Education

Technologies for Green Storage

Alan G. Yoder, Ph.D. NetApp

Author: 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

- Technologies for Green Storage
  - Hardware efficiencies are essential to reducing the amount of power used by storage. Equally real savings are obtained by reducing the number of copies of your data that must be made, kept and managed. This talk presents a number of technologies, ranging from thin provisioning and virtualization to bypass and flywheel UPSs, that each address part of the problem, and illustrates the impact that each technology can have on your data center footprint.

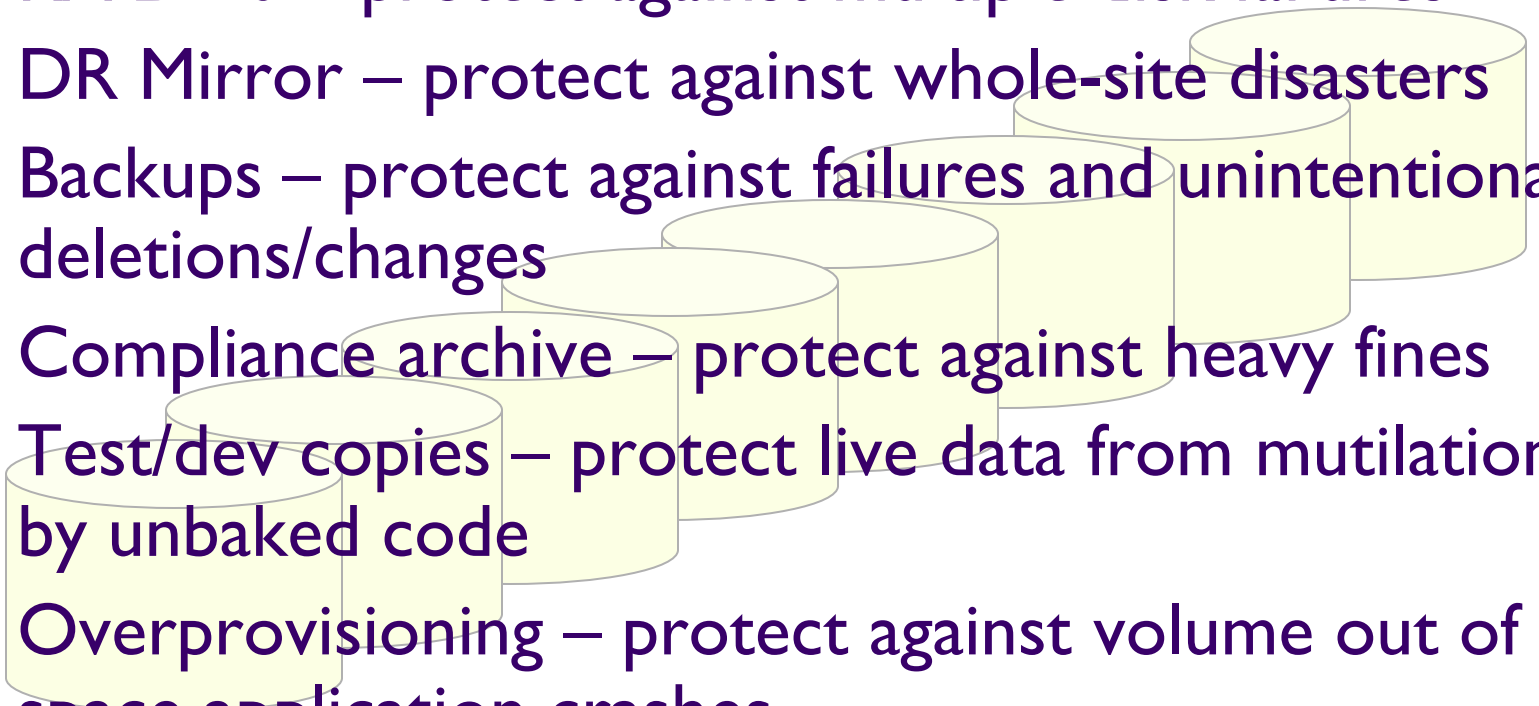
Outline

- Objectives
- Need for redundancy (extra copies of data)
- Software technologies for reducing excessive redundancy
- Other technologies for energy saving
- Data center technologies
- Typical savings

Objective

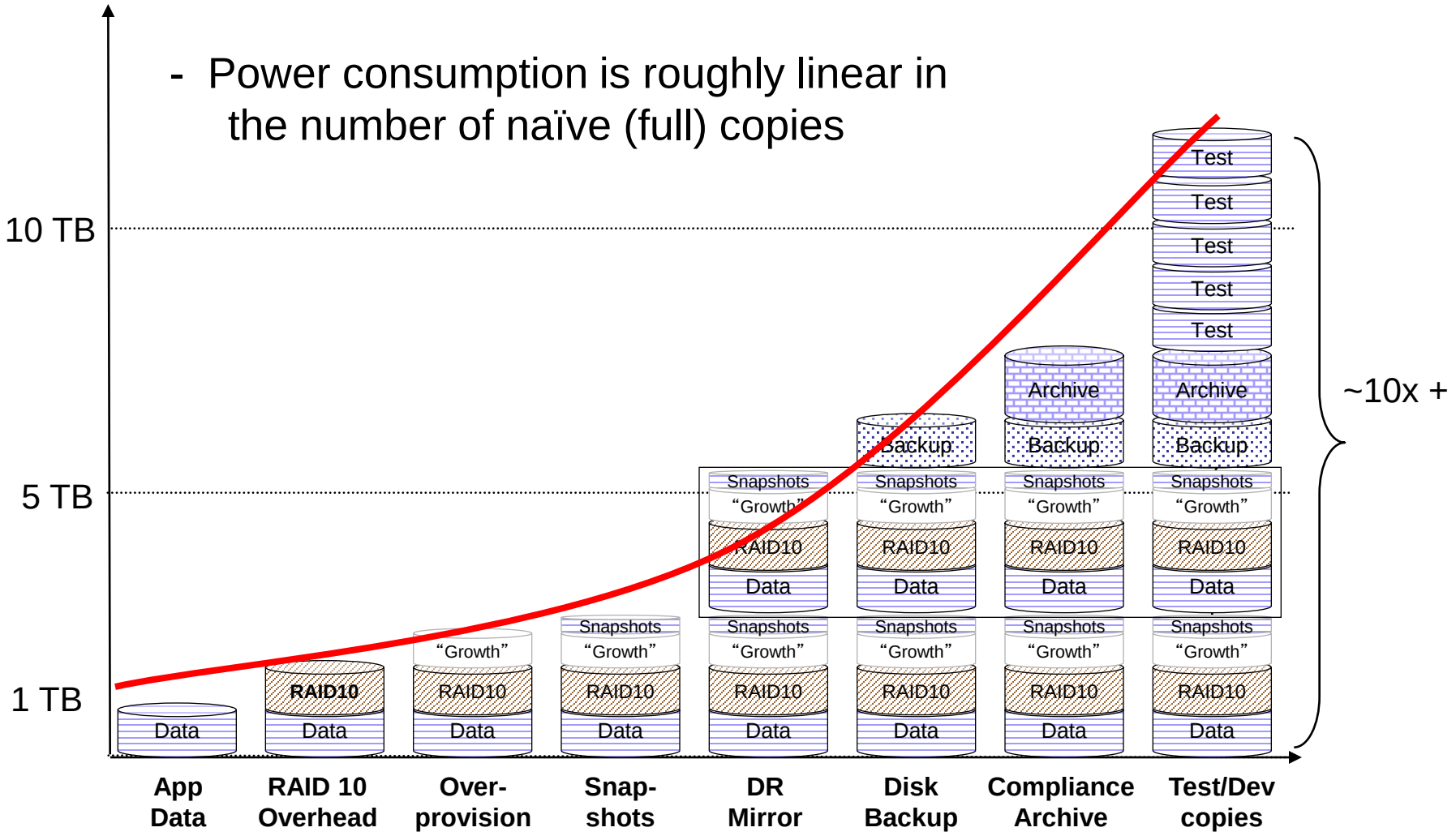
- Get more work done
- Reduce data center footprint
  - in space
    - less storage equipment to buy
  - in energy
    - more energy-efficient equipment
    - less equipment to cool
    - better cooling methodologies
    - better power management
  - in administrative costs
    - less storage equipment to manage

Need for redundancy



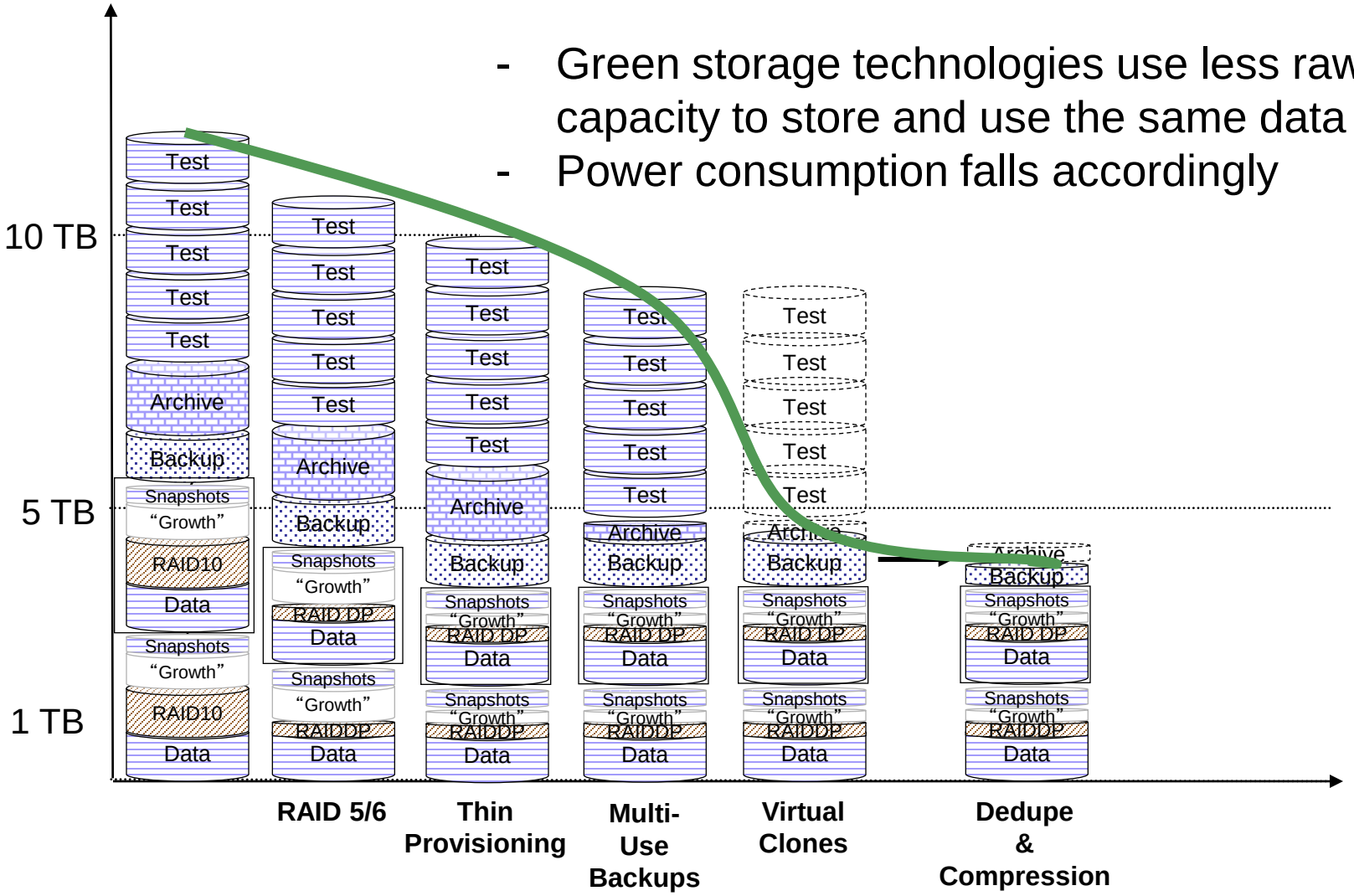
- RAID 10 – protect against multiple disk failures
- DR Mirror – protect against whole-site disasters
- Backups – protect against failures and unintentional deletions/changes
- Compliance archive – protect against heavy fines
- Test/dev copies – protect live data from mutilation by unbaked code
- Overprovisioning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

Result of redundancy



Effect of green technologies

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



Green Storage Technologies

- Enabling technologies
  - Storage virtualization
  - Storage capacity planning
- Green software
  - Compression
  - Delta Snapshots
  - Thin provisioning
  - Parity RAID
  - Deduplication and SIS

Green storage technologies (cont.)

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

Enabling technologies

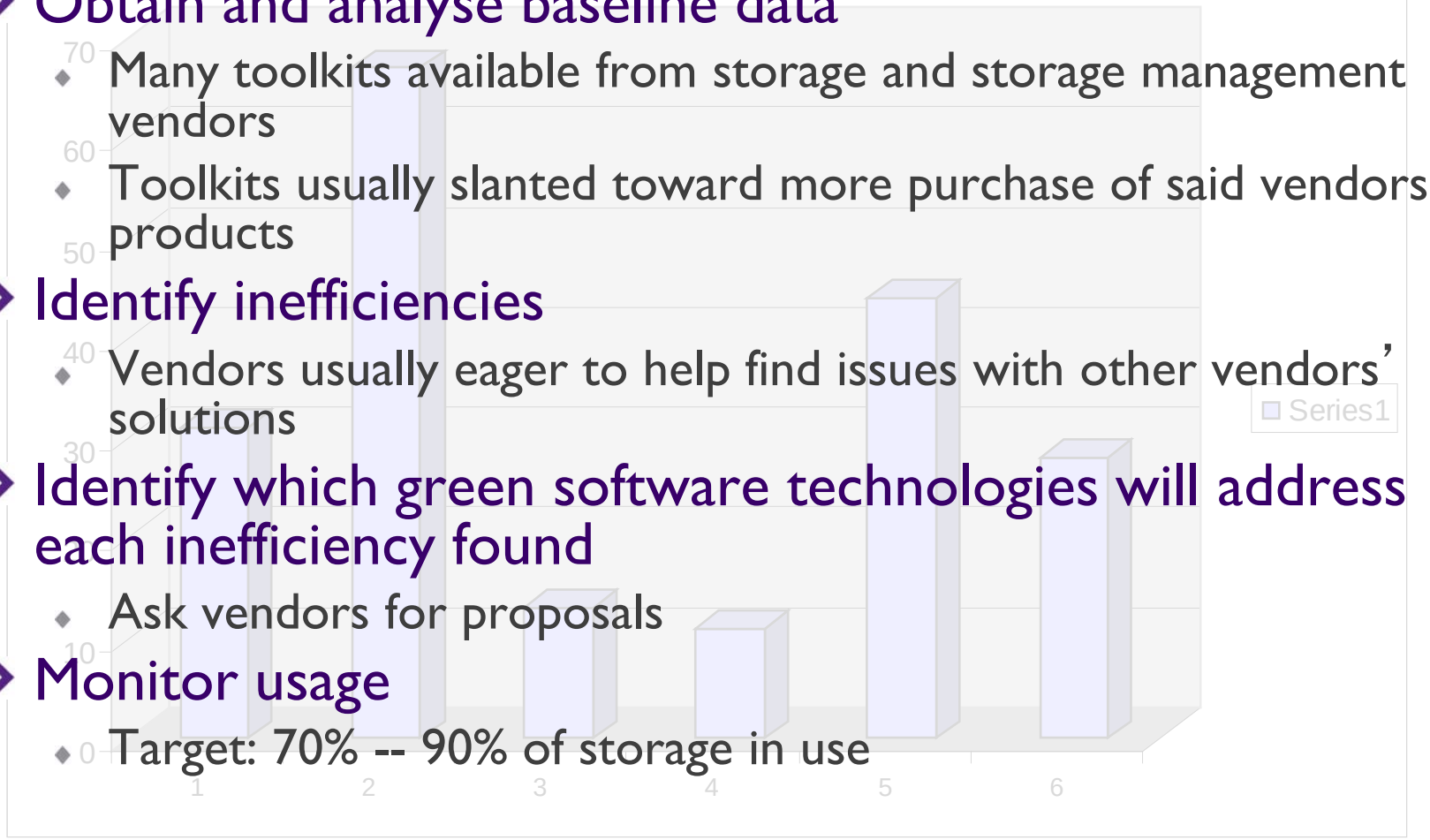
- Storage virtualization
- Storage capacity planning and monitoring

Storage Virtualization

- 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
  - *Very green* wrt servers
- But foundational for some green technologies
  - Thin provisioning
  - Resizeable volumes
  - Snapshots and virtual clones (depending on implementation)
- Also contributes in other areas
  - Flexibility, manageability, etc.

Storage capacity planning & monitoring

- 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
- Monitor usage
  - Target: 70% -- 90% of storage in use

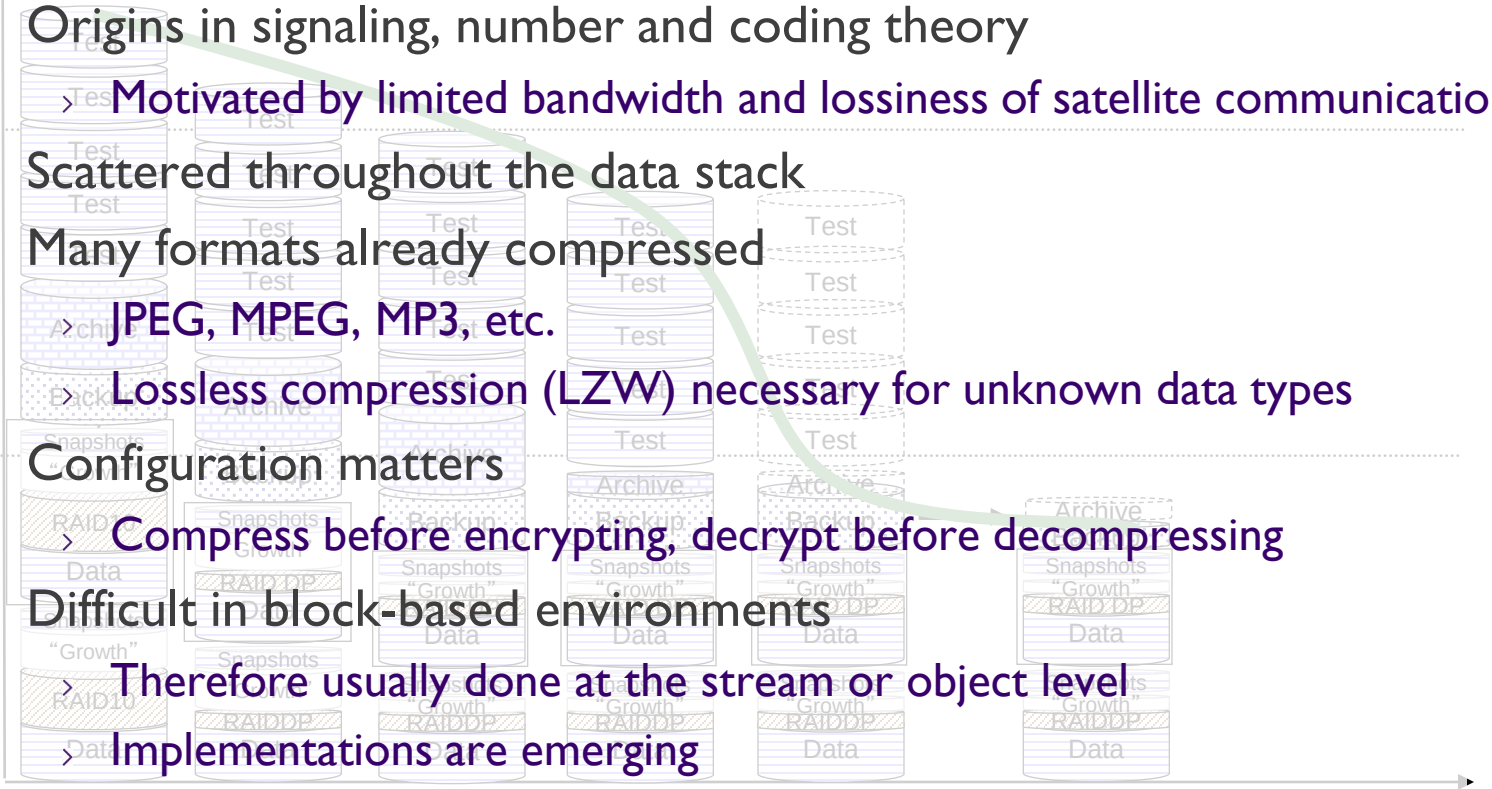


Green software technologies

- Compression
- Delta Snapshots
- Thin provisioning
- Parity RAID
- Deduplication and SIS

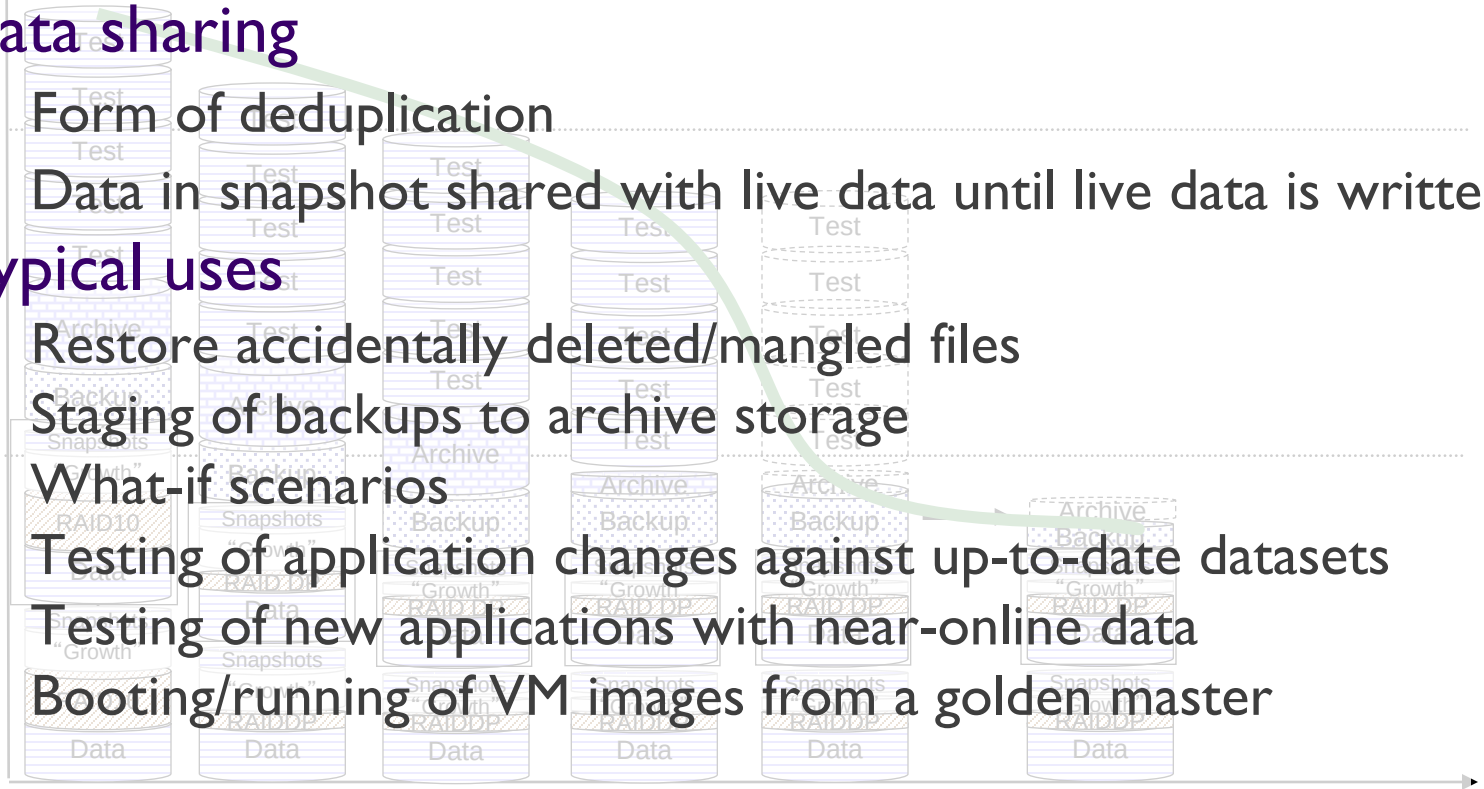
Compression

- 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
    - Therefore usually done at the stream or object level
    - Implementations are emerging



Delta Snapshots

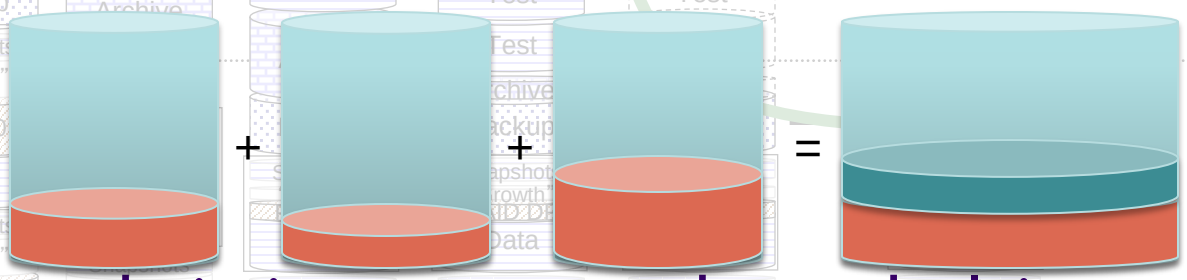
- NOT just wholesale copies of the data
  - We call those point-in-time (PIT) copies
- Data sharing
  - Form of deduplication
  - Data in snapshot shared with live data until live data is written
- Typical uses
  - Restore accidentally deleted/mangled files
  - Staging of backups to archive storage
  - 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

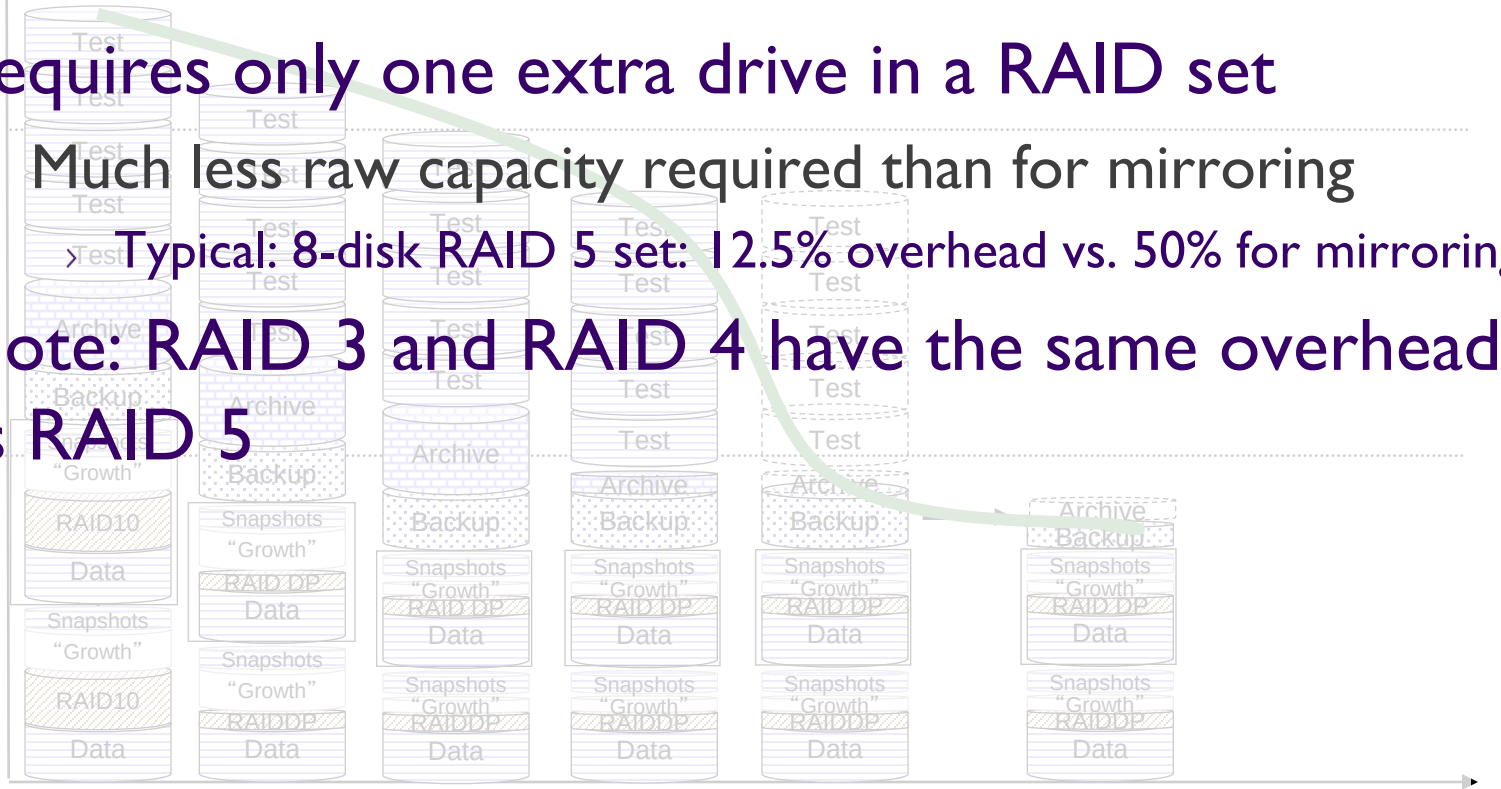
Quotas	Thin provisioned volume
Quota limit	Nominal volume size
Hit quota limit: writes fail	Fill up volume: writes fail
Increase quota	Increase volume size
Need more physical storage	Need more physical storage



- Storage admin just manages the underlying storage

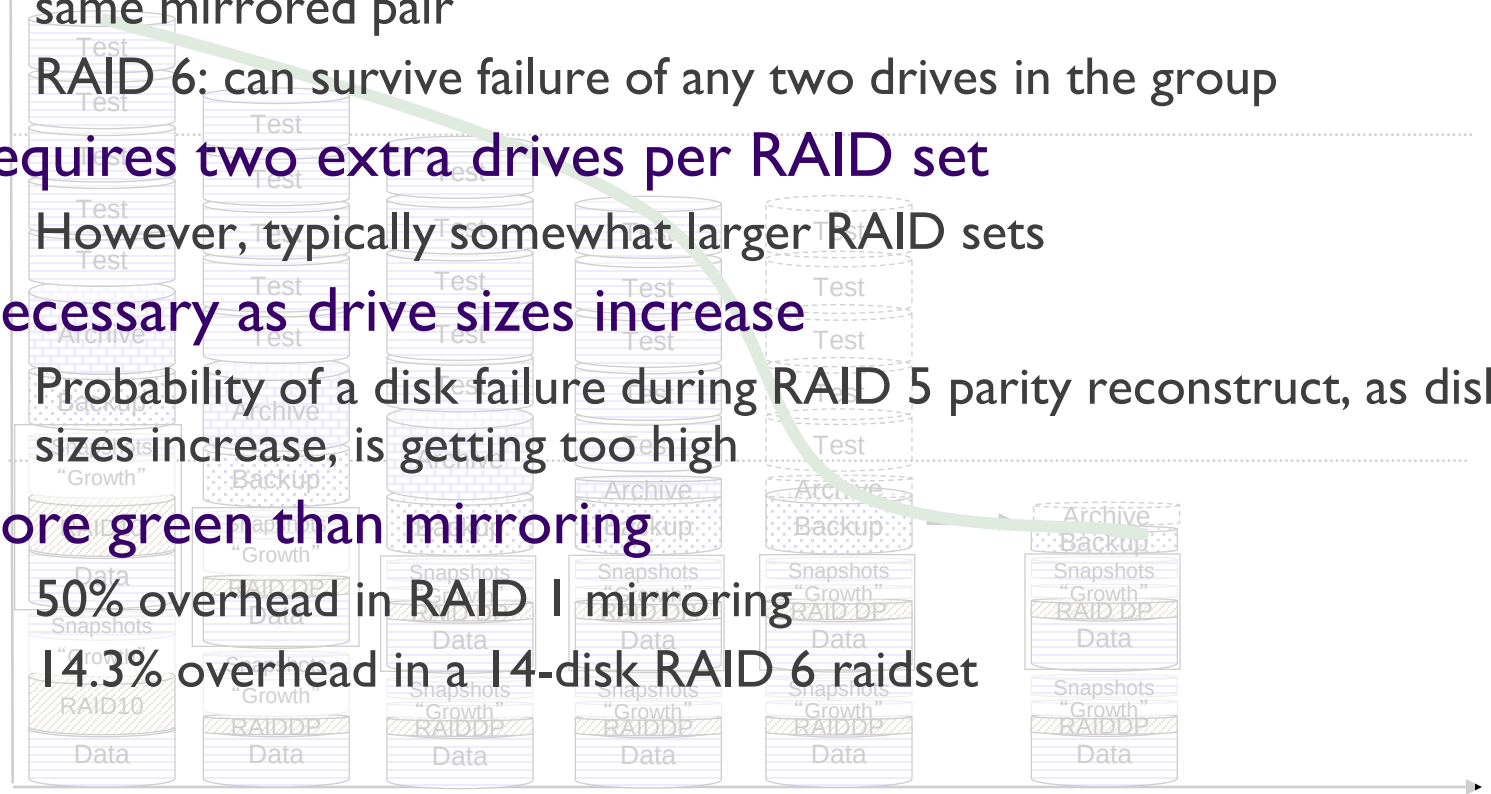
RAID 5

- 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



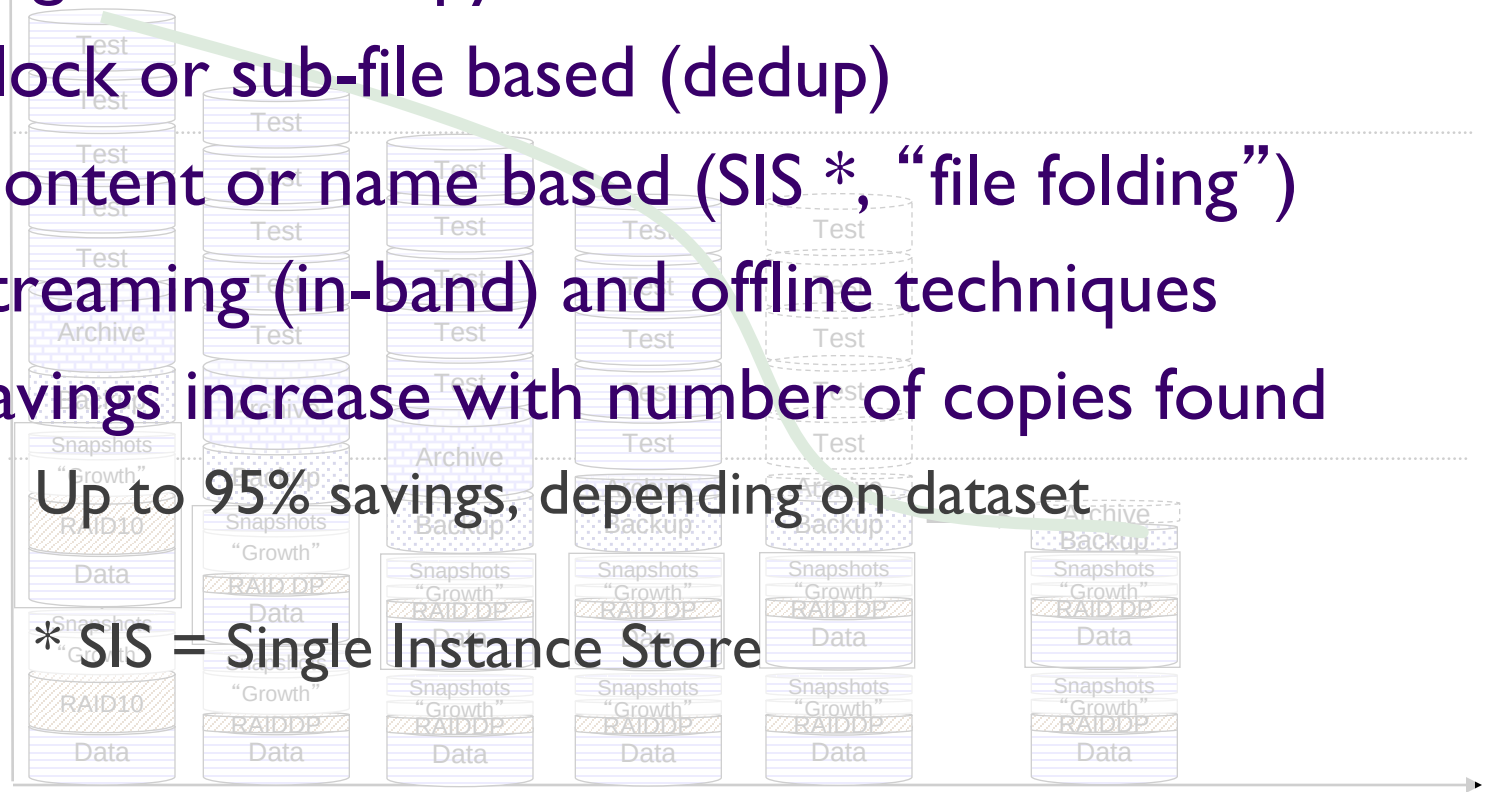
RAID 6

- 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, as disk sizes increase, is getting too high
- More green than mirroring
  - 50% overhead in RAID 1 mirroring
  - 14.3% overhead in a 14-disk RAID 6 raidset



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")
- Streaming (in-band) and offline techniques
- Savings increase with number of copies found
  - Up to 95% savings, depending on dataset



* SIS = Single Instance Store

Other technologies

- Capacity vs. high performance drives
- ILM / HSMHSM / Tiering
- “flash and stash”
- MAID
- SSDs
- Power supply and fan efficiencies

Capacity vs. High Performance

- Capacity
  - focused on GB/watt at rest
    - 1 TB SATA: 15W
    - 4 x 250 GB SAS: 64W
  - also tend to have better $/GB
    - Areal density vs. reliability tradeoff
  - NOTE: power use is 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 FC: 3 – 4 ms
    - SSD: .01 ms read 10-15 ms write
  - also designed for higher RAS * environments

* RAS = Reliability, Availability, Security

ILM / HSM / 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 automation
    - Cost of administration
    - Cost of data movement
- Practice
  - Storage declines in value as it ages
  - Manual movement of data sets
- New wrinkle
  - Footprint considerations preclude keeping older gear

Check out SNIA Tutorial:

What's Old is New Again – Storage Tiering

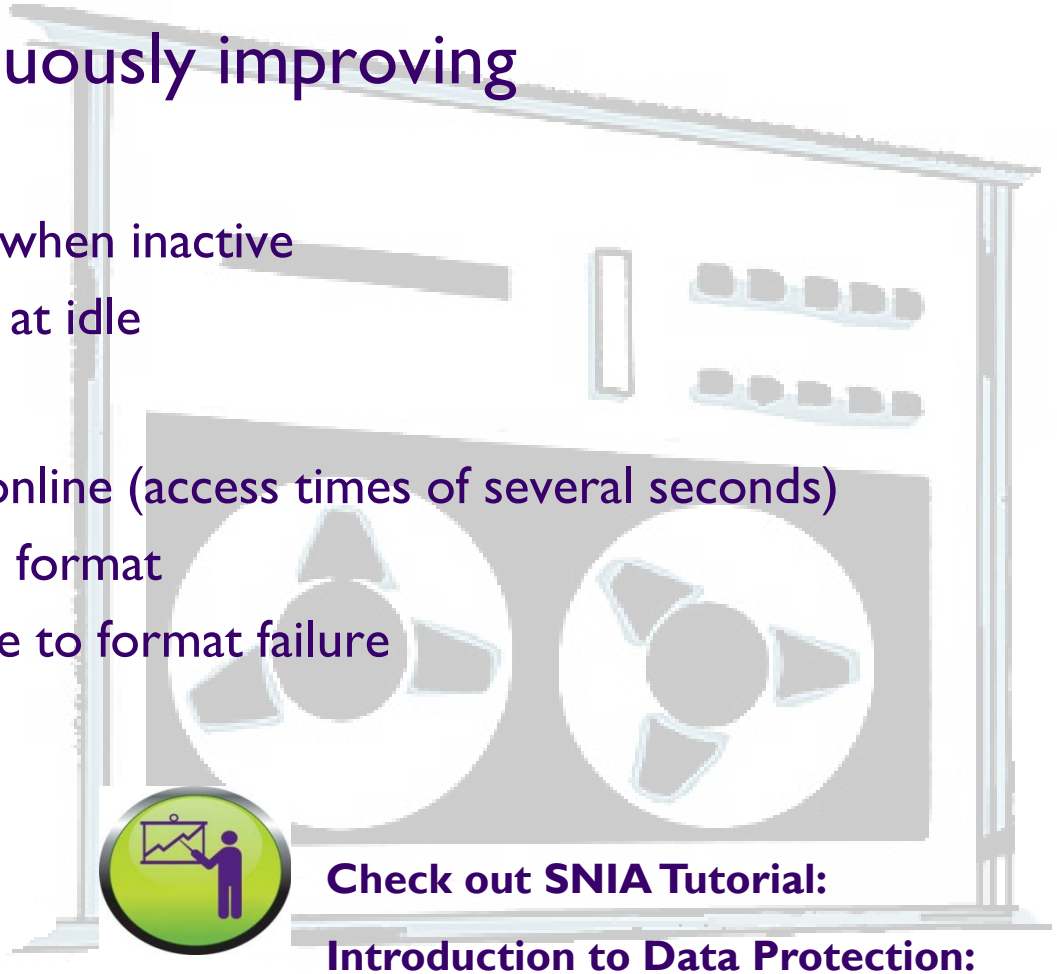
* ILM = Information Lifecycle Management
HSM = Hierarchical Storage Management

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

- Technology is continuously improving
  - 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



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
      Oops, FAIL!
      Recent data shows much background housekeeping activity
    - No access time penalty when idle (cf. MAID)
    - No need to keep some disks spinning (cf. MAID)
  - Cons
    - WRITE performance < 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

Power supply and fan efficiencies

- Efficiency of power supply an up front waste
  - Formerly 60-70%
  - Nowadays 80-95%
    - Climate Savers
    - 80+ group
    - Note: Efficient PSs *are* more expensive
- Variable speed fans
  - Common nowadays
  - Software (OS) control
  - Energy consumption is quadratic in the rotational speed

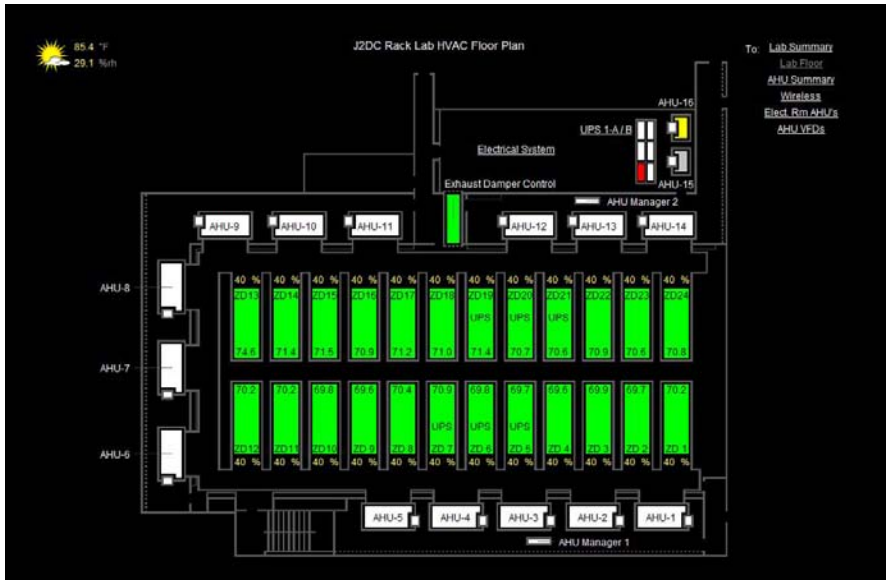
Facilities-side technologies

- Monitoring
- Hot aisle – cold aisle technologies
- Water and natural cooling
- Flywheel UPSs
- Bypass UPSs
- DC Power

UPS = Uninterruptible Power Supply

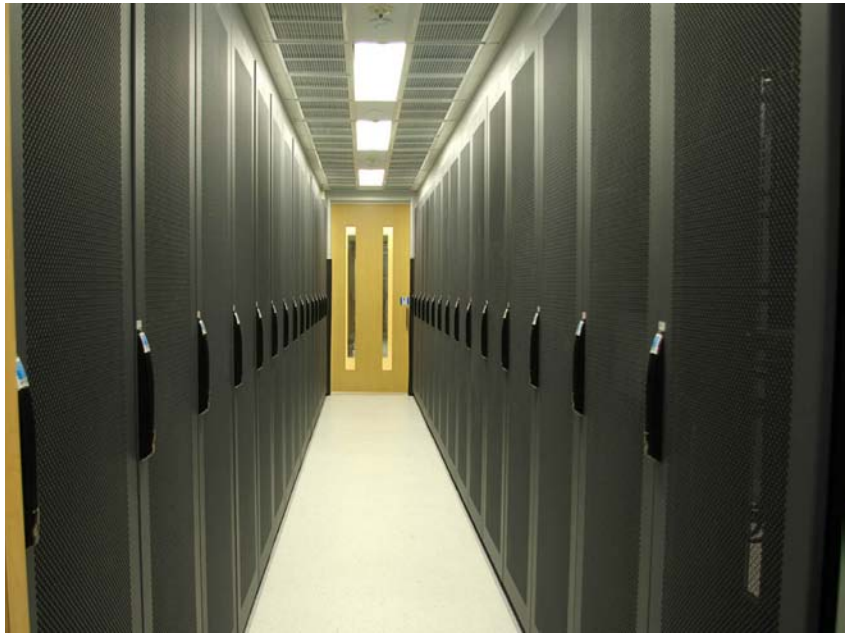
Monitoring

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



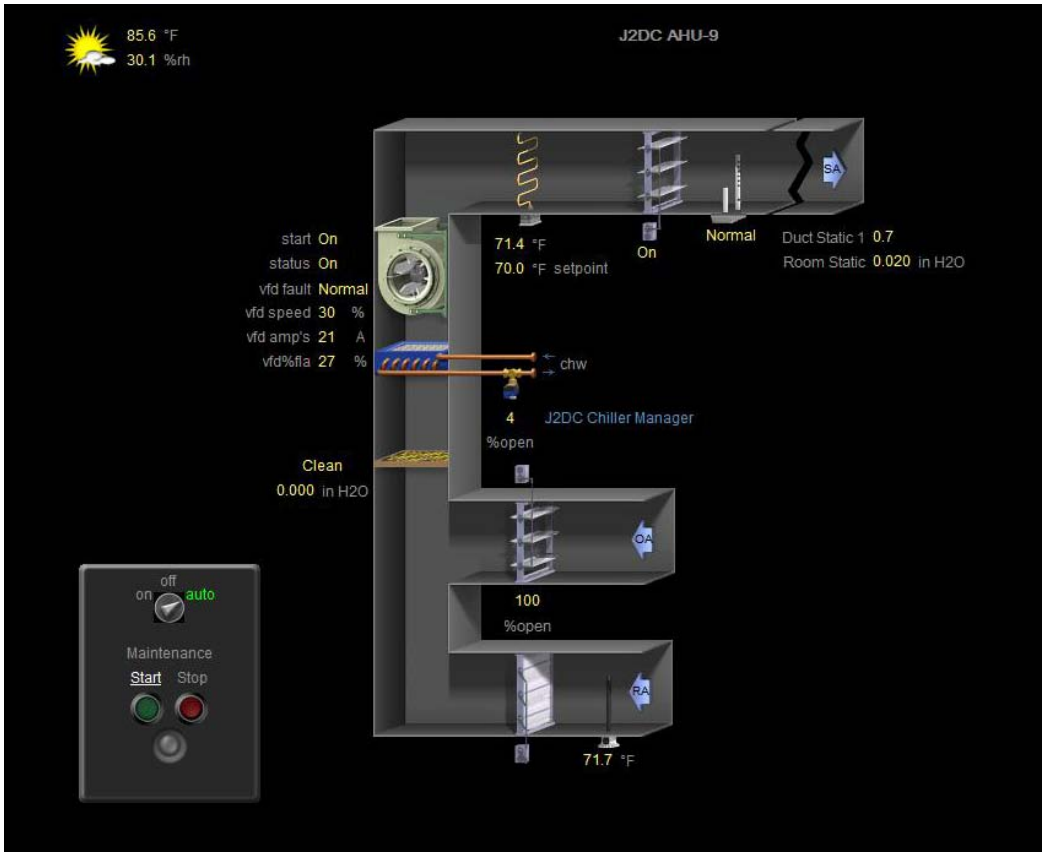
Hot aisle / cold aisle technologies

- Segregate airflows into hot and cold aisles (backs and fronts of servers)
  - More precise control
  - Allows higher input temperatures
  - Several emerging approaches
    - Hot air plenum
    - Hot air containment
    - Complete separation



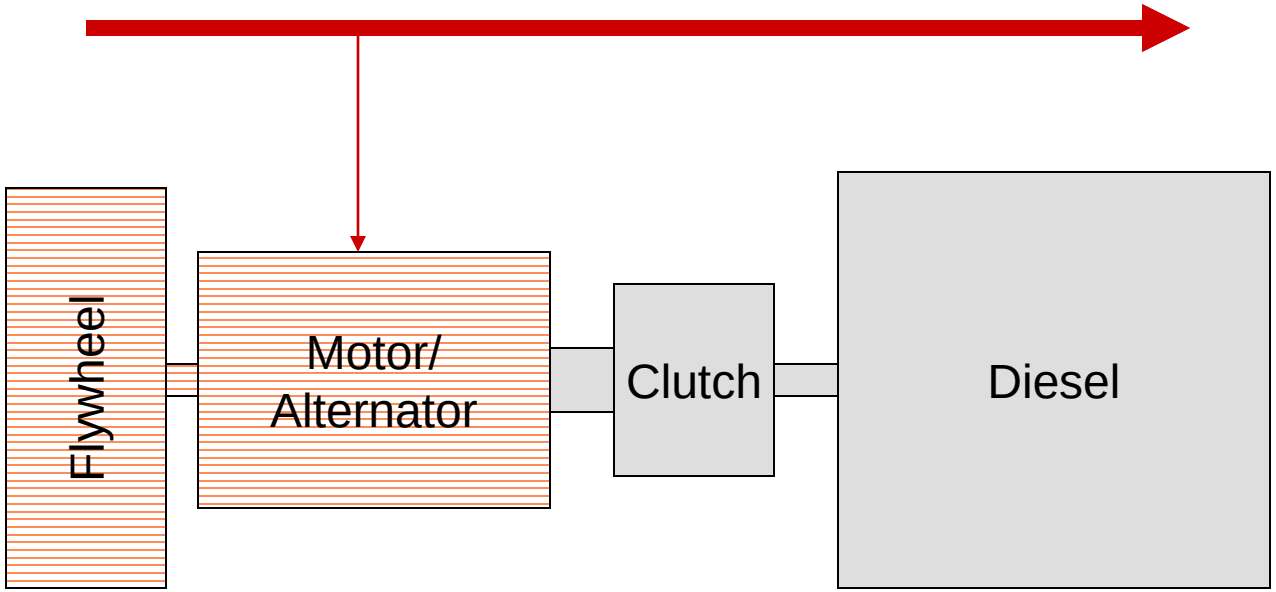
Water and natural cooling

- Water cooling of increasing interest
  - Issues include corrosion, warm water disposal
- Natural cooling
  - "Economizers"
  - Exhaust hot aisle air when temp > outdoor air
  - Use outdoor air when temp < cold aisle air
  - May need humidity + pollution control



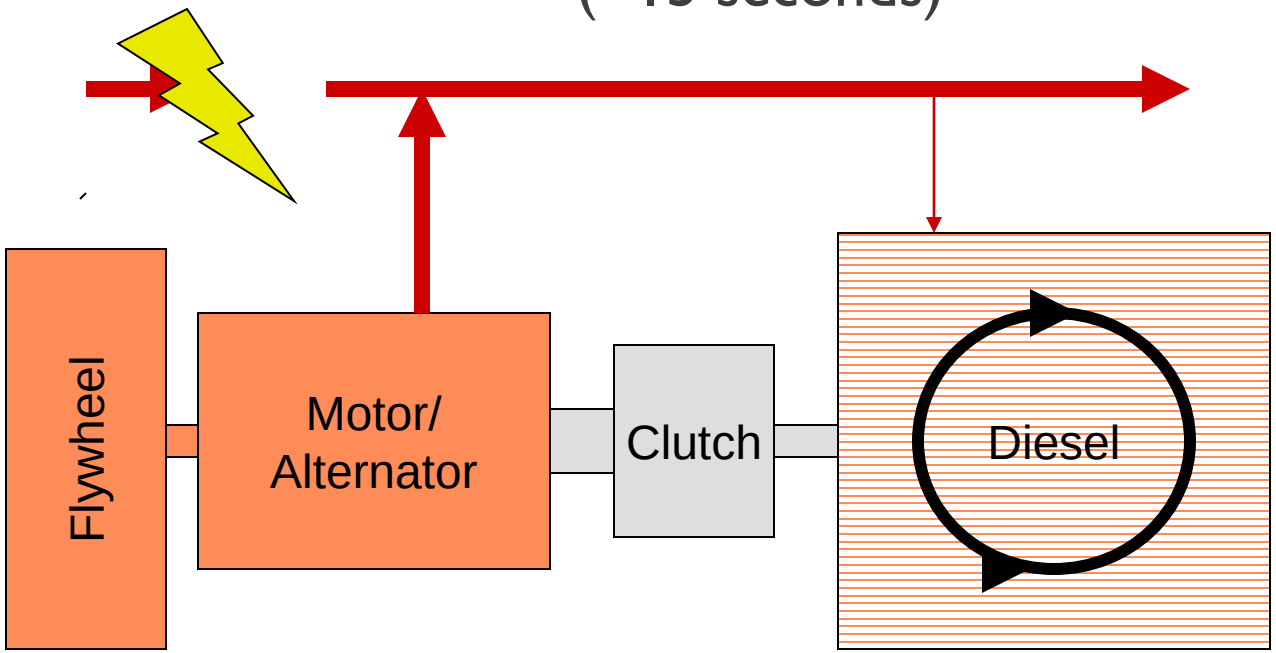
Flywheel UPSs

- Operational flow: normal condition
  - Grid power keeps flywheel turning



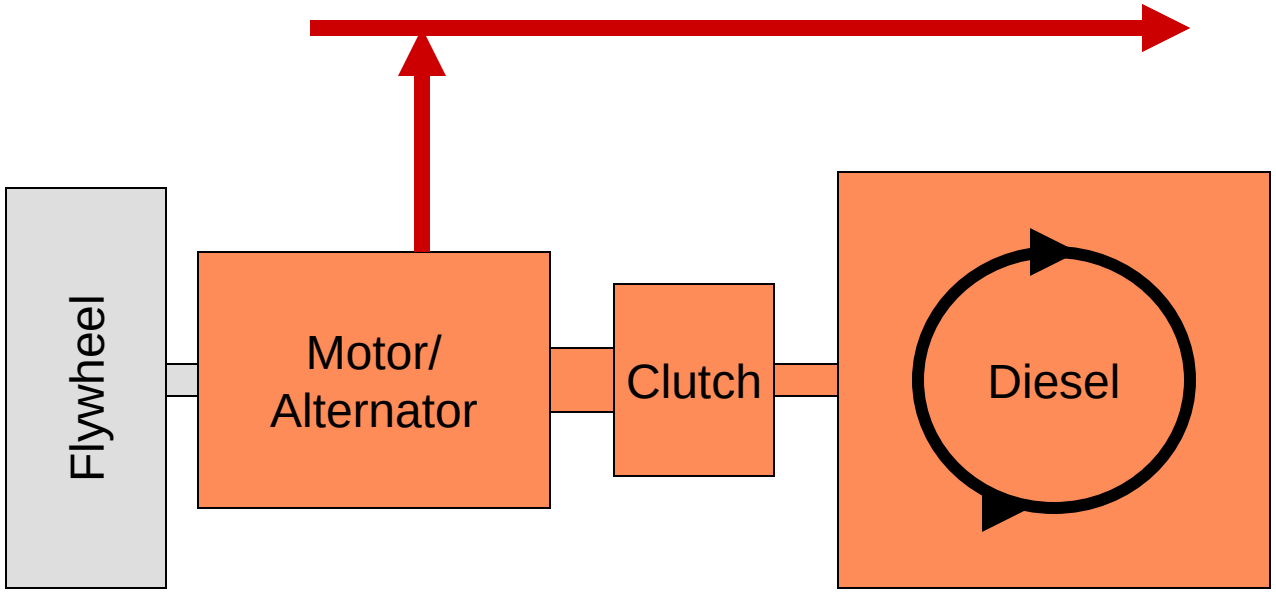
Flywheel UPSs

- Operational flow: outage
  - Flywheel powers alternator while diesel starts up (~15 seconds)



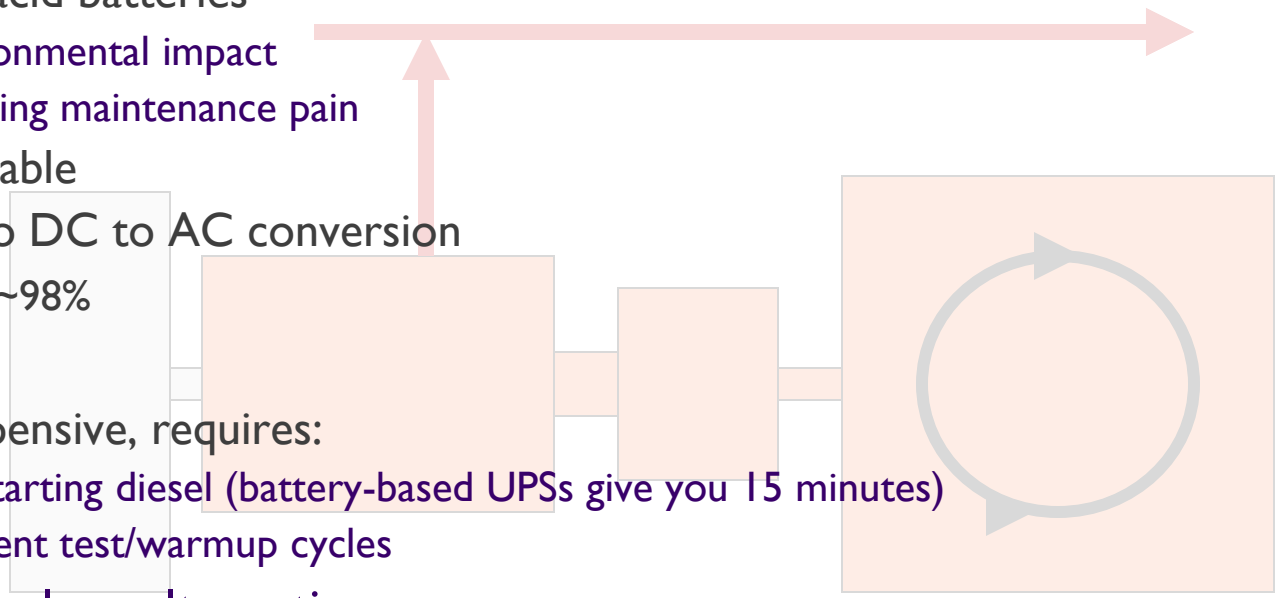
Flywheel UPSs

- Operational flow: diesel power
  - special clutch transfers mechanical linkage from flywheel to diesel



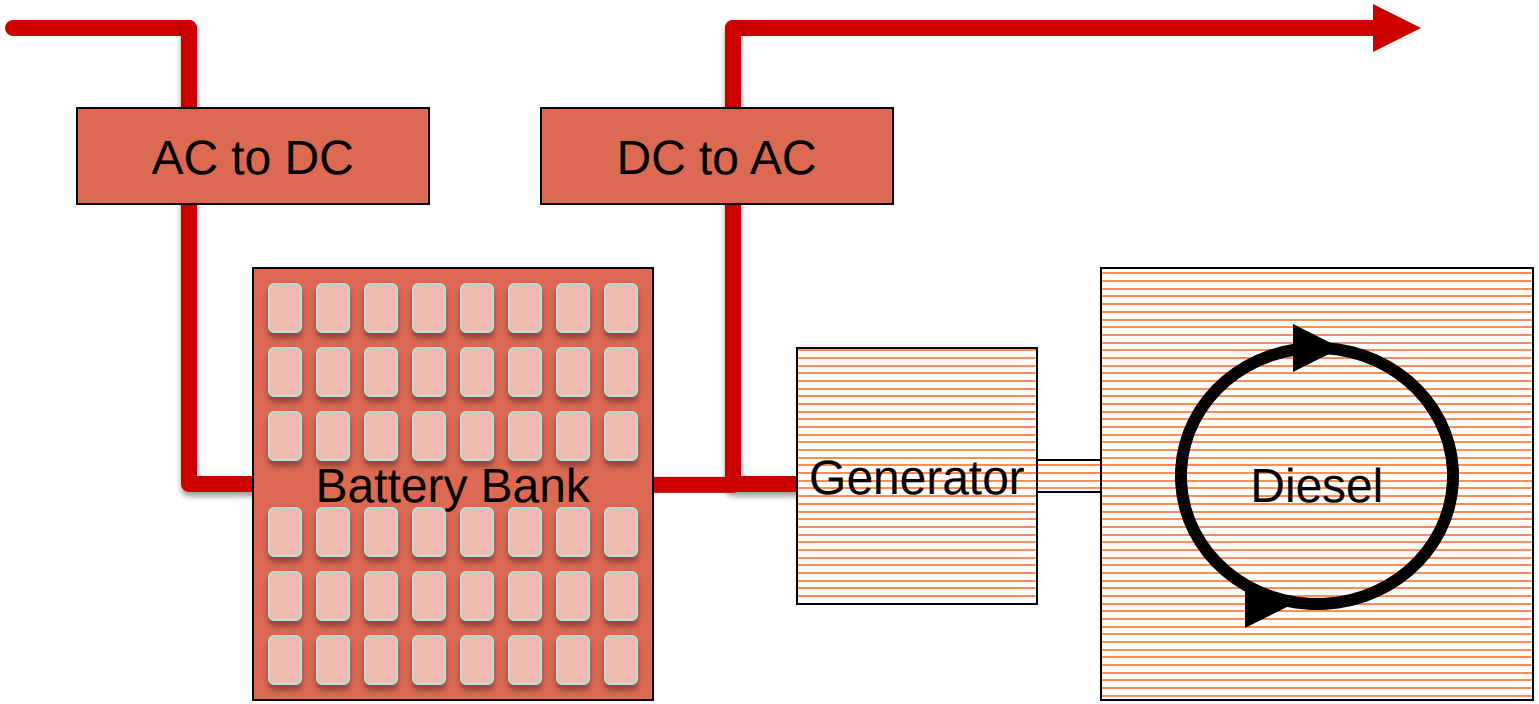
Flywheel UPSs

- Pros
  - No lead-acid batteries
    - Environmental impact
    - Ongoing maintenance pain
  - More reliable
  - No AC to DC to AC conversion
  - Efficient: ~98%
- Cons
  - More expensive, requires:
    - fast-starting diesel (battery-based UPSs give you 15 minutes)
    - frequent test/warmup cycles
- Similar technology alternatives
  - CAES (compressed air "flywheels")



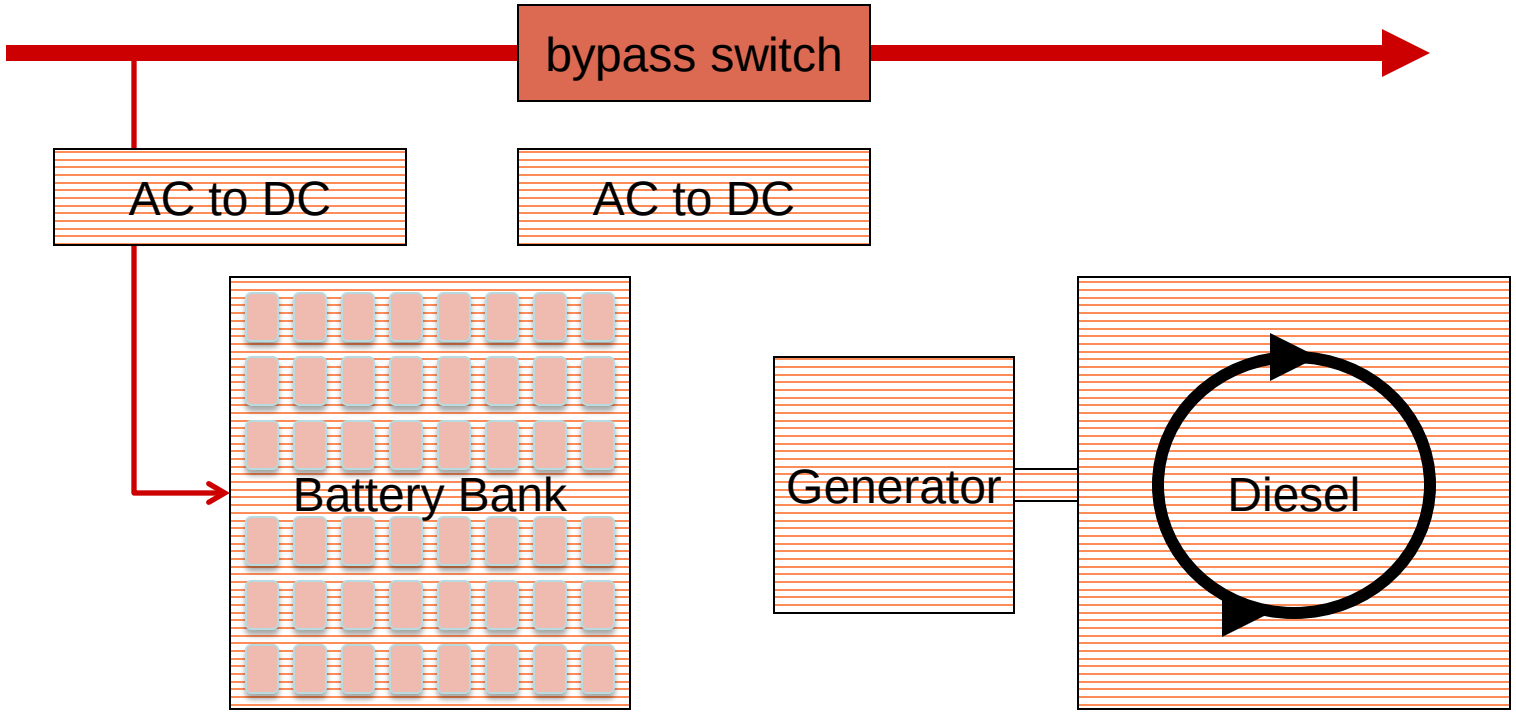
Old-fashioned UPSs

- Normal power flows through UPS
  - Significant conversion losses -- 85% - 94% efficient
  - Power conditioning as a side benefit



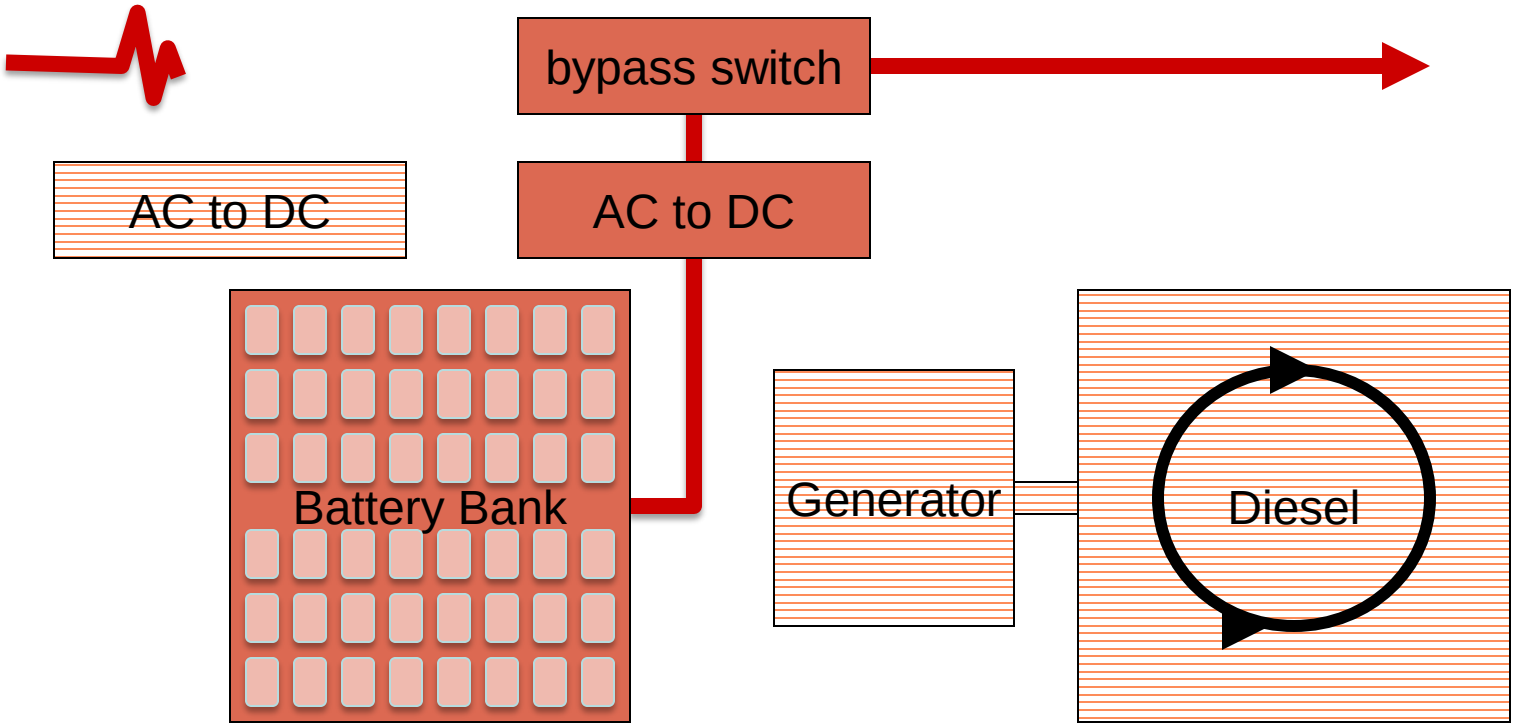
Bypass UPSs ("ECO mode")

- Normal power flows directly from grid
  - No conditioning, but also no conversion losses
  - ~98-99% efficient



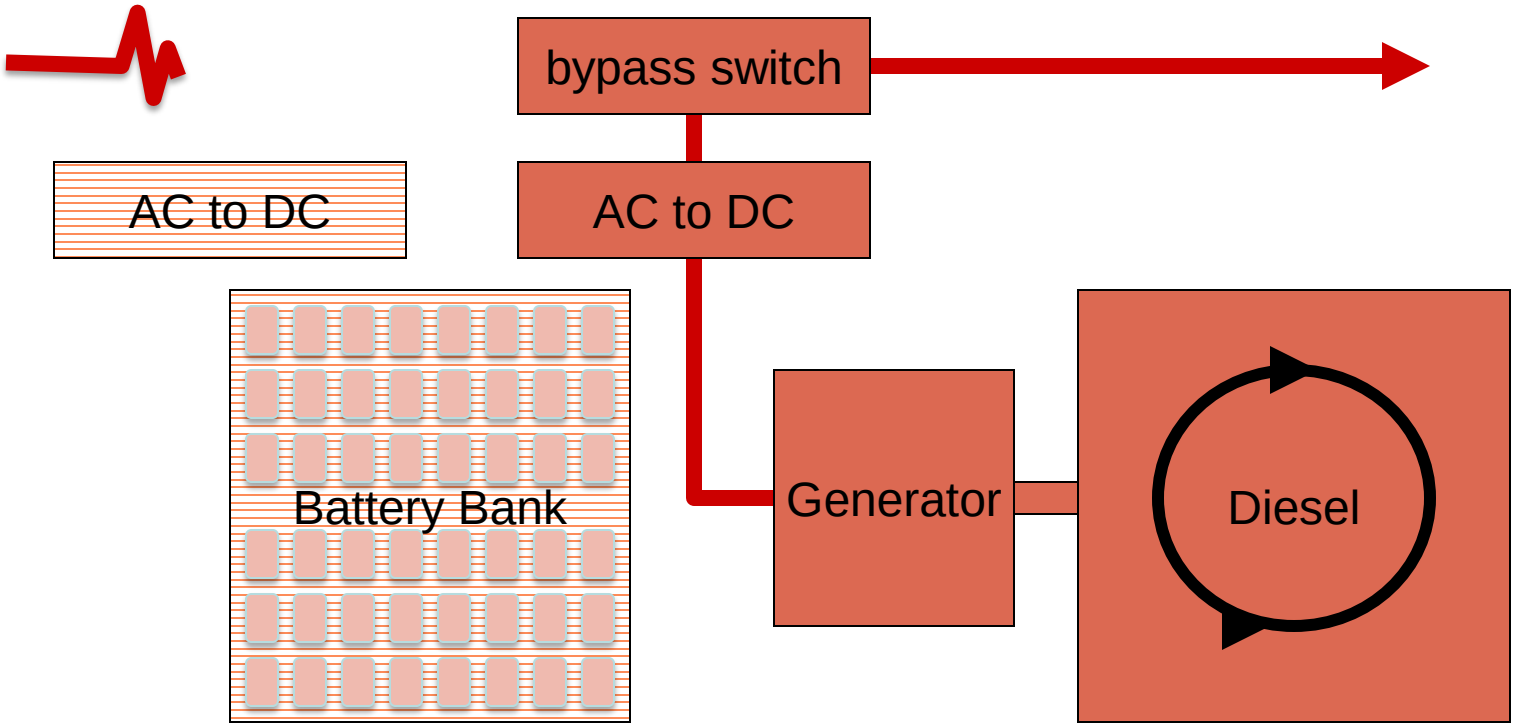
Bypass UPSs

- Static transfer switch activates upon fault (< 20ms)
  - IT equipment needs to provide ~20ms "holdup time"
    - a.k.a. "ride-through time"



Bypass UPSs

- Backup power as usual



DC power

- 380v DC (study by Lawrence Berkeley Labs/DOE)
  - Avoid AC/DC and DC/AC conversion losses at UPS
  - Avoid equipment power supply conversion losses
  - ~5% improvement over best-in-class AC systems
  - ~25% improvement over typical data center equipment

  - Caveat: "best-in-class" used traditional UPS
    - Tough to improve on 98-99%
    - Savings left to be gotten are mostly in the equipment

Savings calculations

- Equipment power savings
- Facilities power savings
- Space savings

Equipment power savings

- Server virtualization
  - up to 80% savings
  - much depends on load
- Power supply efficiencies
  - 80+/EPRI "bronze" already above 80% efficient
  - push your vendors for silver/gold/platinum (especially in servers)
- Variable speed fans
  - up to 80% savings
  - power consumption quadratic in rotational speed
  - interesting interaction with data center cold aisle temperatures
    - note possibility of gaming PUE by using equipment fans to exhaust hot air

Facilities 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
  - Be aware – ASHRAE is pushing facilities temperature limits upward
    - WAAAAAY up
- What you need to do
  - Maintain those blanking plates
  - Specify rack-level power and temperature monitoring for storage equipment

Savings calculations (storage)

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

Typical savings

- Compression
  - 15 – 30%
  - Remember, no savings from already compressed file formats
- Delta Snapshots
  - 80 – 95% per snapshot, compared to raw PIT copies
  - Only data written since snapshot needs to be copied

Typical savings

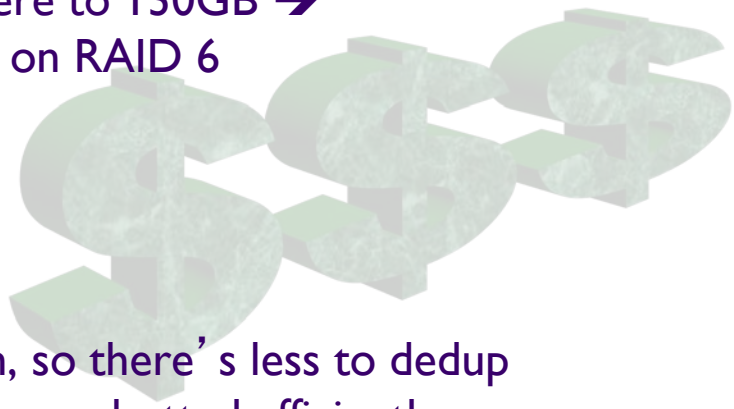
- Thin provisioning
  - 40 - 60%
  - Average 30% utilization ➔ over 80% utilization
- RAID 6
  - 35%
  - For 14-disk RAID 6 set, compared to RAID 1/10
- Deduplication
  - 40 – 95%, depending on dataset and time interval
  - ~ 40 – 50% average over time

Caveats

- Savings estimates are real, but best taken as anecdotal
  - YMMV – your mileage may vary
  - Make your vendors prove their claims
- 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 SAS – 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

Are these savings multiplicative?

- Sometimes yes
  - RAID 6 + writeable clone – 1000GB
    - **2000GB** needed for a raw writeable copy on RAID 1 storage
    - 90% writeable clone savings takes us to 200GB
    - 35% RAID 6 savings takes us from there to 130GB ➔
    - **130GB** needed for a writeable clone on RAID 6
- Sometimes not so much
  - Snapshot + deduplication
    - Can't dedup readonly snapshots
    - Snapshots are a form of deduplication, so there's less to dedup
    - OTOH, already deduped data can be snapshotted efficiently
- In general, diminishing returns with multiple technologies
  - Amdahl's Law in reverse



Takeaways

- Heavy hitters
  - Facilities improvements
    - hot aisle/cold aisle
  - Software technologies
    - especially thin provisioning, dedup
  - Tape
  - RAID 6
- Utilization
  - A practice, not a technology
  - But none of this means much if the storage is empty and idle

Q&A / Feedback

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

Check out our other SNIA Green Tutorials!

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

- SNIA Education Committee

SNIA GSI Marketing Committee
SNIA reviewers

Larry Freeman – NetApp