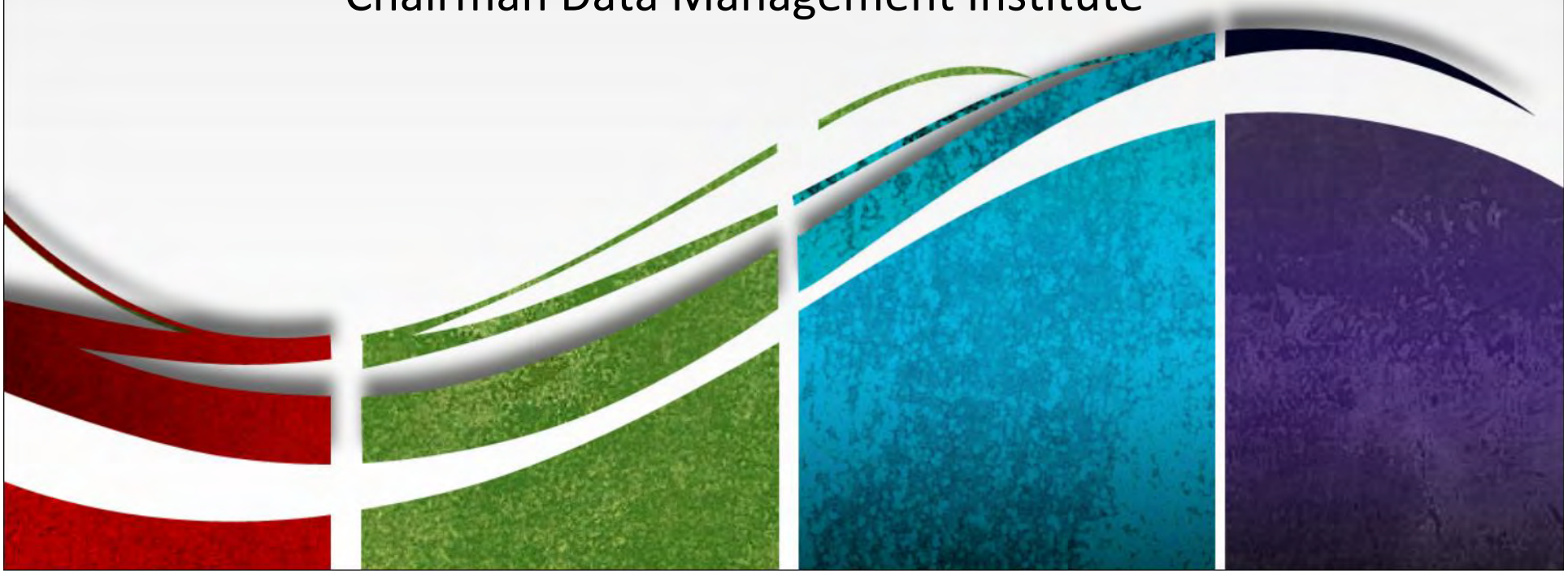
Building an Efficient Storage Operation

(and How to Cope with the Earworm Problem)

Jon Toigo
CEO Toigo Partners International
Chairman Data Management Institute



WARNING

WARNING

The presentation you are about to see contains references to "earworms" that may evoke memories of recursive or repeating lyrics, melodies or jingles that you may not be able to dislodge from your conscious mind.

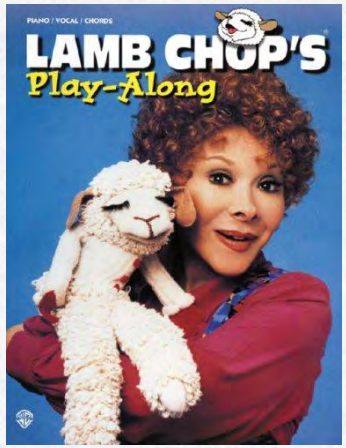
Management takes no responsibility for the discomfort – or, in rare cases, neuralgia, schizophrenia or insanity – that may result.

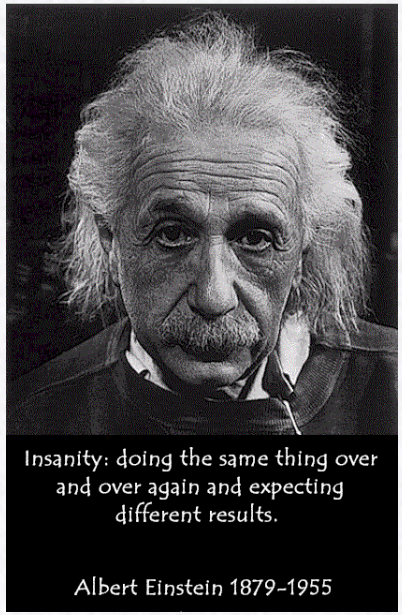
You have been warned.

EARWORMS

"The Song that Never Ends"

- Favorite of my kids in their early years from *Lamb Chop's Play-Along* on PBS in 1992
- It's cute at first – a single verse long, looping infinitely
- Once you hear it: it sticks in your head…forever… repeating and repeating
- If you have been in IT for three decades or more (like me), you may be hearing some tunes begin to repeat in recursive loops
- They are called "earworms"
- They drive me crazy.





EARWORMS

Here's One I Want to Get Off My Chest at the Outset

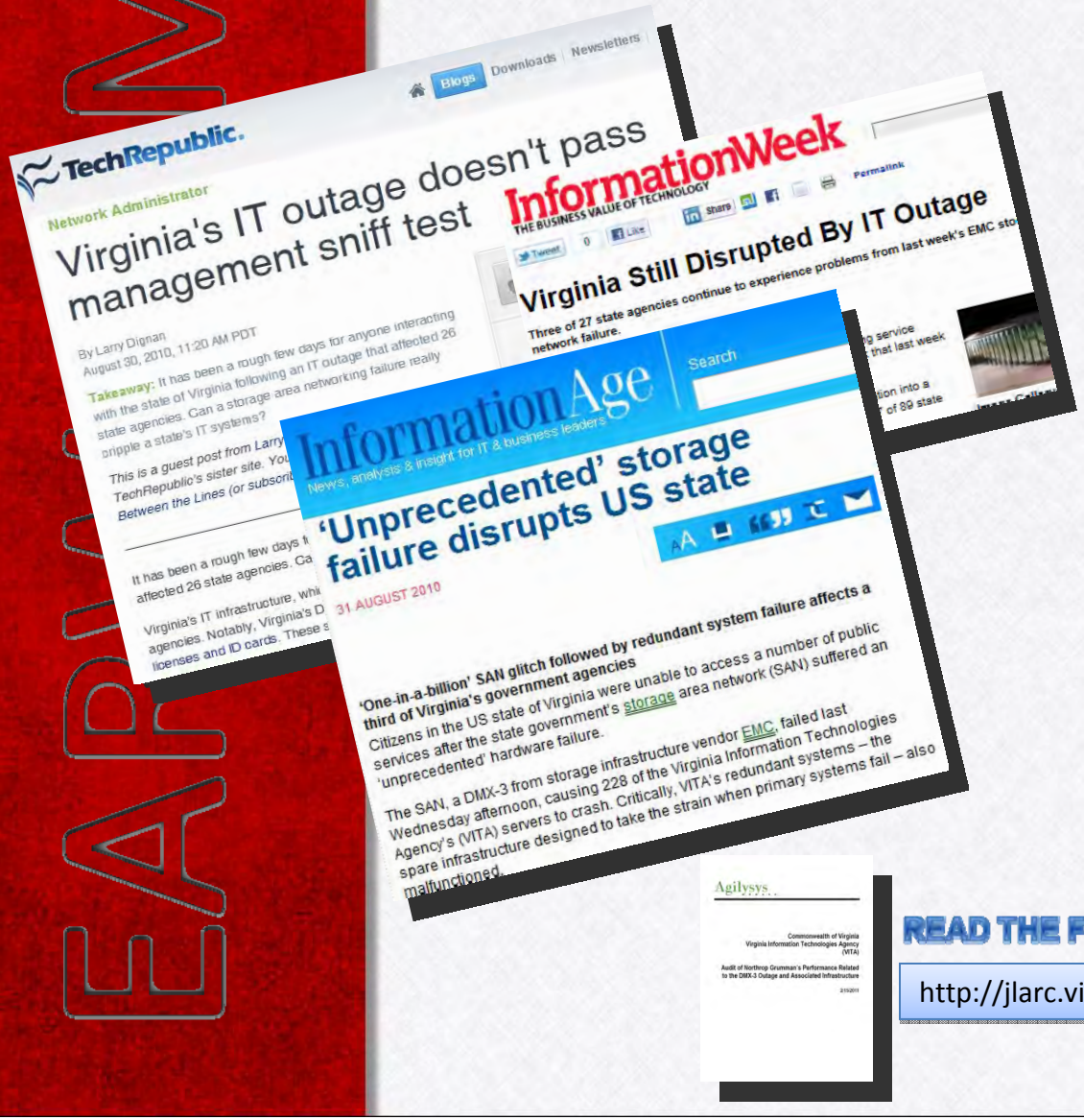
- "Tape is dead."
- Gartner was paid big bucks to reach this conclusion in the late 1990s…by EMC.
- Now, with the acquisition of Data Domain (a vendor of de-duplicating Virtual Tape Libraries), the messaging is back in full throat.

EARWORMS

I Saw It On the Back of a Car at a Recent Storage Conference

- Honestly, the owner may have been using it to hold the paint together…

Apparently, the Driver Wasn't Up on Current Events…



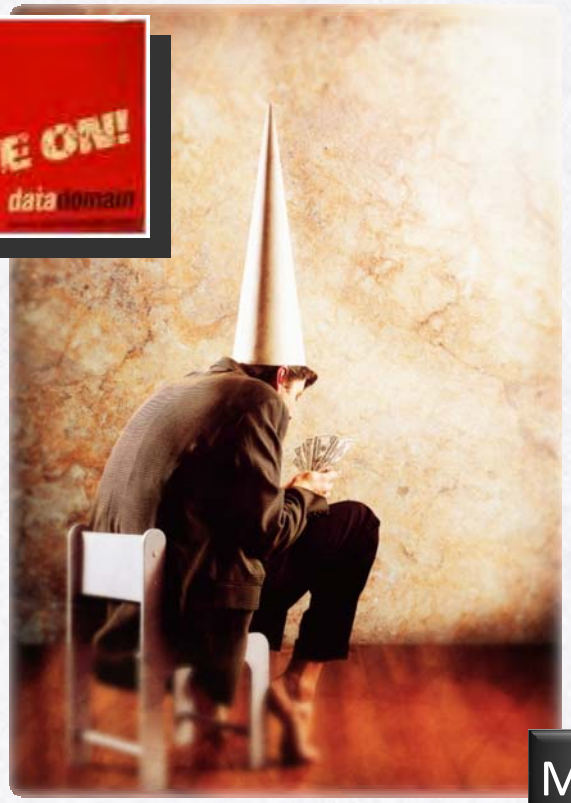
- EMC array and fabric failure, compounded by EMC repair service error, crashes 228 servers. Denies services to 24 of 27 agencies and 45,000 citizens.
- All primary systems disk backups fail. SRDF/PIT Data corrupted.
- **Data could only be restored from tape. TGFB!**

READ THE FULL AUDIT REPORT HERE:

http://jlarc.virginia.gov/other/Nortrop%20Grumman%20Audit.pdf

EARWORMS

“Move on.” *Indeed…*



Moving on…

EARWORMS

Psychiatric researchers say that the only way to fight an Earworm…

- Is with another earworm.
- So, here's my new bumper sticker…

EFFICIENCY

Other Memes of 2011

- "Do more with less" – The cost- containment meme
- "Virtualize it!" – The CAPEX/OPEX reduction meme
- "OMG! Data is exploding!" – The BIG DATA meme
- "Cloud computing" – The outsourcing meme
- "Stop those Wikileaks" – The data protection meme



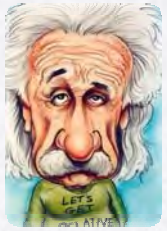

All of Which Require

Storage Efficiency

What is Efficient Storage?

(and who's saying my storage isn't efficient?)

- Three common interpretations …



The ratio of the output to the input of any system (Engineering Perspective)

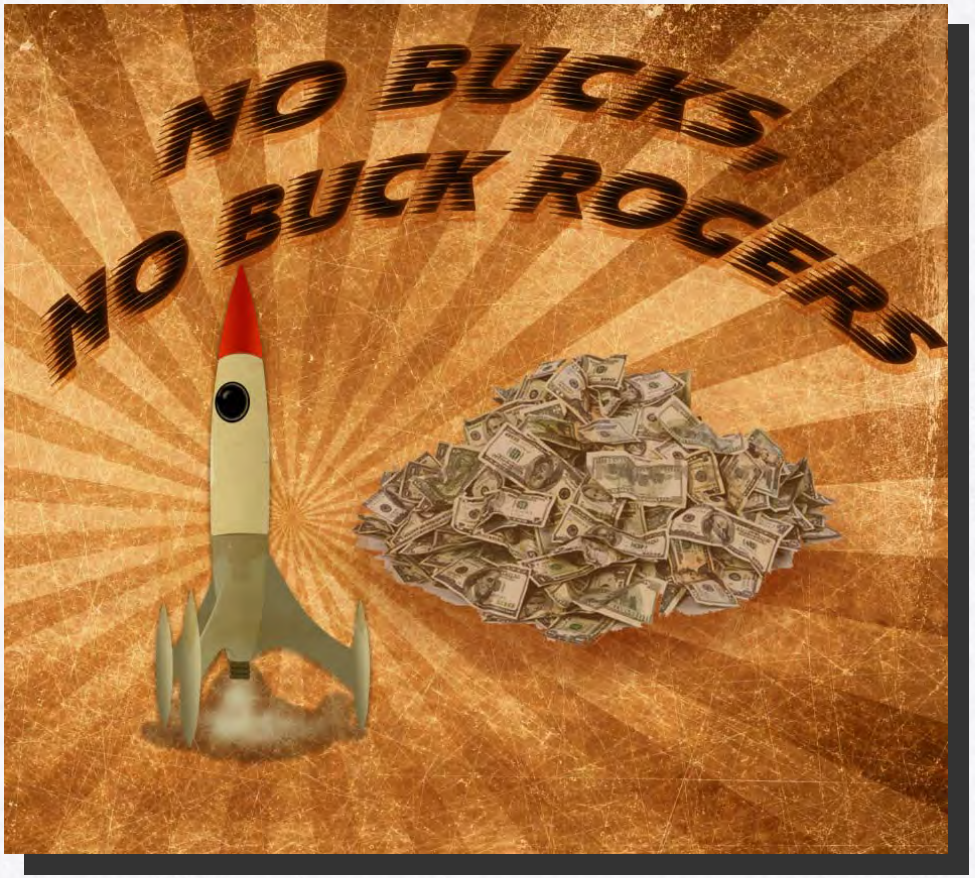
Comparison of what is actually produced or performed with what can be achieved with the same consumption of resources; "productivity" (Business Perspective)

Skillfulness in avoiding wasted time and effort (Operational Perspective)

EFFICIENCY

Why?

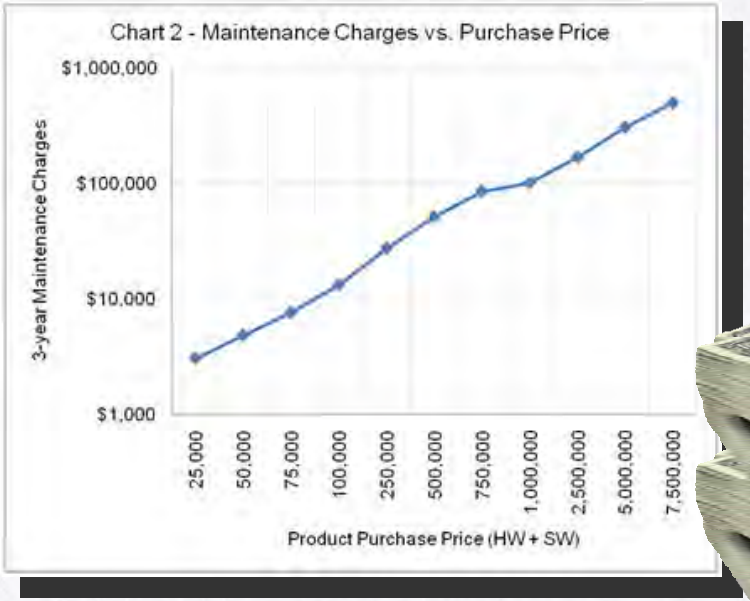
- Simple. They control the budget.



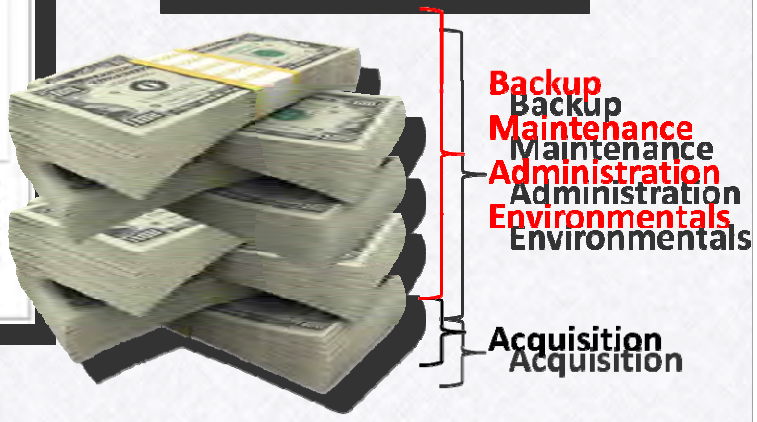
EFFICIENCY

They See Storage Infrastructure as a Set of Investments with Costs and Internal Rates of Return

- The key facts:
  - Storage hardware costs
  - Storage operations costs



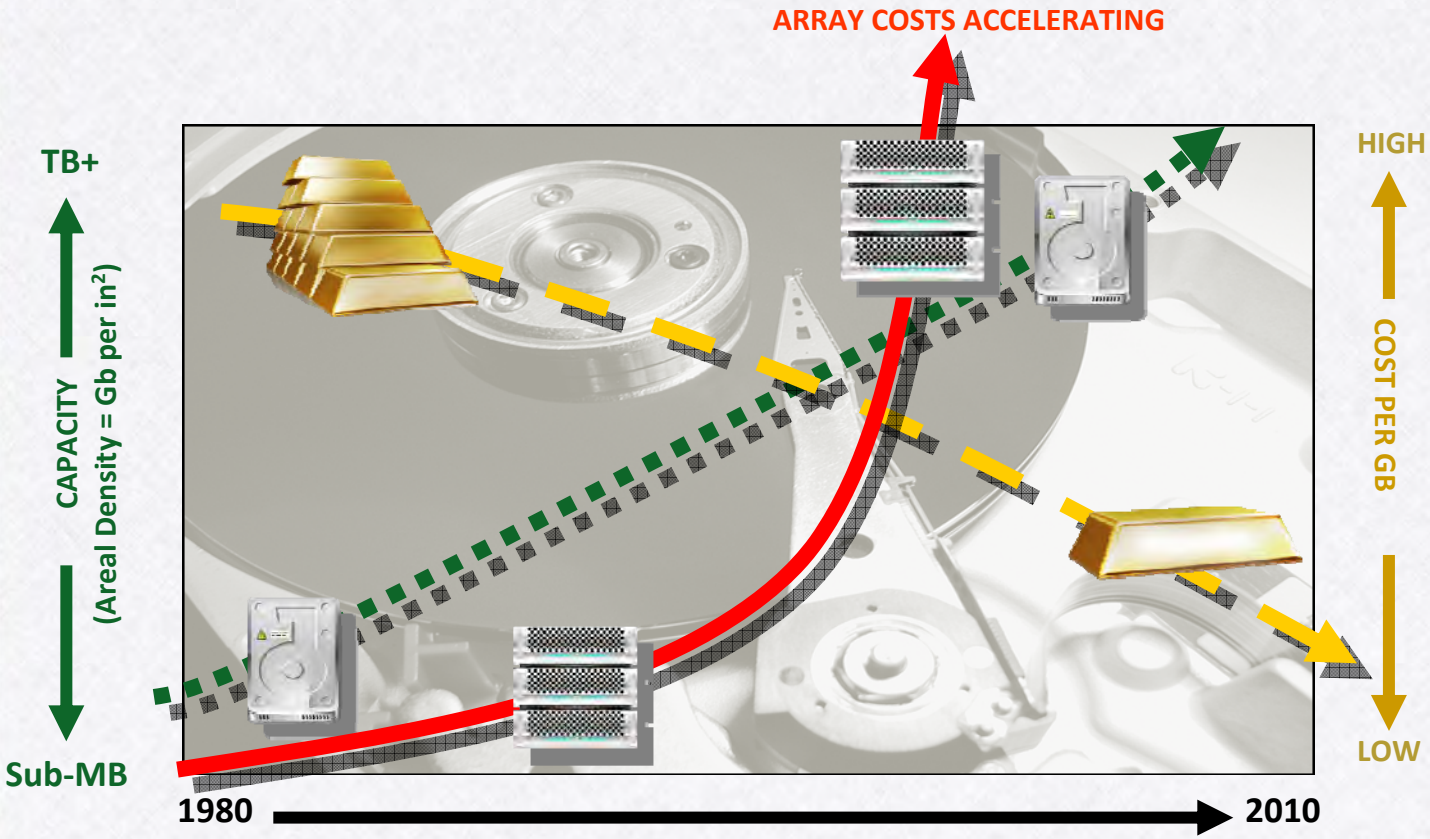
Annualized cost of ownership is ***5-8x*** acquisition



Storage seems like an *inefficient* investment!

EFFICIENCY

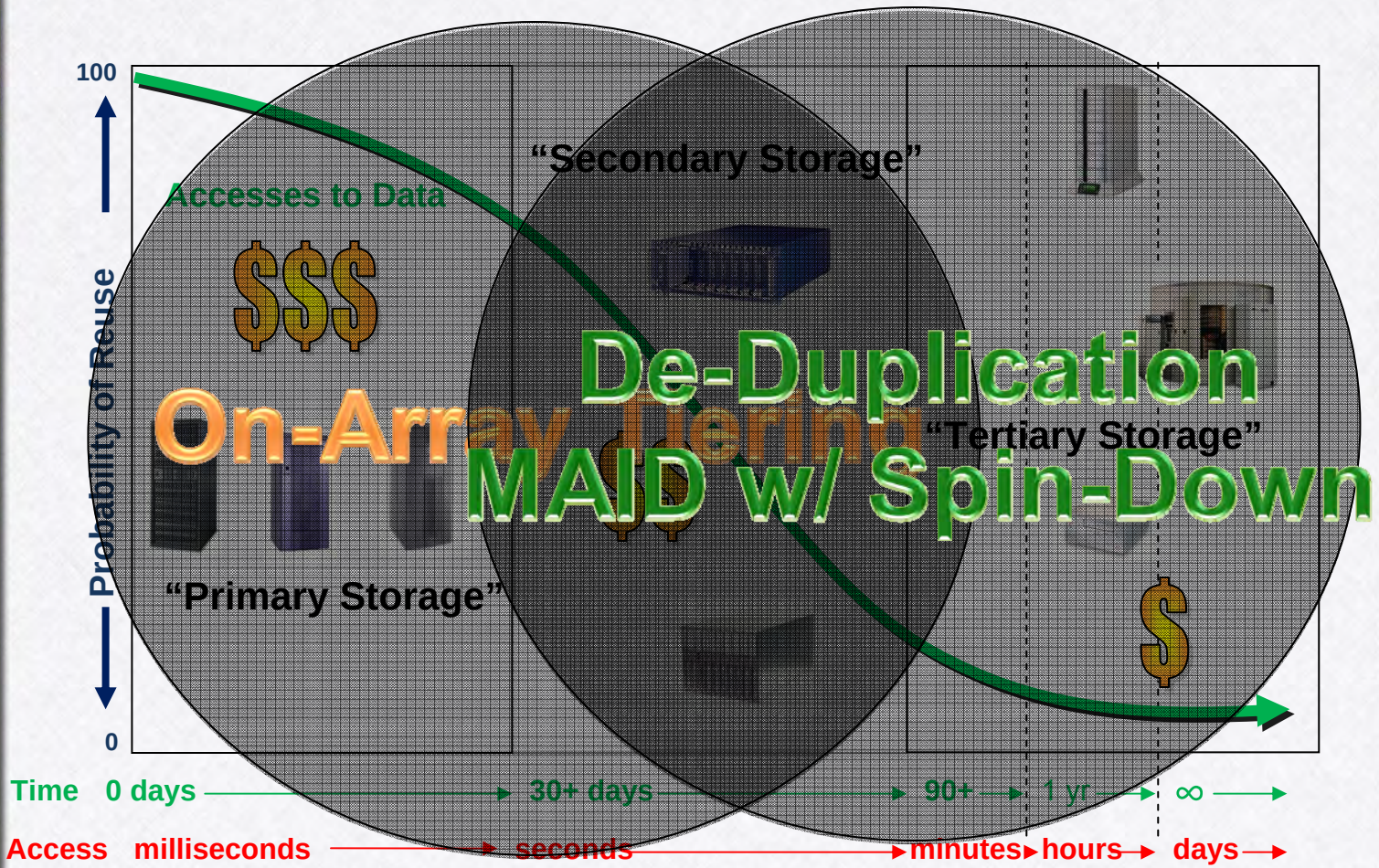
Perhaps, Because It Is

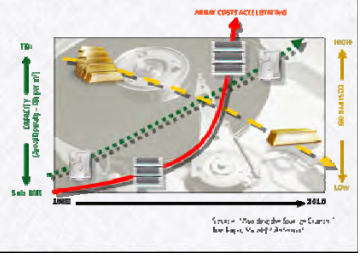


Source: "Avoiding the Storage Crunch,"
Jon Toigo, *Scientific American*

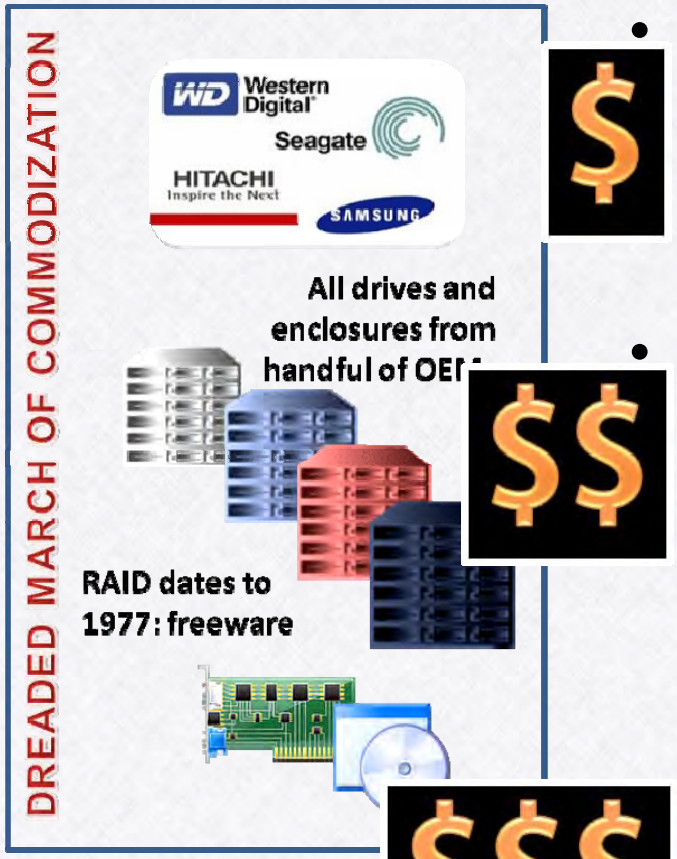
EFFICIENCY

Why? *Creeping Featur-itus*

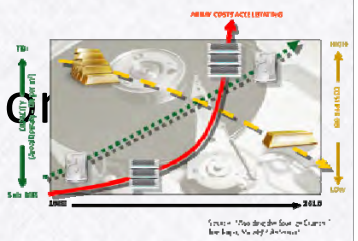




The Disk Array Vendors Call This "Smart" Storage

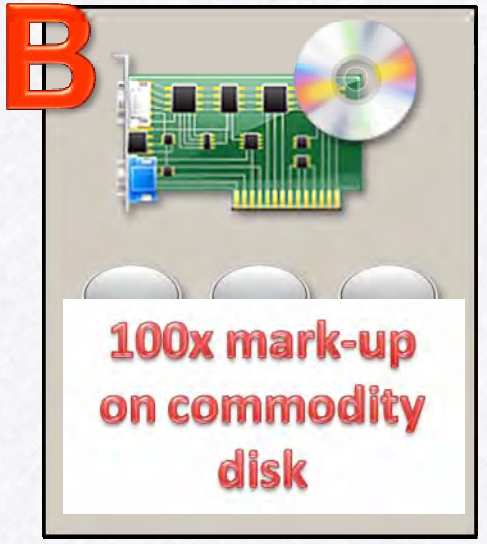


- Commodity components "customized" for performance and reliability
- Controllers with "value add" functionality that automates capacity, performance, and data protection management so that more storage can be managed by fewer people
- One stop shop, c
to choke



Case in Point

- The de-duplicating VTL
  - A. Software only
    - Licenses software separately for $2500 per TB
    - Run on a server, use any storage you have
  - B. Software plus hardware
    - 32 TB of commodity SATA disk storage with embedded software = $410,000 MSRP (or $12,812.50 per TB)
    - 1 TB SATA disk costs $79 - $140 on NewEgg.com (~$3200 for 32 TB)
    - Vendor's "value-add" rig costs $80K for software, $330K for disks or ***100x the cost of the disks themselves***



Vendor argues that the cost is justified…of course.

- You get "70:1 reduction ratios" with their rig, so…

*"**Phenomenal cosmic powers.***
Itty, bitty living space."

Genie from *Aladdin*

(***We like fairy tales, too.***)

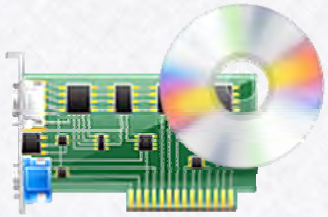
Another example: thin provisioning

- Thin provisioning: “virtualized” capacity oversubscription, but with “intelligent” demand forecasting
- Vendor claims better capacity allocation efficiency
- Starter rig:
  - 8.7TB (raw) mix of (6) 460GB 15k SAS and (6) 1TB 7200 SATA drives at $38.1K with single controller and value-add software for tiering and thin provisioning.

Q: On what planet are $2600 in drives worth $38.1K???

6 SAS Drives
~$350 each without OEM volume purchase discount ($2100)

6 SATA Drives
~$85 each without OEM volume purchase discount ($480)

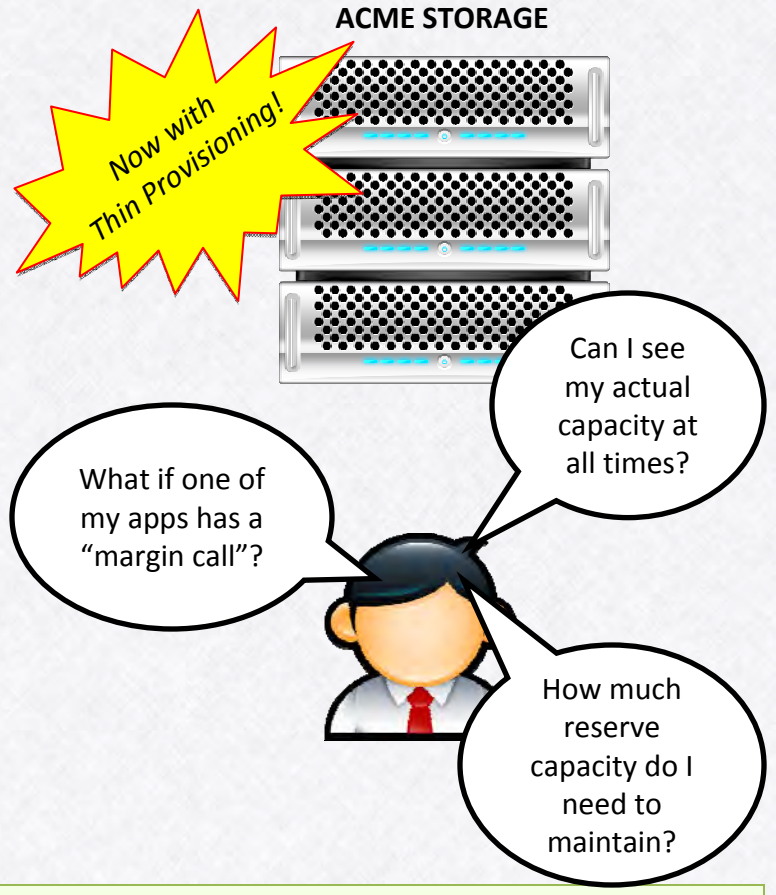


Head, chassis, software, margin = ~$35.5K???

EFFICIENCY

Again, the vendor argues that the cost is justified…

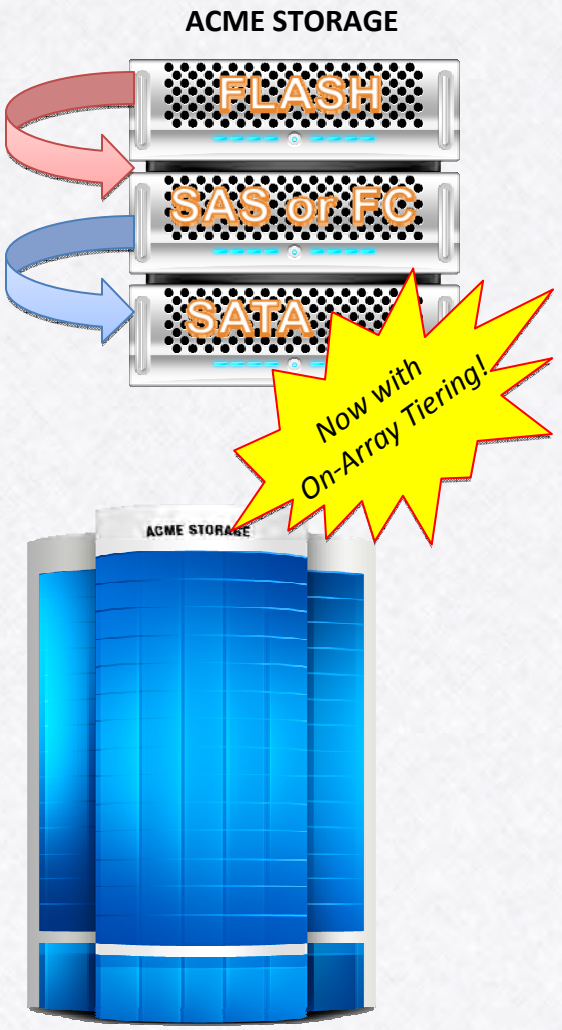
- "Thin provisioning reduces the labor cost associated with capacity provisioning"
- Better capacity management reduces downtime
- Enables "just-in-time" capacity purchasing



No one is arguing the value of thin provisioning, but valid questions need to be addressed regarding the wisdom of placing thin provisioning technology on the array controller, rather than operating it as a generalized service…

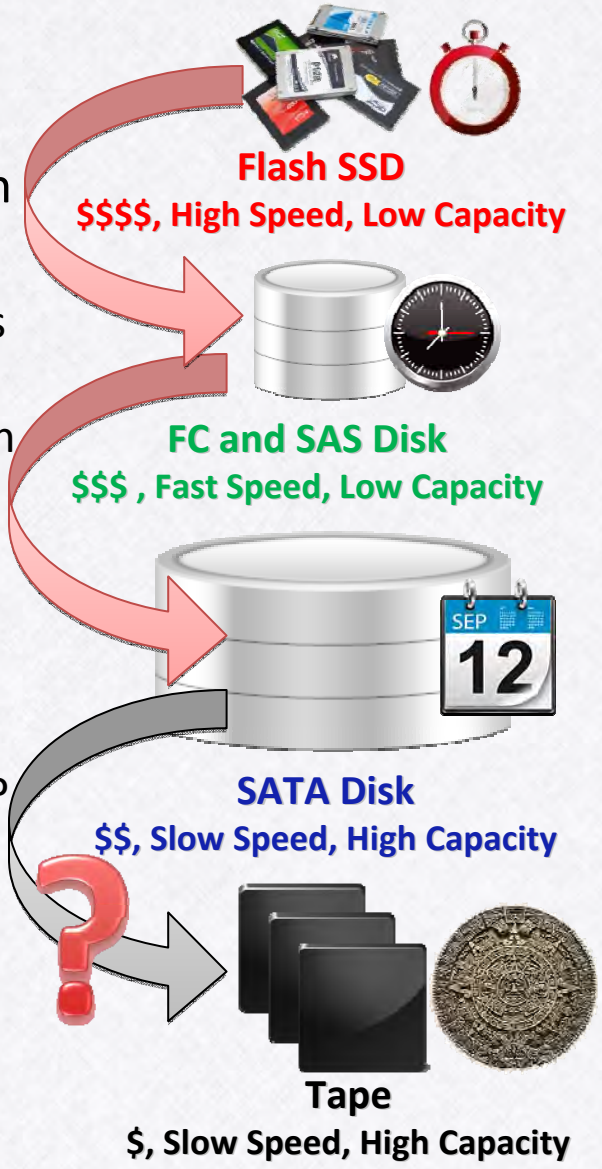
Third case: on-array tiering…

- Value proposition: Migrate older data automatically from expensive "capture" storage (FLASH SSD, FC Disk and/or SAS Disk) to less pricey "retention" storage (SATA Disk) for
  - More cost-effective capacity management
  - Improved array performance
  - Less downtime due to capacity issues
- Implemented inside the array itself using software functionality embedded on the array controller
- Pre-integrated and scalable "data management"!



Third case: on-array tiering…

- What does On-Array Tiering do to array cost at the disk drive level?
- How "smart" is the data migration technology?
  - Does migration consider business context of data or use simplistic non-granular FIFO rules (based on watermark or DATE LAST ACCESSED or DATE LAST MODIFIED metadata)?
  - How is capacity provisioned if data on RETENTION disk must be re-migrated to CAPTURE storage?
  - How are data protection services able to keep up with changing data locations?
  - Is tape part of the "data management" path?



On a related note…

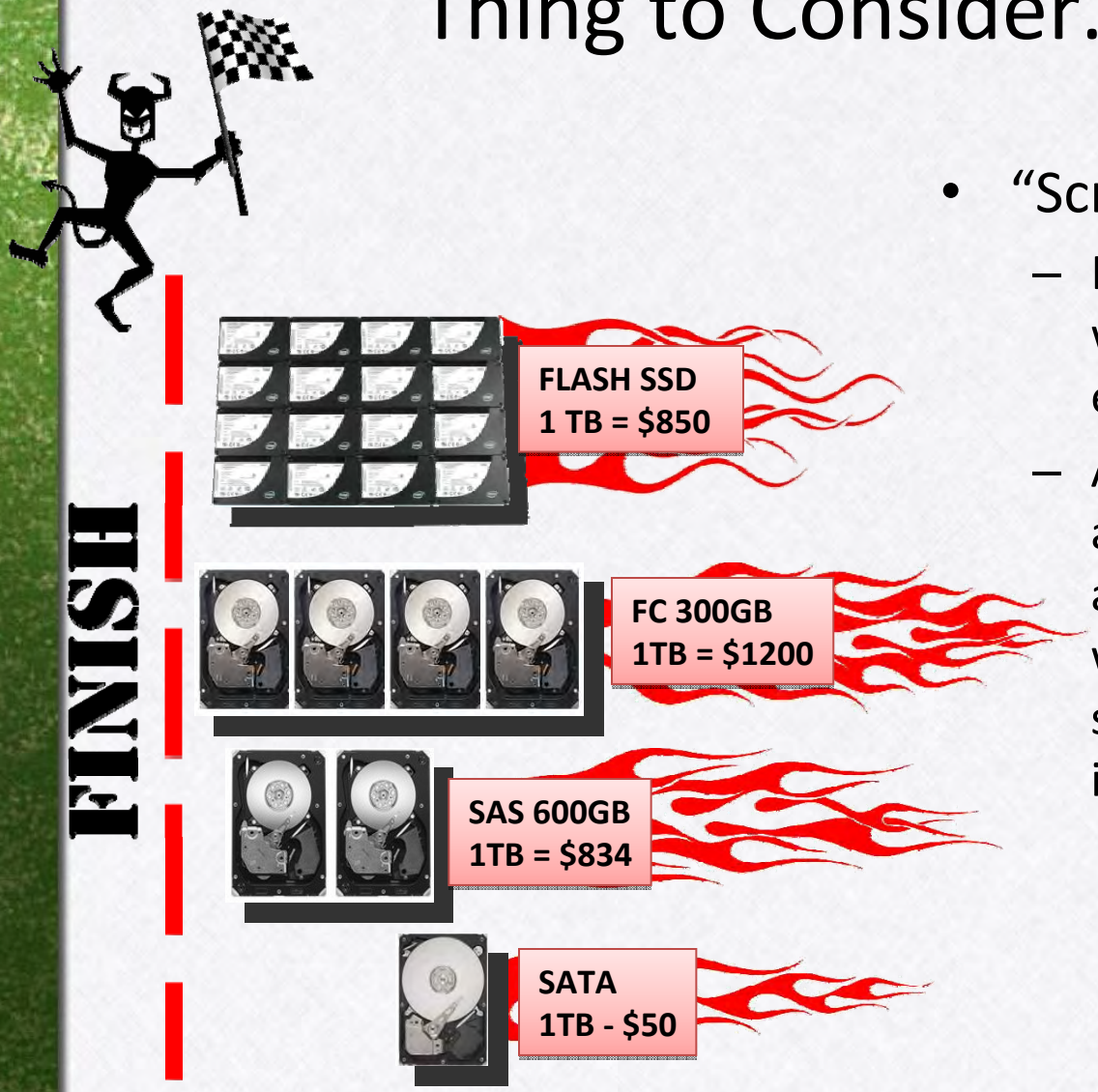
- Is Flash SSD really the breakthrough that the vendors are claiming?
  - "2/3 less power and less BTUs than disk drives"
  - "Wear leveling and block placement technology ameliorate memory wear and non-linear performance"
  - "1.5 million hours MTBF"
  - "Up to 21,000 IOPs"

EFFICIENCY

Depends on how you apply it…

- Write-intensive, transaction processing (Flash SSD as Tier 0)?
- Read-intensive applications (Flash SSD as Read Cache)?
- Typical PC operations?
- Exotic/specialized loads?
  - File tree crawling in scale-out NAS
  - “Hot sheet” swapping

Oh, And There's That Other Little Thing to Consider…



- "Scratch." (Cost.)
  - Is the speed worth the expense?
  - Are there alternative approaches that will deliver satisfactory improvements?

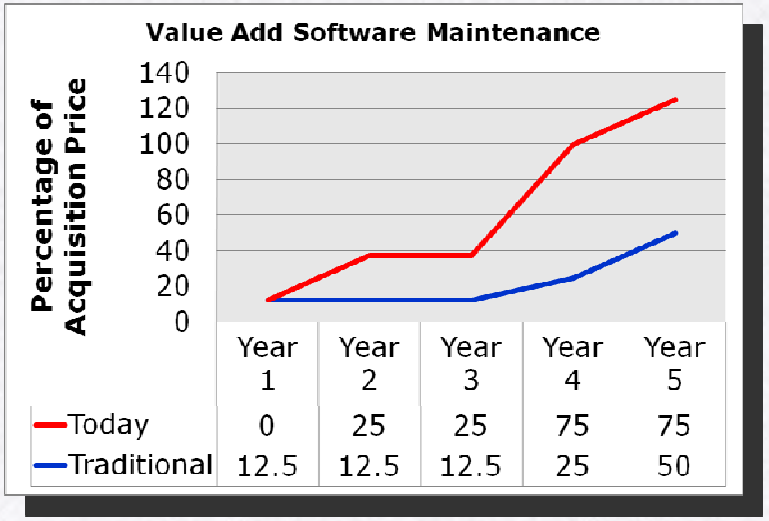
Bottom Line: Does Value-Add Contribute Meaningfully to Storage Efficiency…

- By
  - Automating tasks for reduced labor burden
  - Helping to cope with data growth
  - Simplifying data protection
- Or are value-add arrays
  - Introducing array controller complexities that increase likelihood of downtime events
  - Isolating data into islands of disk without a workable scalability or unified management model
  - Supporting the vendor's objectives more than your needs…

EFFICIENCY

Considering the storage acquisition and maintenance cost-acceleration…

- Smarter storage may be a not such a smart investment…



- The only other industry with a comparable sales model…

Storage is starting to look like a nail in search of a cost-cutting hammer

- In these days of budgetary belt-tightening, you can be sure that management is paying attention to what things cost…even if they don't understand all of the technology
- They are looking for the right investments to meet business needs…

EFFICIENCY

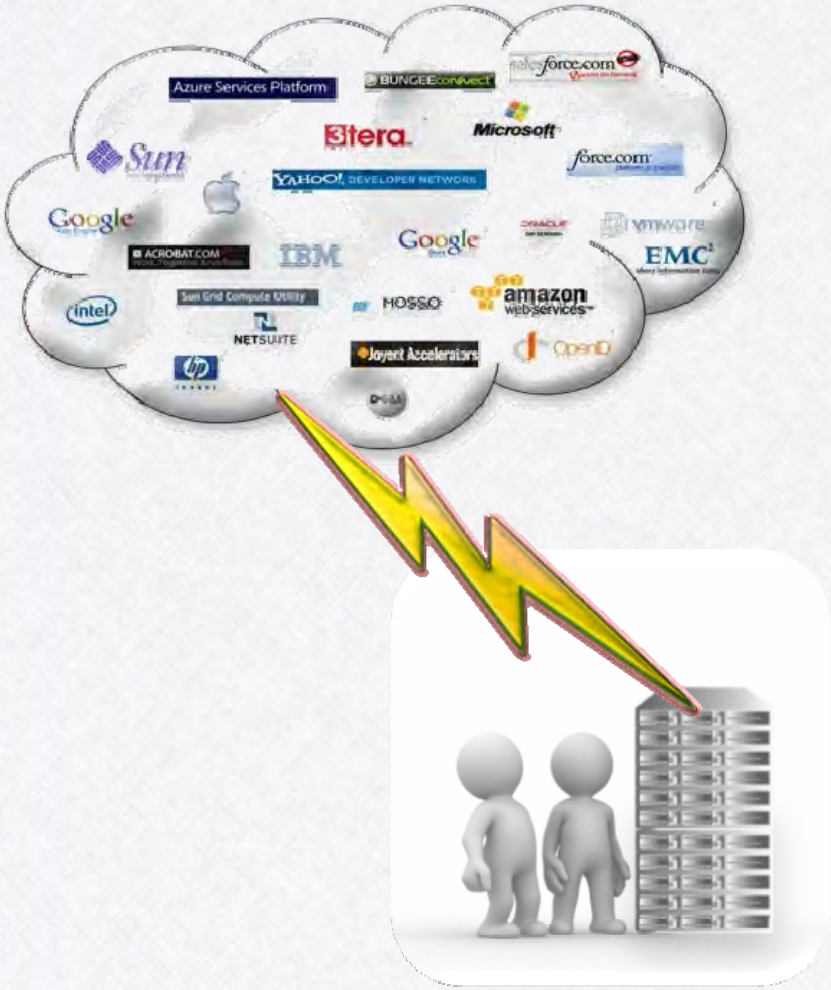
Their assessment of business value is straightforward…

What are you doing today to contain the growth of storage expense and what's your plan for containing costs going forward?

How are you minimizing the risk to data and compliance with legal and regulatory requirements?

How are our investments in storage increasing our productivity and competitiveness to drive top line growth?

Ignoring these concerns is inviting trouble…



- Management is reading articles about "Cloud Storage" in *Forbes*, *Barrons*, *The Economist*, etc.
- Vendors claim it is
  - Cheaper per TB
  - Better managed
  - Less labor intensive
  - Better protected
  - More compliant
- *Is your resume up to date?*

EFFICIENCY

Delivering storage efficiency isn't just a financial matter…

- It is the lynchpin of other initiatives
  - Server and desktop virtualization…
  - Managing data growth, security and compliance…
  - Safeguarding the corporate data asset from loss or destruction…
  - Reducing power demand and data center energy costs…

EFFICIENCY

Let's start with virtualization…

- Please roll video 1…

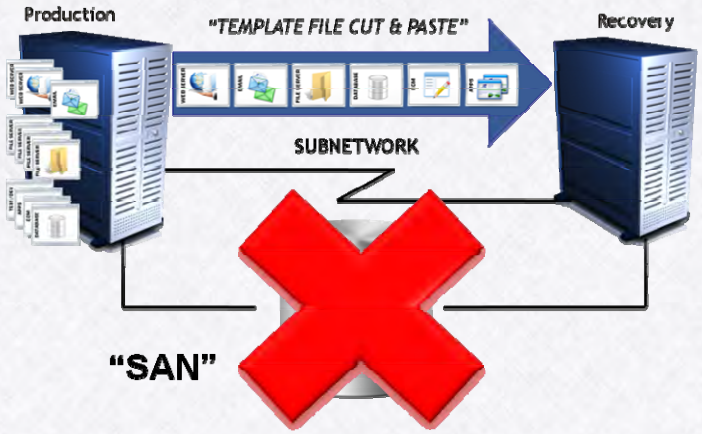
Entire presentation to be delivered at *Storage Decisions* next week…

- A quick summary
  - Storage infrastructure is creating the "Big Stall" in storage and desktop virtualization projects
  - Isolated functionality on island arrays cabled together inflexibly into FC fabrics – combined with lousy I/O handling by hypervisors – is creating a nightmare both in terms of capacity sharing and performance
  - Many IT planners confront a need to forklift upgrade storage to meet new workload profiles: a huge cost that was not anticipated in virtualization plans!

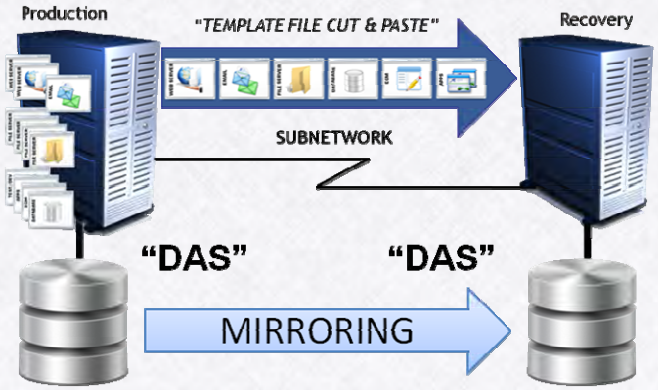
EFFICIENCY

Server and desktop virtualization driving attention to storage efficiency

- Efforts to make storage cooperate with new workloads and I/O profiles are making many companies reconsider hardware and topology choices…



"Abandon SANs for DAS"

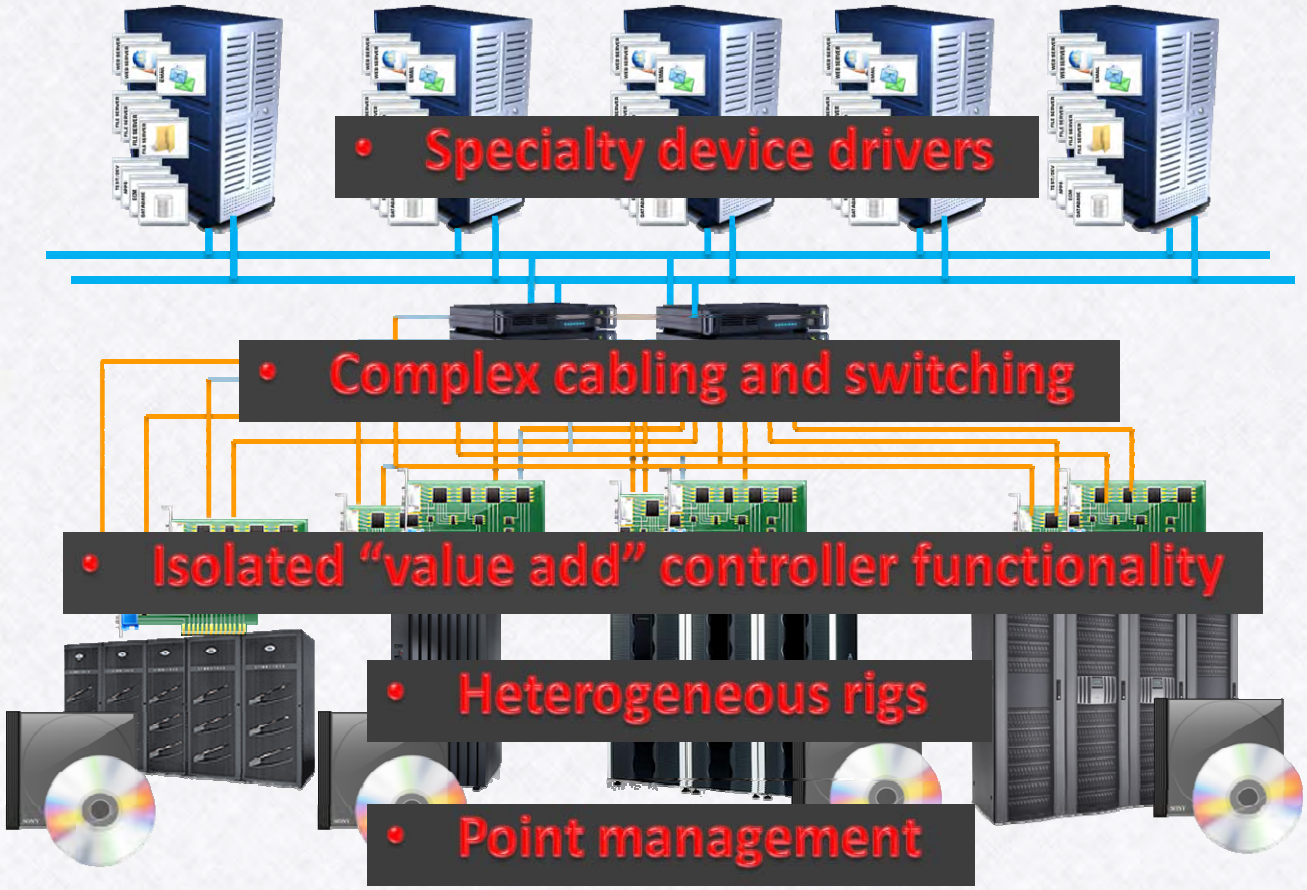


"Mirror data on the back end"

Those SAN things never delivered on their value proposition anyway…

EFFICIENCY

If only it were that easy…



In short, virtualization has underscored problems that already existed in storage…

EFFICIENCY

And it has re-opened our eyes to the basics of efficient storage infrastructure design…

- Is the best storage architecture
  - Isolated array-based islands of data…
  - Each island providing a specialized compliment of value-add services…
  - Each island managed with point management utilities?
- Or
  - Storage as Building Blocks: units with predictable performance that scale at least linearly in terms of performance…
  - Offloading value-add to an externalized software layer for ease of allocation and sharing…
  - Having a common management model?

versus

Simpler components beginning to emerge…

Xiotech's entire family of Intelligent Storage Element (ISE) arrays eschews value-add software on the array controller. Implements only what needs to be near the disks, opens architecture for 3rd Party vendors with in an ecosystem…

NetApp has begun moving some of its ONTAP software functionality off of its box and making it available as server hosted software…

Announced in January the "deconstruction" of Clariion and Celerra, moving value-add software to a server head…

EFFICIENCY

Credit where due: Xiotech showed the way

Basic building block (or "brick") the ISE array – several varieties optimized for performance, capacity, "balanced" and hybrid, place in FC or SAS fabric…

"Surface" storage using their Block Storage Controller or a File Storage Controller (an x86 server hosting required presentation functionality and any 3rd Party value-add software you think you need) or choose a 3rd party head…

Manage everything via Web Services REST and HTTP using your iPad, iPhone, laptop or desktop browser, etc. All management utilities and software stack being made free to storage industry and developers at CORTEXDeveloper.com…

EFFICIENCY

And storage virtualization is at last finding its voice…

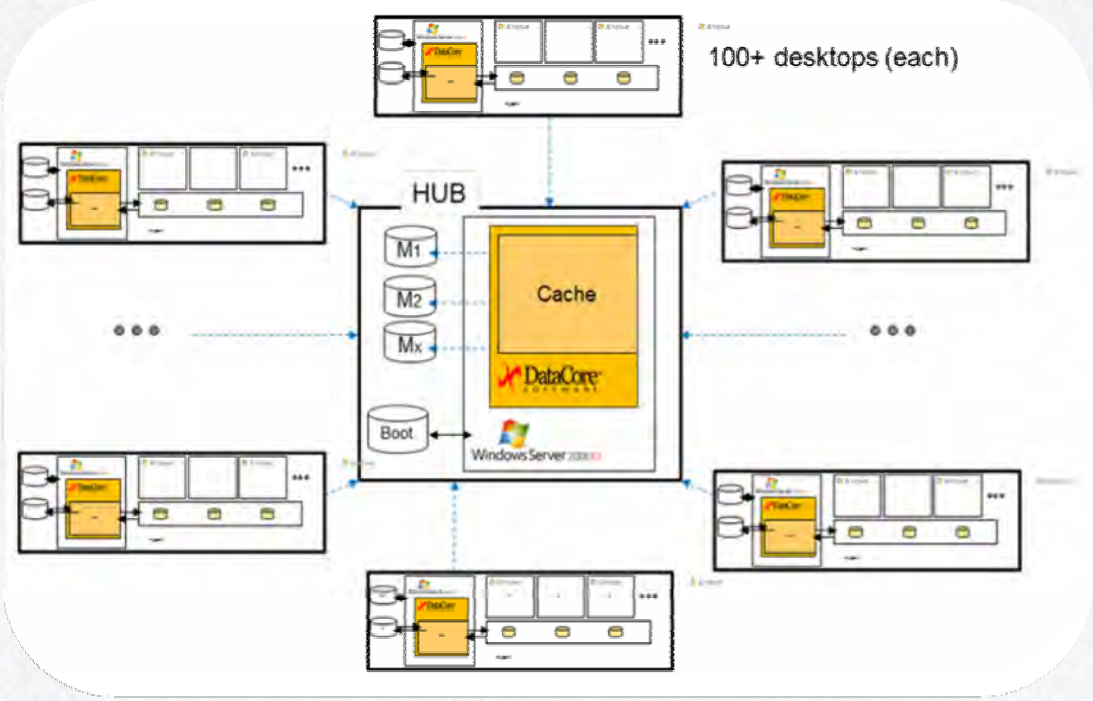
And groundbreaking work by DataCore Software is advancing the goal of defining "atomic units" of storage to support applications like VDI…

View my interview with Aral at C4Project.org or on DrunkenData.com

EFFICIENCY

A Building Block Model for VDI?

- With predictable performance and cost, and incremental scalability…



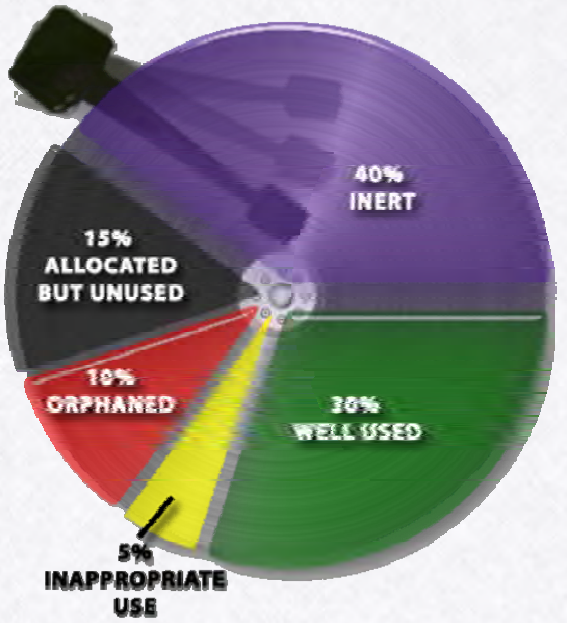
http://www.datacore.com/Libraries/Document_Library/VDI_benchmark_paper_FINAL_21-Dec_2010.sflb.ashx

EFFICIENCY

Efficiency isn't just about investments and infrastructure…

- Data is growing…though I hesitate to use the term "exploding"
- Companies generally do not know what their rates of data growth are, but they do know how much capacity they are adding annually…often to their chagrin
- We also know three other things:
  - Poorly managed and unmanaged data are adding to what must be stored and protected
  - Data is mostly poorly managed or unmanaged
  - Nobody wants the job of sorting out the junk drawer!

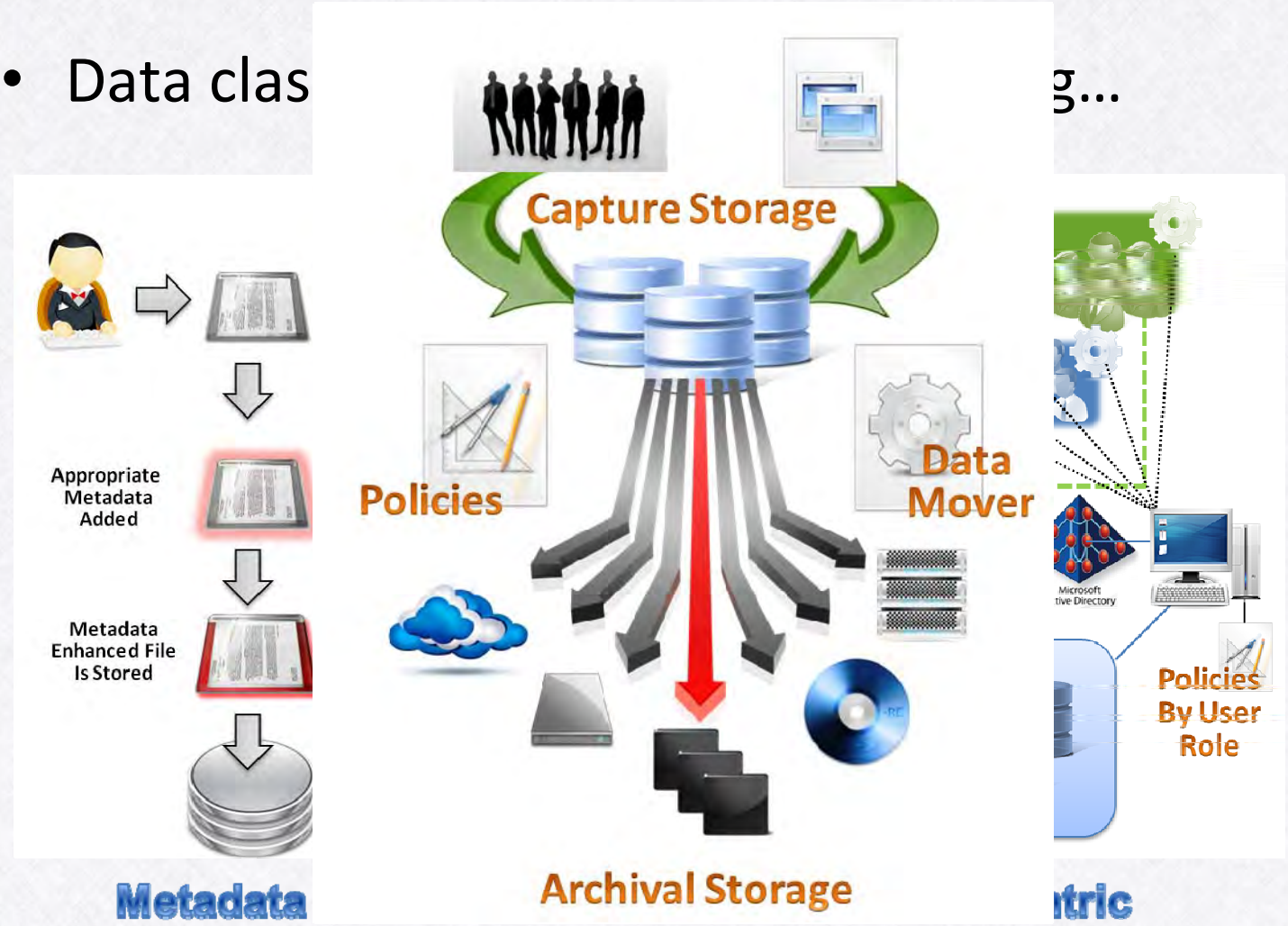
Unmanaged data has serious repercussions for storage costs, data security and discoverability and disaster preparedness.



Source: *Making IT Matter*
(Toigo, 2011 at MakingITMatter.com)

Outlines of a solution appearing

- Data clas g…



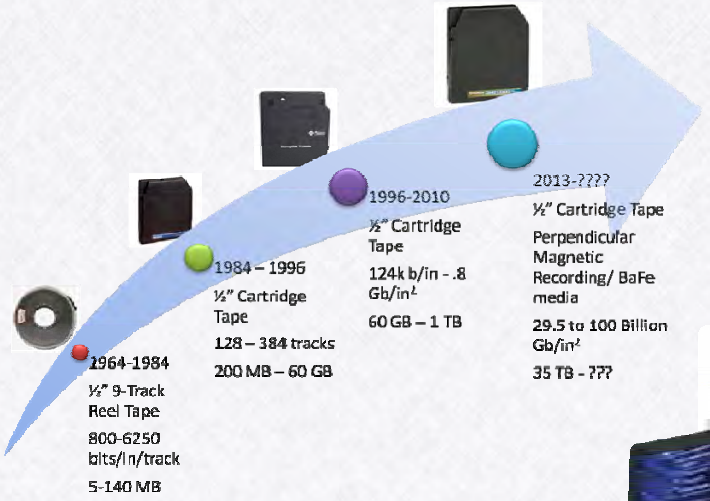
Feeding an Archive Strategy

EFFICIENCY

Benefits of better data hygiene and archive

- Capacity reclamation, reducing storage spending requirements
- Faster application performance
- E-discovery capabilities suited to legal and regulatory requirements
- Reduced burden on data protection and disaster recovery processes: faster data recovery

Potential Role for Tape Is Huge!



Big Tape for Big Data: 35 TB Tape Cartridges

Tape + File System A New Archive System?

EFFICIENCY

These days, that last point about data protection and DR, is front-of-mind…

- Please roll video 2…

EFFICIENCY

The full presentation will be made at Storage Decisions next week…

- A quick summary
  - [illegible]
  - [illegible]
  - [illegible]
    - [illegible] capability

Oh, and don't sweat the Mayan apocalypse…

ROLE OF TAPE

The role of tape in DR persists…

- Despite what the tape-is-dead crowd may say, tape still delivers…

Protection for nearly 80% of all data…

Fastest magnetic media write speeds…

Restore target agnosticism…

Highest storage density per raised floor tile…

Unmatched portability…

Unrivaled reliability…

Air gap protection for stored data…

Lowest cost per GB of all magnetics…

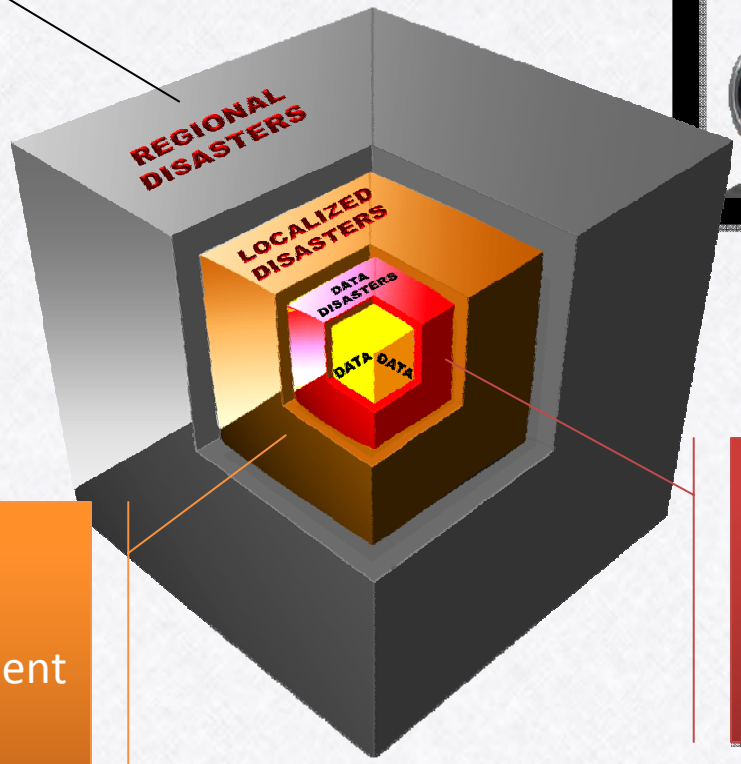
Lowest power consumption…

ROLE OF TAPE

Even with virtualization and VTLs…

- Tape backup is a key part of a defense-in-depth strategy for data protection…

Regional Interruptions caused by natural and man-made disasters



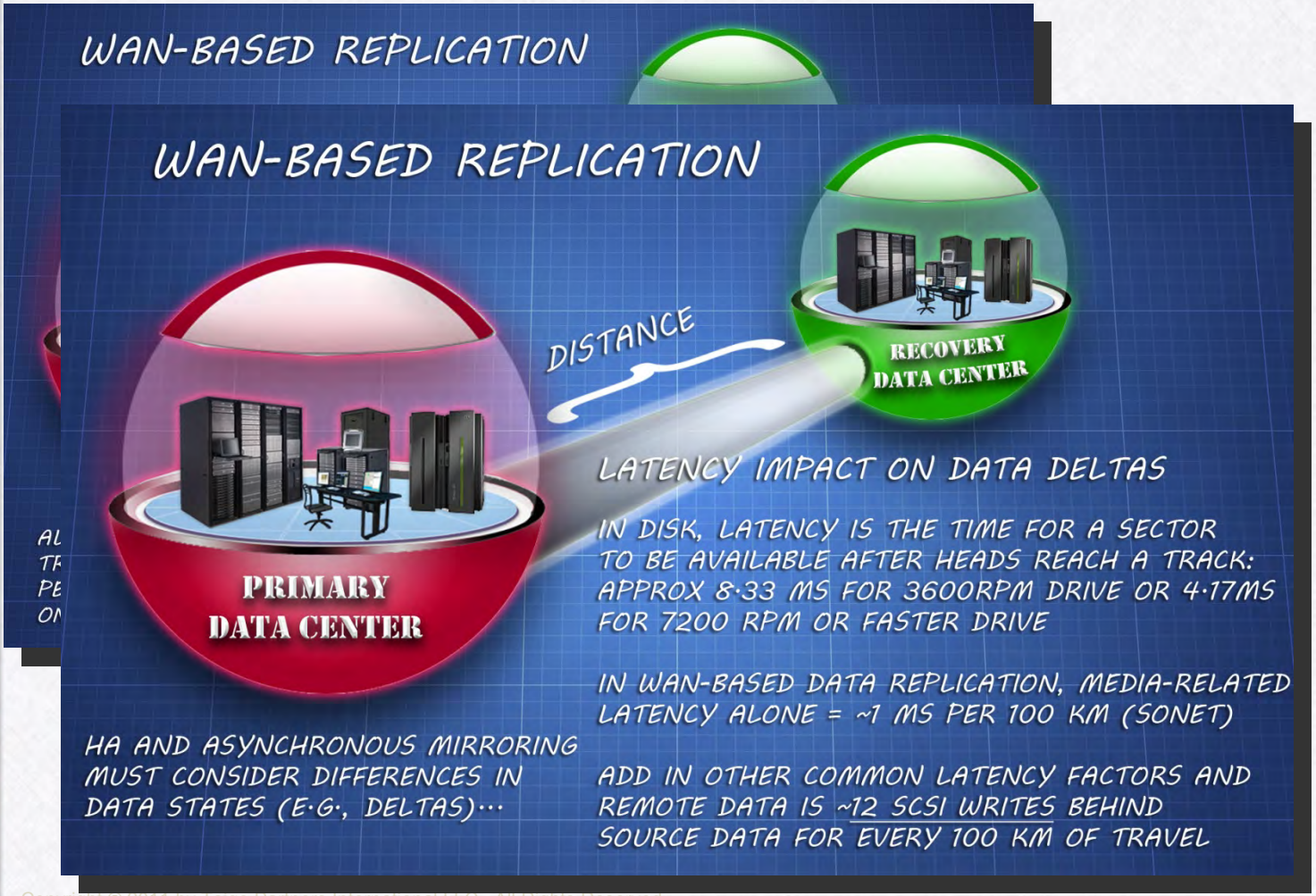
Localized Interruptions involving equipment and facilities

Data corruption or deletion due to viruses or malware, application and user errors

ROLE OF TAPE

Regardless of bandwidth, WAN-based VTL to VTL replication…

- Still confronts LATENCY: producing "data deltas"

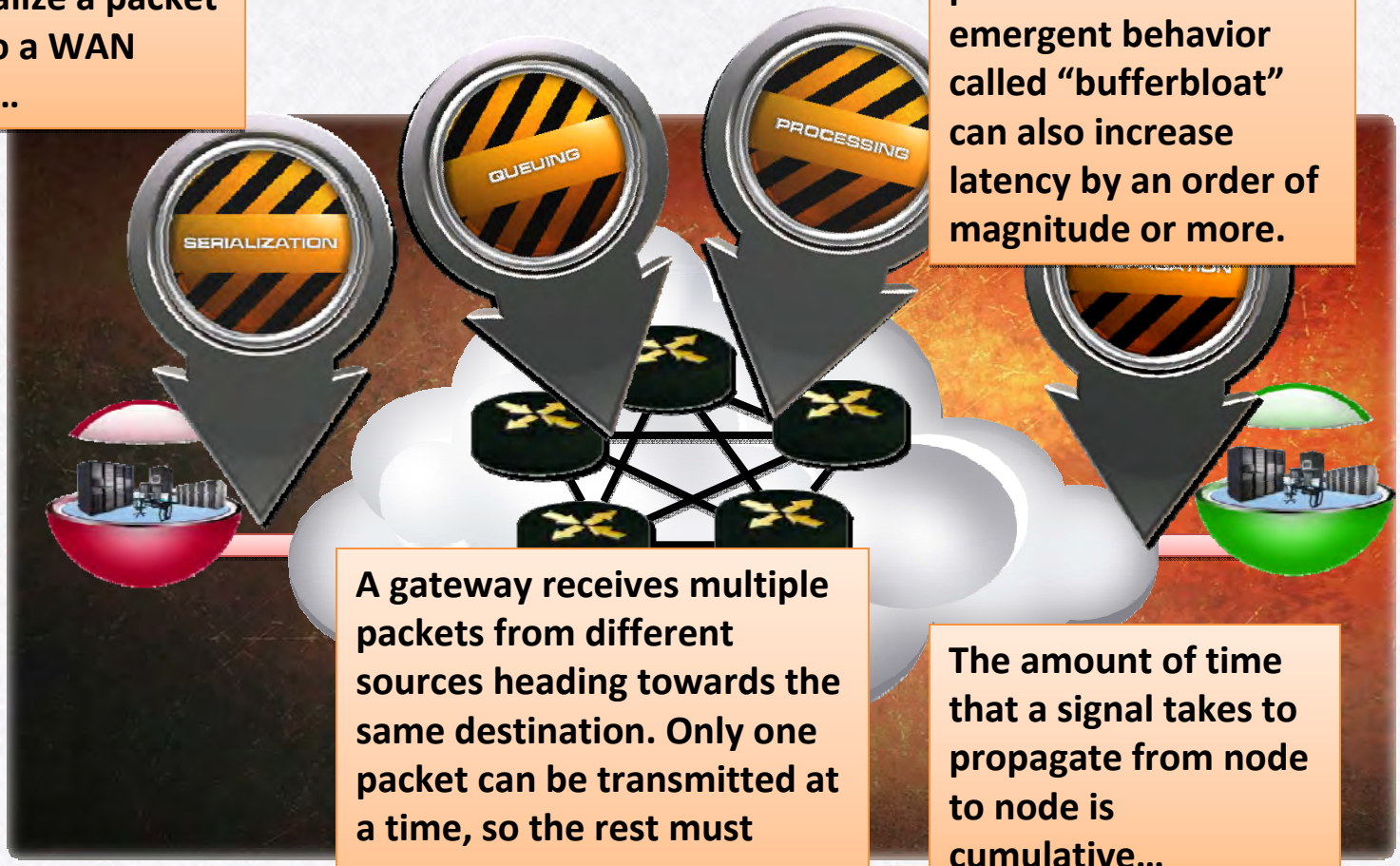


ROLE OF TAPE

And not just “distance induced” latency…

Time is required by a router to serialize a packet onto a WAN link…

A gateway adds delay while determining what to do with a packet. A new and emergent behavior called “bufferbloat” can also increase latency by an order of magnitude or more.



A gateway receives multiple packets from different sources heading towards the same destination. Only one packet can be transmitted at a time, so the rest must queue…

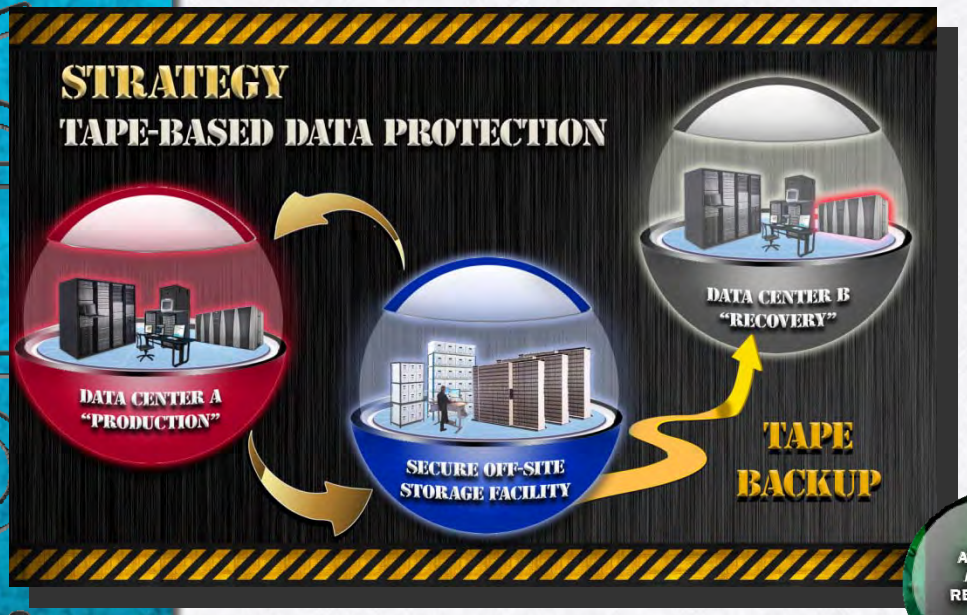
The amount of time that a signal takes to propagate from node to node is cumulative…

ROLE OF TAPE

If you want to use a VTL, make it complementary to your tape operation…

ROLE OF TAPE

Instead of Looking for Ways to Replace Tape, the Smart Money is On…



Sticking with what works…

And refining the processes for using it efficiently…

CHALLENGES

Earworms persist, I know…

- Getting to storage efficiency may just require you to stop listening to the song that array vendors keep playing…

Tape is dead. Move on!

Tape is so 20th Century!

De-duplicating VTLs are better!

1 in 10 tapes fails on restore!

Bandwidth is ubiquitous. Use the WAN!

…And choose your own song!

Questions?

- Thanks.
- jtoigo@toigopartners.com