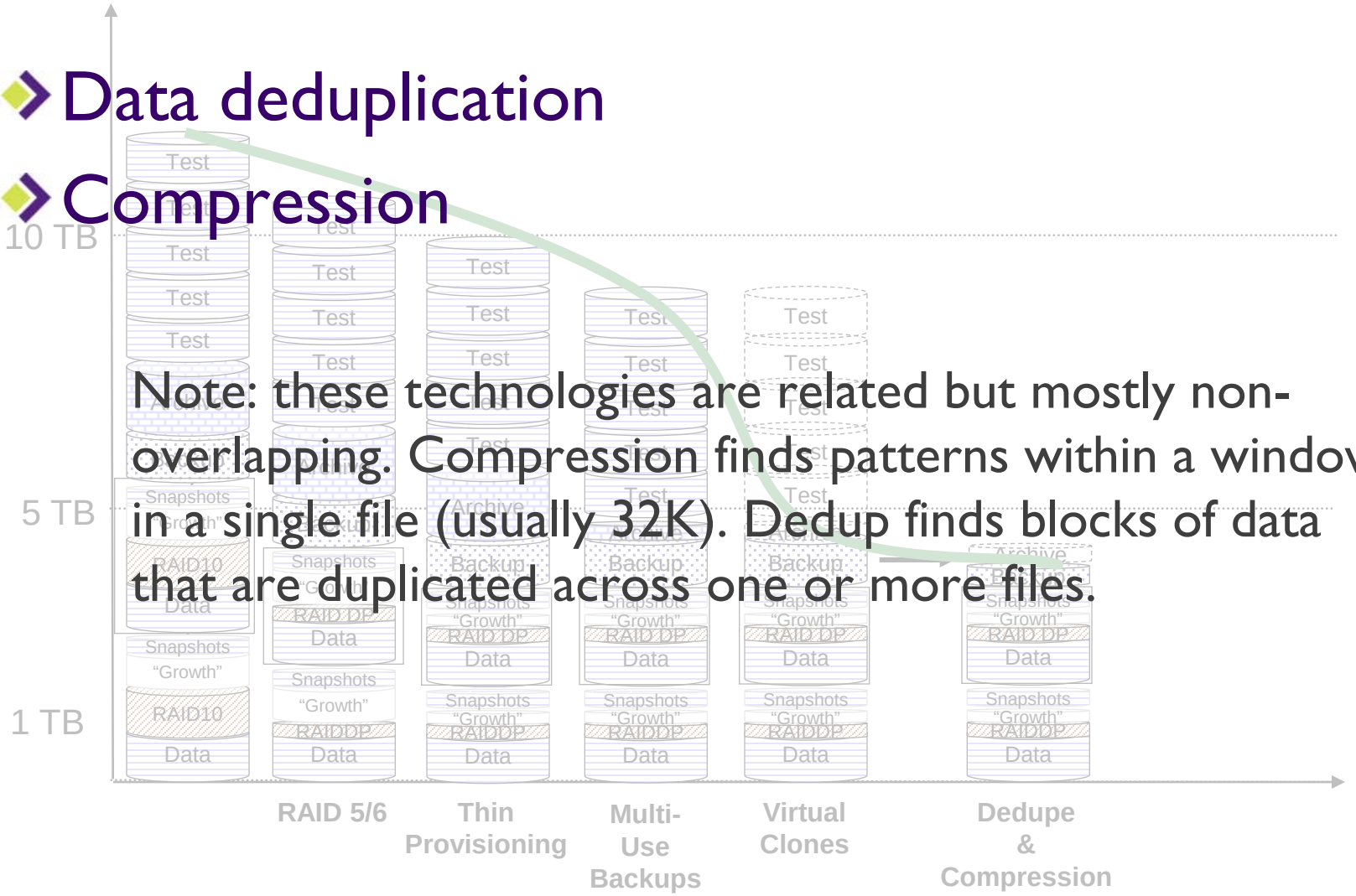
Thin provisioning

- "Just in time provisioning"
- Similar in concept to filesystem quotas
- Storage admin just manages the underlying storage
  - Storage is *assigned* virtually
  - Storage doesn't have to be actually *allocated* until it is used
  - Reduces overprovisioning of LUNs to avoid running out of space for data
  - Reduces overprovisioning of systems to avoid running out of space for more or expanded LUNs
- Big wins--doubling of capacity utilization common



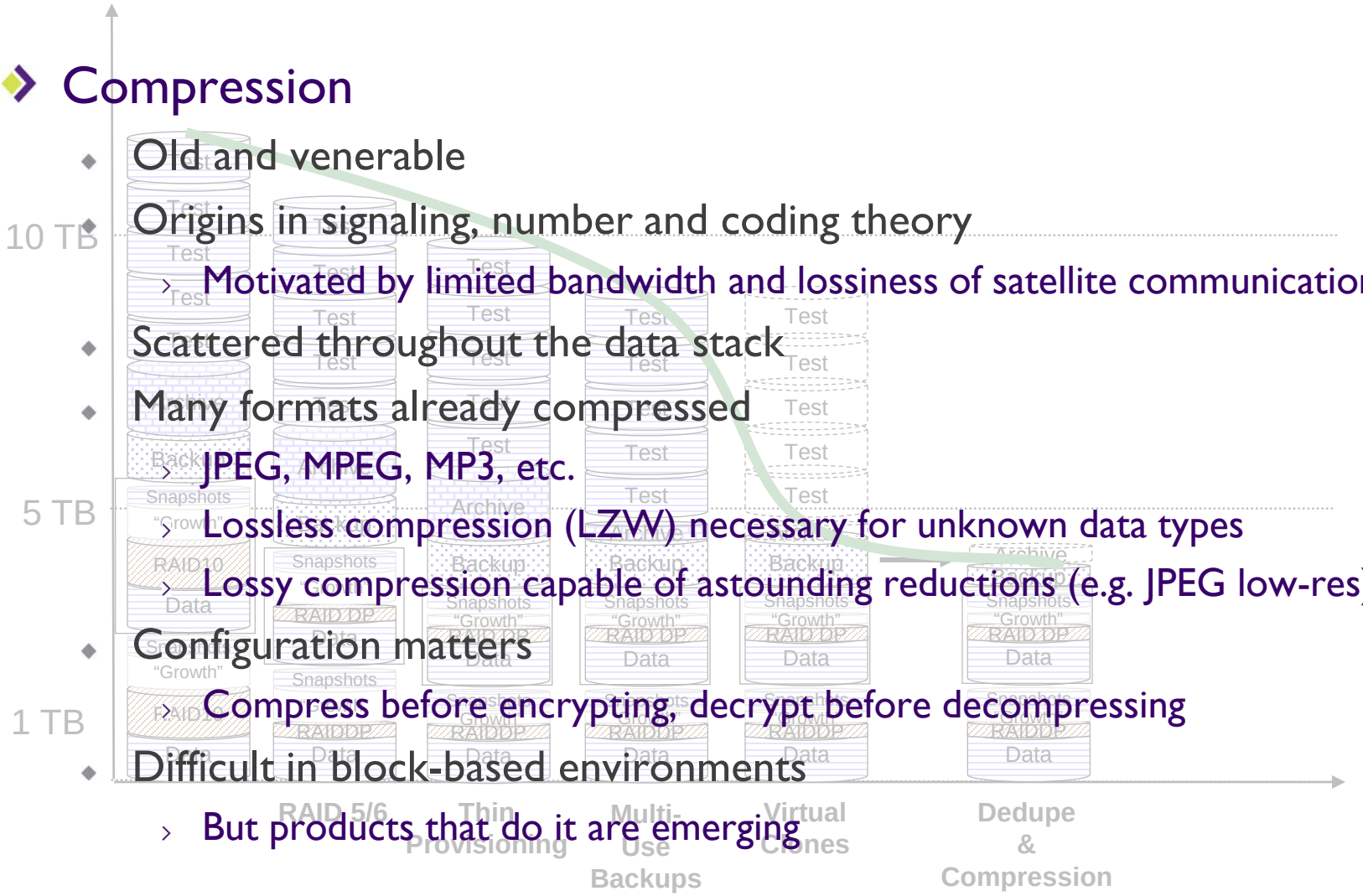
Squeezing in more data

- Data deduplication
- Compression



Note: these technologies are related but mostly non-overlapping. Compression finds patterns within a window in a single file (usually 32K). Dedup finds blocks of data that are duplicated across one or more files.

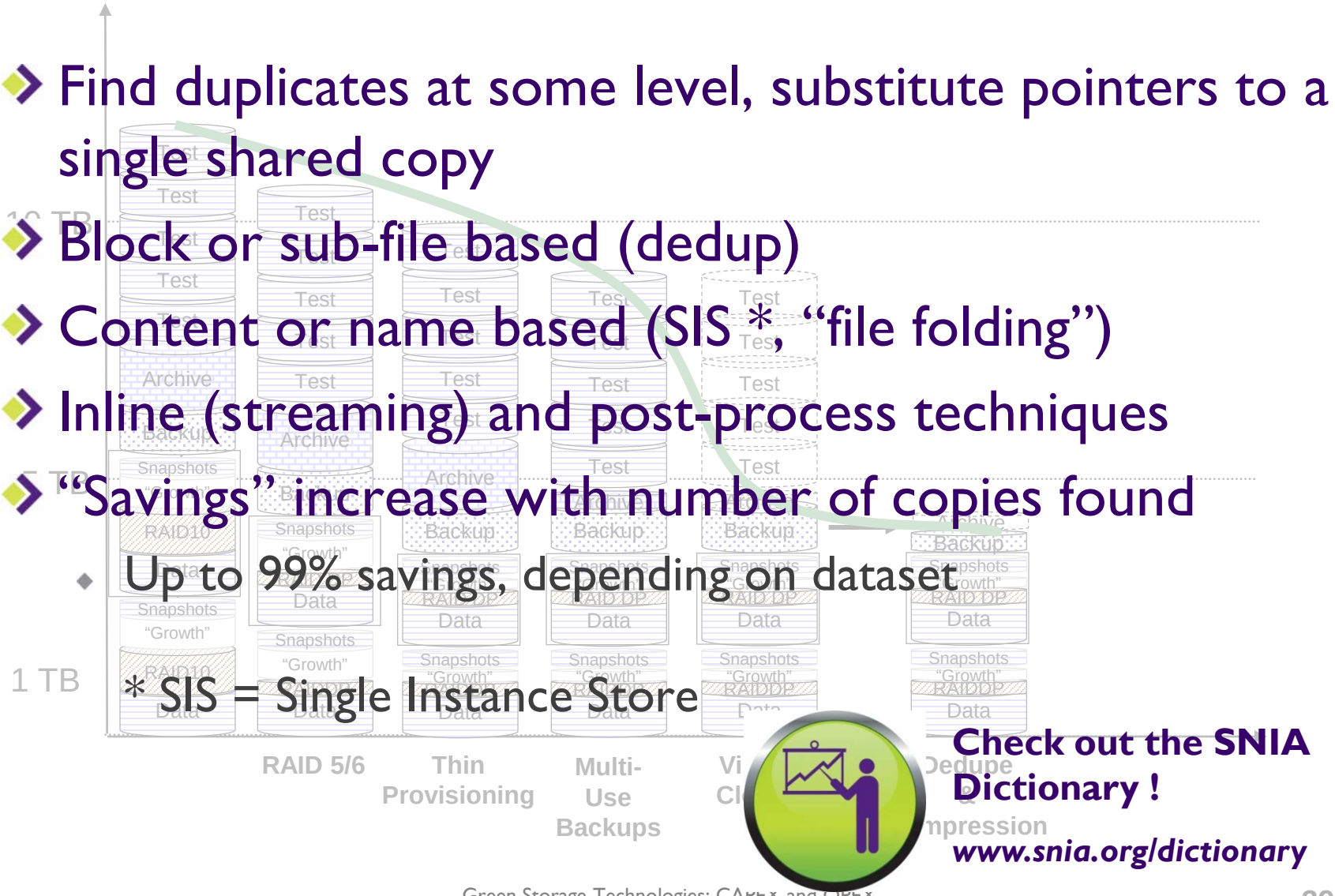
Compression



- 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
    - Lossy compression capable of astounding reductions (e.g. JPEG low-res)
  - Configuration matters
    - Compress before encrypting, decrypt before decompressing
  - Difficult in block-based environments
    - But products that do it are emerging

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
  - Up to 99% savings, depending on dataset



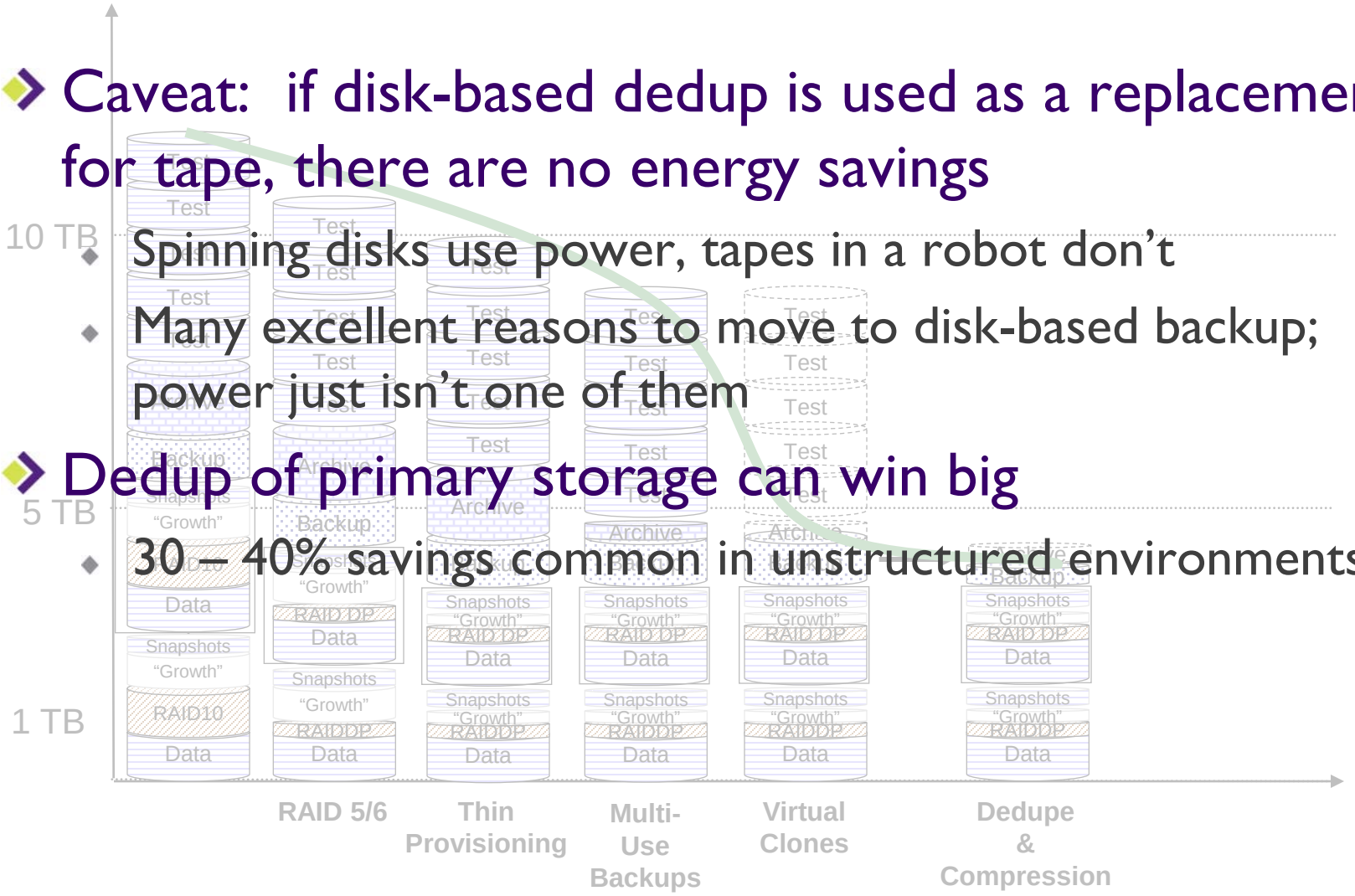
* SIS = Single Instance Store

Check out the SNIA Dictionary !

www.snia.org/dictionary

Dedup and energy

- Caveat: if disk-based dedup is used as a replacement for tape, there are no energy savings
  - Spinning disks use power, tapes in a robot don't
  - Many excellent reasons to move to disk-based backup; power just isn't one of them
- Dedup of primary storage can win big
  - 30 – 40% savings common in unstructured environments



Savings calculations

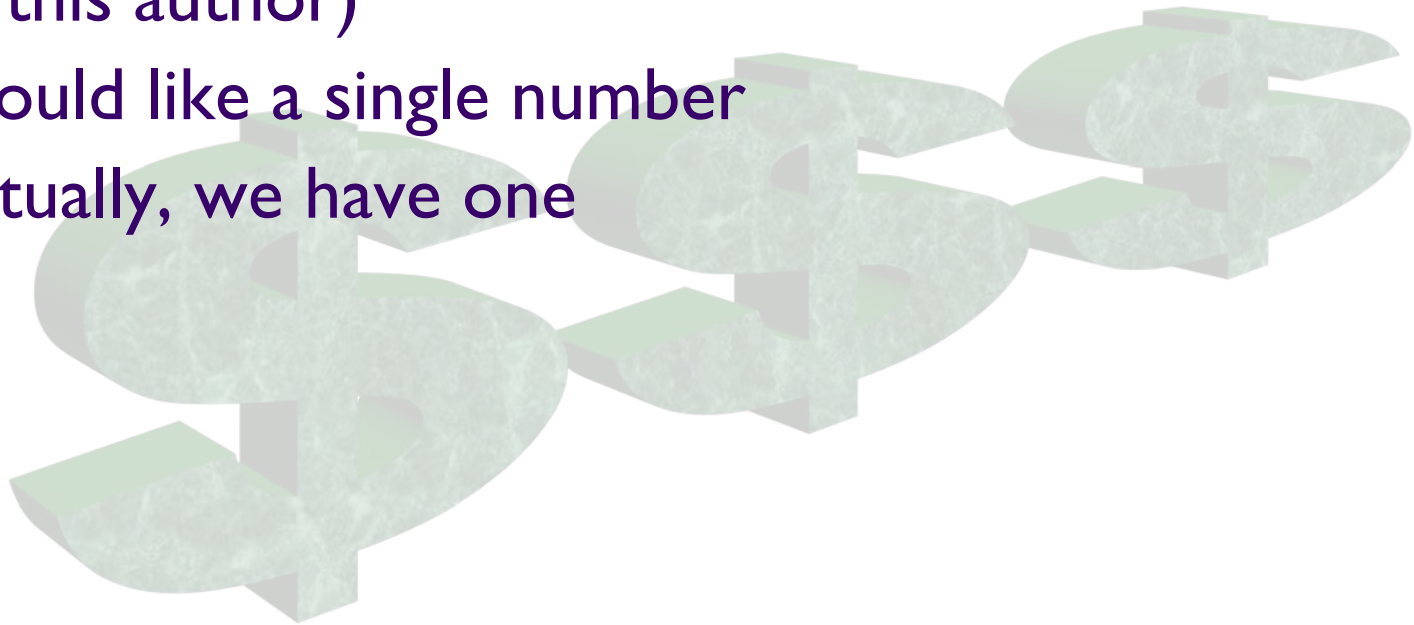
- Facilities power savings
- Equipment power savings
- Capacity savings
  - Dedup and compression
  - Thin provisioning
  - Delta snapshots, advanced RAID

Typical savings

- Facilities
  - ~50% typical when moving to state of the art
- Equipment power savings
  - 20 – 30% overall (EPA says low 20's)
- Capacity savings
  - Dedup and compression
    - Primary storage (unstructured data): 25 – 40%
    - Secondary storage (streaming compression only): 50%
  - Thin provisioning
    - Move from 30% utilized to 80% utilized
  - Delta snapshots, advanced RAID
    - Savings linear in the # of snapshots
    - 10 – 20% overhead for RAID 5/6 vs. 50% for RAID 1/10

Problems with measuring gains

- Difficult to evaluate how the various capacity-saving technologies work together (see previous tutorials by this author)
- Would like a single number
- Actually, we have one



Storage efficiency

- A way to look at--and actually measure--overall gains
- Three key quantities
  - How much data did I store?
  - How much unused usable space do I have left?
  - How much raw capacity did I start out with
- Ratio of the first two to the third is my *storage efficiency*
  - $$\text{Storage efficiency} = \frac{\text{size of data + free space}}{\text{raw capacity}}$$

Storage efficiency cont.

- Why are we including free space?
  - Can (theoretically) always store as much data as there is free space
  - Doesn't seem right to have an efficiency of zero for an empty system
  - Do want to reward full use of systems
    - Using 10% of your storage very efficiently is not efficient
    - (Partial calculations ignore this reality)

$$\text{Storage efficiency} = \frac{\text{size of data + free space}}{\text{raw capacity}}$$

Storage Efficiency - definitions

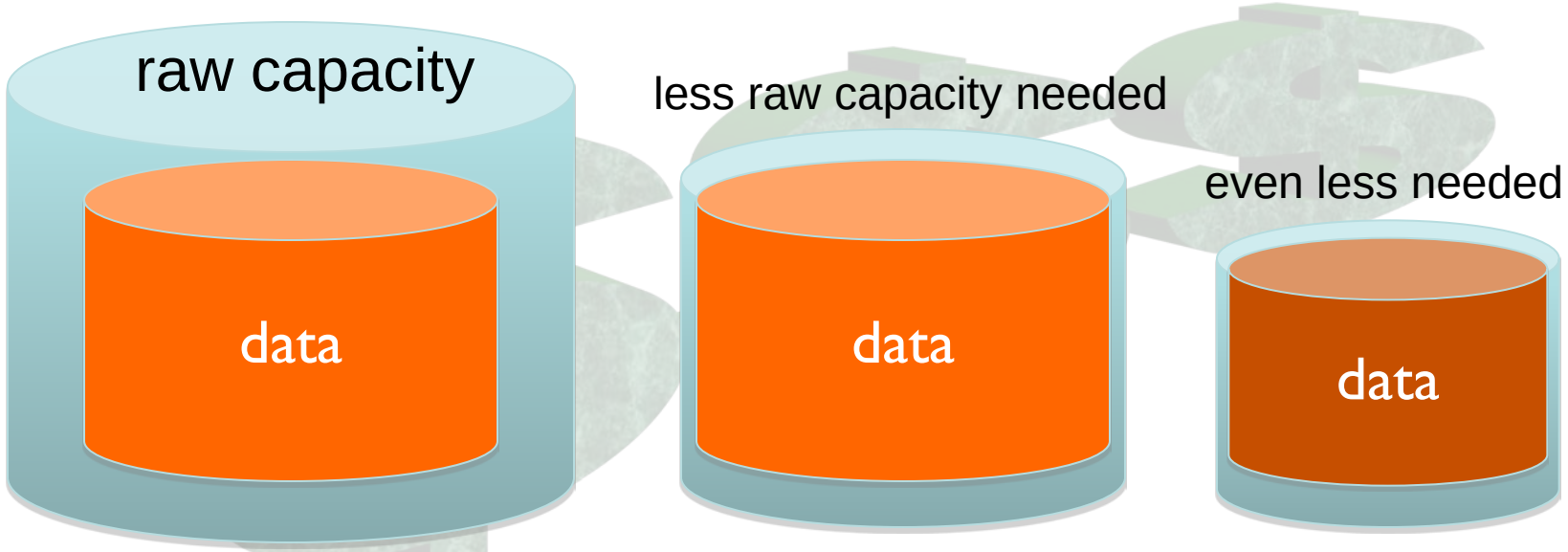
- Raw capacity
  - What the manufacturer says
- Formatted (usable) capacity
  - Raw capacity minus system overhead
  - Anything that can be assigned by the storage admin for application-level use
- Assigned capacity
  - Nominal size of a thin provisioned container
  - Can exceed formatted capacity
- Effective capacity
  - Amount of data that has been crammed into a container plus leftover free space
- Storage efficiency
  - Effective capacity / raw capacity

Check out the SNIA Dictionary !

www.snia.org/dictionary

Storage efficiency - implications

- Large positive impact to IT budget
  - Energy savings help the facilities budget as well



E.g.	SE = 50% (RAID 1)	SE = 80% (RAID 6)	SE = 104% (add dedup)
Cost	$1,000,000	$625,000	$481,000

Takehome: the heavy hitters

- Facilities
  - Resources: The Green Grid, DOE
- Thin provisioning
  - Your kit probably has it: USE IT!
  - Document and compare before and after utilization, if possible
  - 100% improvement in capacity utilization often achievable
- Fat slow drives
  - Archive data, 2nd tier storage
  - Greatly expanded possibilities when fronted by large cache or SSD tier
    - Most unstructured data
    - Structured data with medium or low performance requirements
- Parity RAID
  - If your vendor says RAID 6 doesn't work, it means they don't have it
  - 40% reduction in raw capacity requirements

What's up at the EPA?

- Not much progress on a storage spec
- Things in the way:
  - Federal budgeting issues causing internal stress
  - Insistence on test data
    - Test fatigue by product vendors
    - EPA unwilling to sponsor testing
  - Uncertainty over definition of a "product"
    - *n* Controllers + *m* drive shelves + *p* cache modules + *q* HBAs ...
- Alternatives
  - SNIA Emerald™ program
  - European CoC
  - China, Japan, Korea, AU

Q&A / Feedback

- Please send any questions or comments on this presentation to *trackgreenstorage@snia.org*

Check out our other SNIA Green Tutorials!

Many thanks to the following individuals for their contributions to this tutorial.

- SNIA Education Committee

Joseph Glider
Wayne Adams
SW Worth
Larry Freeman