BACK TO THE FUTURE

⟶ LTFS

How the Linear Tape File System and Other Innovations
Are Returning Tape to a Production Storage Role in the Data Center

Agenda

- Brief history of time, space and developments in storage media…

- Developments in tape technology

- New (old) roles for tape

- How LTFS promises to reinvigorate tape as a production file storage medium

Presented by
Jon Toigo
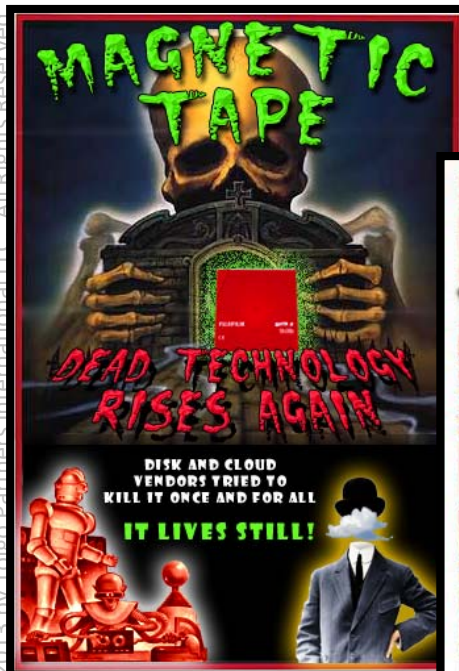Toigo Partners International
Data Management Institute
IT-SENSE.org

TOIGO PARTNERS INTERNATIONAL   DMI   IT-$ENSE

"Tape? We Thought You Was Dead…"

"Yeah…I get that a lot."

You may have seen the movies…

Tape killed by cheap disk drives, or by lack of capacity or resiliency, or by SANs, or by clouds…

Bottom line: tape has been declared "as dead as the De Lorean" too many times to count, **but never accurately**…

A Litany of Past Wrong Projections…

- Tape declared dead in late '70s, replaced by *automated disk*
  - *Then, automated tape libraries appeared…*

- Tape declared dead in early '80s, replaced by *RAID and disk-based replication*
  - *Then, the limitations of WANs kicked in…*

- Tape declared dead in late '90s, replaced by *SANs*
  - *But #1 stated reason for SANs was to share an enterprise tape library…*

- Tape declared dead in early '00s, replaced by *de-duplicating Virtual Tape Libraries (VTLs)*
  - *But no one got the reduction ratio promised by vendors from overpriced rigs…*

- Tape declared dead in the '10s, replaced by *clouds*
  - *Latest nonsense from vendor quoted in CIO magazine – "60% will use clouds for backup and archive in three years…"*
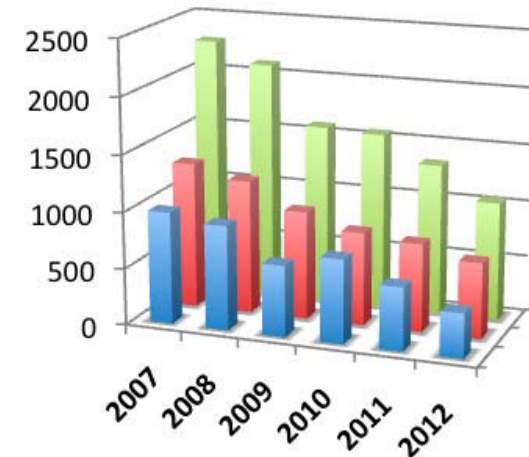
Recent Prognostications…

Tape Drive & Media Revenues
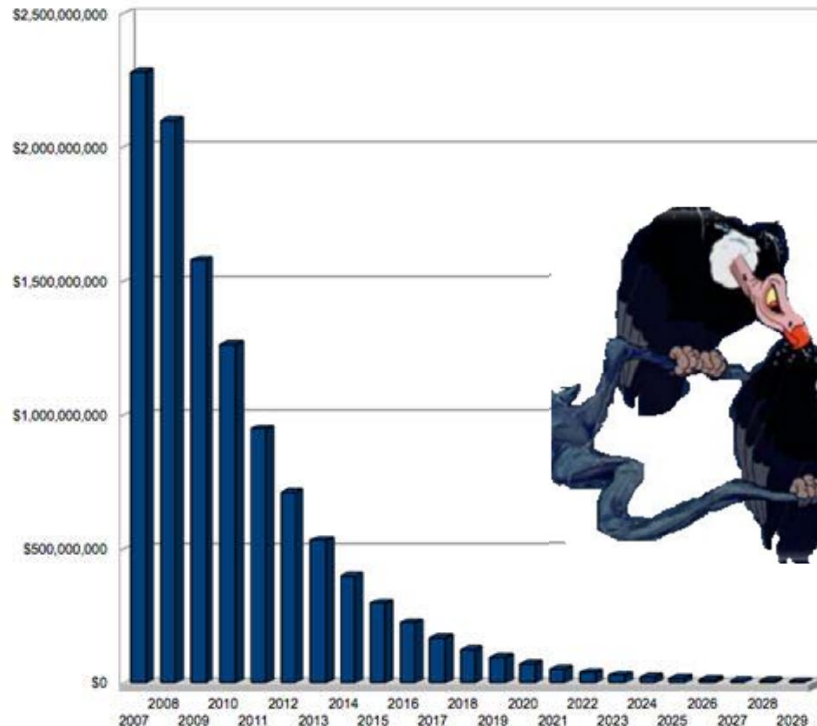
Source Unattributed from *The Register*, 10/19/12

- Tape Drives
- Tape media
- Drives and media

2007 2008 2009 2010 2011 2012

LTO drives have the most favorable outlook for hardware products over the forecast period. Their units sales are expected to grow at a CAGR of -5.09% from 2012 to 2017. LTO cartridges have the most favorable outlook among media products with a CAGR of -.08% leading to sales of $606.45 million in 2017.

Santa Clara Consulting Group, 5/6/13

Let's Stipulate that Tape Has Had Its Limitations...

- Tape operations were labor intensive (much less so now with automated libraries)

- Tape went through a period of competing formats and cartridge sizes (now largely resolved with LTO, and two generations of backward compatibility)

- Backup software was proprietary and a poor performer -- de-dupe, after all, was just "a waste management system for backup" (but LTFS is changing the storage methodology)

- Tape lacked reliability (said analysts, who were paid by disk vendors to say so. Tape is actually much more reliable than disk.)

- Tape does not do random access as quickly as disk (though its read and write speeds blow the socks off of any disk array)
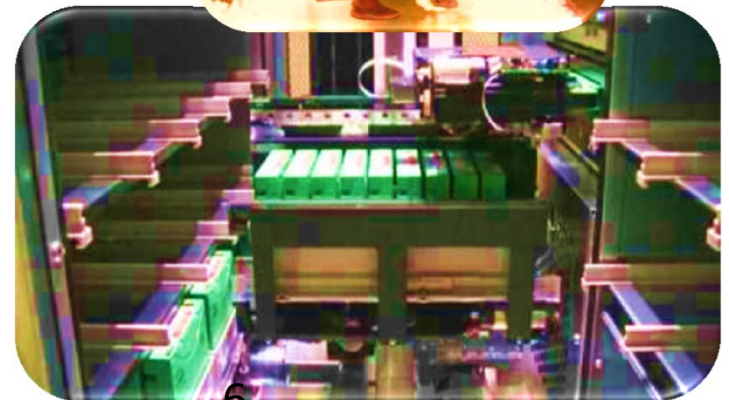
But Progress Is Ongoing...Like the DeLorean

1984-96
½" Cartridge Tape
128-384 Tracks
200 MB to 60 GB

1964-84
½" 9-Track
800-6250 bits/in/track
5 to 140 MB capacity

2014-Future
½" Cartridge Tape
BaFe Media
Type II PMR
29.5 to 100 Billion
Gb/in2
32 TB to ???

1996-2010
½" Cartridge Tape
124K bits/inch to 800 Mb/inch
60 GB to 1 TB

DeLorean
Motor Cars
DMC 12

DeLorean
Motor Cars
Time Machine 1

Anti-Gravity
Flight Capability

Barium Ferrite

FUJIFILM ULTRIUM 6
BariumFerrite 2.50TB/6.25TB

Credit Where Due

Electronics / October 20, 1982

Iwasaki: inventor of vertical recording

Floppy-disk drives using vertical recording to achieve 10 times the density of present longitudinal-recording drives should hit the market next year. They will be the first of many products that incorporate the technology proposed at the 1977 Intermag meeting by Professor Shun-Ichi Iwasaki of the Research Institute of Electrical Communication, Tohoku University, in Sendai, Japan. He predicts that "this will open an era of perpendicular recording that will last 30 years, just as the era of longitudinal recording lasted for 30 years."

Iwasaki received his undergraduate and graduate degrees at Tohoku University and has been an assistant and then a full professor there for the past 25 years, spending his entire career working on high-density recording techniques and theory.

His heavy theoretical studies positioned him to take full advantage of the potential of vertical—perpendicular, he says, is more descriptive—recording when the time came. In 1968, he developed a new theory of recording, including a theoretical quantitative computer analysis of internal magnetization. That was followed in 1972 by a model for a circular recording mode, which of course has a vertical component.

Iwasaki detoured into optomagnetic recording, but it enabled him to develop the cobalt-chrome medium used in vertical recording. "Perpendicular magnetic recording is better," he says, emphasizing the "magnetic."

Back on his main track, Iwasaki developed the standard vertical-recording head, a double-layer medium, and a theory that supports this work. The head features a coil wound on a large auxiliary pole because a winding on the slender main pole—a stripe of magnetic material deposited on a nonmagnetic support—would saturate its tip. The double-layer medium has bits in the form of horse-shoe magnets with attraction between poles, rather than the polar repulsion of longitudinal recording media.

Iwasaki says that theory shows that head, magnetization, medium, and all other aspects of perpendicular recording are complementary to longitudinal recording. And as it happens, "while longitudinal recording is suitable for analog signals, perpendicular recording is better for digital signals," he comments.

In his leisure hours Iwasaki paints in oils. He says that "it is similar to my research because I must do the entire painting in parallel. One doesn't complete one side and then move across the canvas."   -Charles Cohen
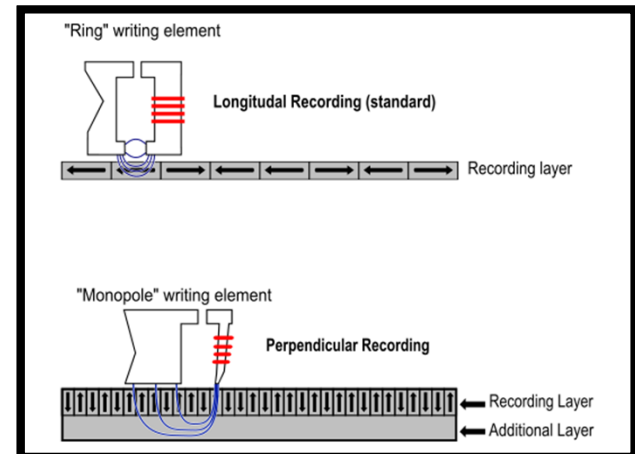
Journal of Magnetism and Magnetic Materials, Volume 287, Feb 2005, Pages 1-8

Shun-ichi Iwasaki

**1976 – Iwasaki, then professor at
the Tohoku University in Japan,
invents Perpendicular Magnetic Recording (PMR)**

Things Take Time

- 1990 Toshiba ships first commercial PMR floppy disk drive (2.88 MB capacity) then, in 2005, first 1.8" HDD with 80 GB capacity

- Seagate Technology ships first 2.5 inch PMR drive in 2006, first Cheetah 15K with 300 GB capacity in April 2006, First Barracuda 7200 3.5 inch with 750 GB capacity

- 2007 Hitachi introduces first 1 TB 3.5 inch drive

- 2009 Western Digital announces first 2 TB SATA drive with PMR technology

- 2011 Seagate releases 4 TB HD

- There have been demonstrations of 800-900 Gb/in^2, WD sees "shingled media" provides interim step to higher capacities enabled by PMR plus Bit Patterned Media (BPM) or Heat Assisted Magnetic Recording (HAMR)…


29 years later

35 years later

34 Years After Iwasaki

- 2010 – FujiFilm and IBM demonstrate tape media coated with Barium Ferrite: initial capacity of LTO style cartridge: 35 TB – "PMR by virtue of particle structure" -- BaFe II media will yield huge capacities (40TB already demonstrated)

- 2010 – Hitachi Maxell and Tokyo Institute of Technology demonstrate ultra-thin nano-structured magnetic film using "facing targets sputtering methods" delivering over 50 TB per tape cartridge – No products yet.
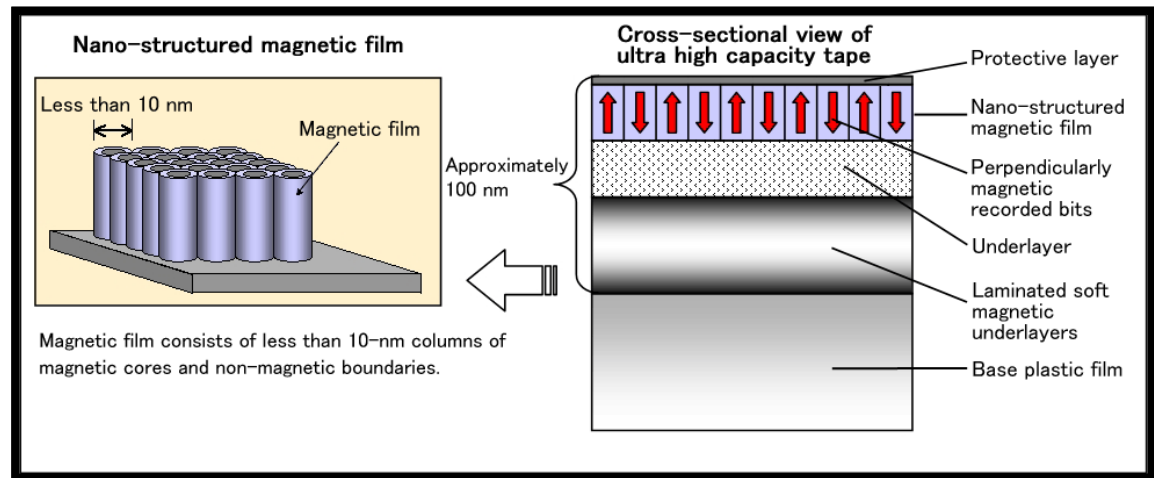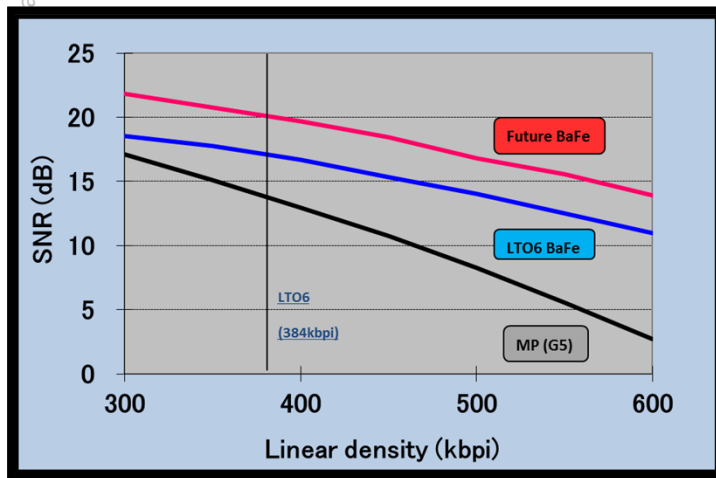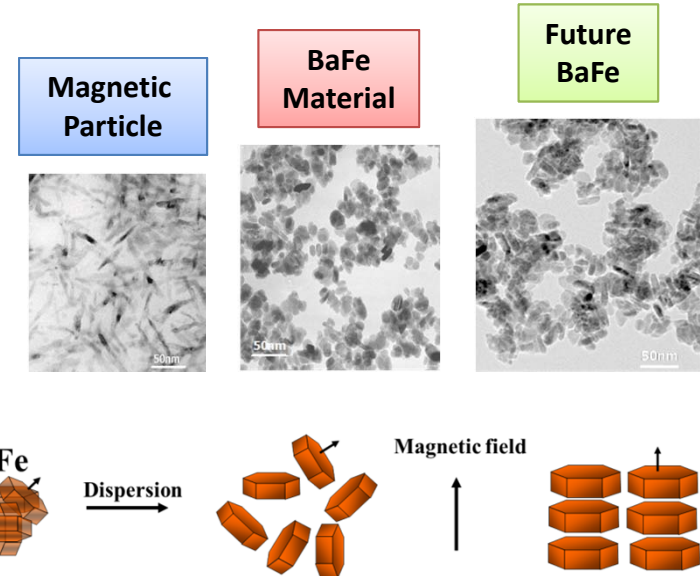
In Addition to Capacity, Improvements in Resiliency…

- ## Extensive testing to assess

  – Amount of de-magnetization over time (60°C 90% relative humidity)

  – Surface and Friction Coefficient (wrapping tape around a drive head, measure debris)

  – Corrosive gas testing ("Battelle evaluation" – tape left in corrosive gas environment for 14 days)
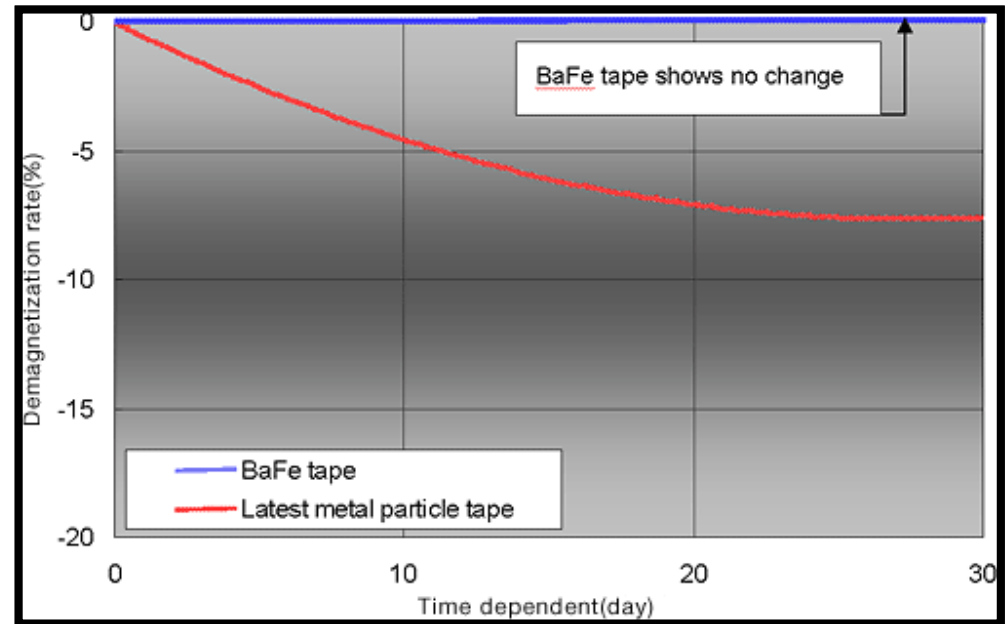
M. L. Watson, R. A. Beard, S. M. Keintz, and T. W. Feebeck, "Investigation of Thermal Demagnetization Effects in Data Recorded on Advanced Barium Ferrite Recording Media" IEEE Transaction on Magnetics, Vol 44, No. 11, Nov 2008.

Not to Mention Data Integrity…

- Cloud vendors may publish old FUD about tape – *72% failure rates for year-old tape, 1 in 10 failures on data restore, etc.* – none of which is true

- The facts: tape benefits from
  - In line error correction
  - Infinitesimal uncorrected bit errors

- No head crashes (LTO tape designed to be in contact with head)

Device	Hard error rate in bits	PB equivalent
SATA consumer	1.00E+14	.11
SATA Enterprise	1.00E+15	1.11
Enterprise SAS/FC	1.00E+16	11.10
LTO	1.00E+17	111.02
Oracle T10000	1.00E+19	1,1102.22
IBM TS1140 Technology	1.00E+19	1,1102.22

Uncorrectable bit error rate of tape is 10,000 times lower than disk.
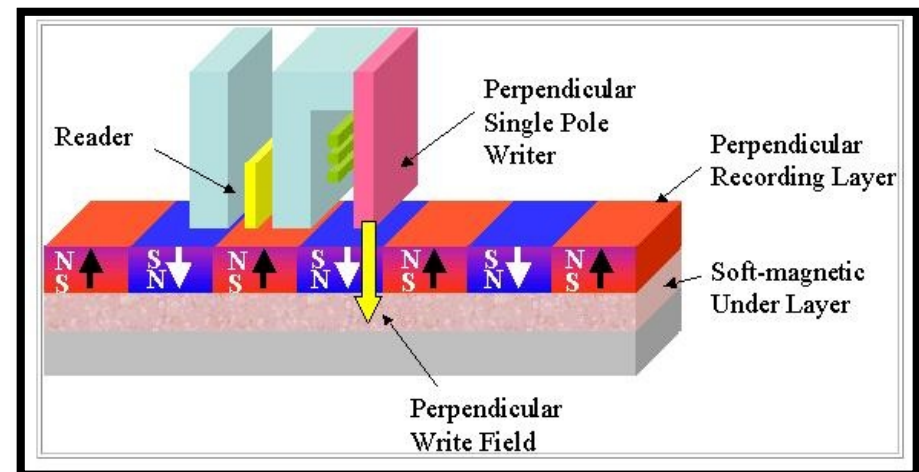
Okay, So We Get It...Tape Doesn't Suck (Technically)

- Despite what the haters say...

The Problem is That The Disk Folks Are Getting All of the Attention…
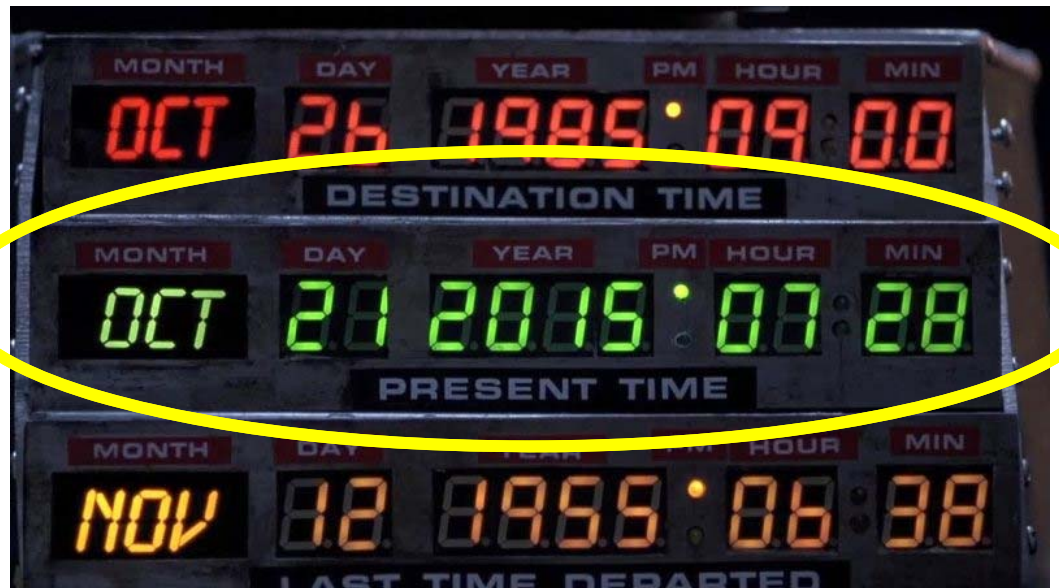
- And backup is so *day before yest...*

Tape Media & Automation Improvements Alone Aren't Going to Sell Tape

- Especially if we need to wait forever for products to appear in the market…

Maybe we need to go back to get to the future…

In the 60s and 70s, Tape Played Three Roles...

- A target for backup data, leveraging media portability, air gap, and restore target agnosticism...
- A home for deep archive - data we needed to preserve but didn't require on disk...
- And as a production data repository (yep, like disk, just another kind of active filestore)...

Backup Retains a Healthy Niche…

- Says Gartner, only 27% of SMBs have abandoned tape for other data protection solutions…

- Medium and larger firms continue to use tape predominantly – some analysts claim it is more difficult for them to abandon tape -- representing between 70 and 80 percent of the world's data…

- "Either/or" thinking is showing some signs of giving way to "why not both?"

According to IDC, tape data CAGR will be 45% from 2010 through 2015…

But backup may not be the primary application.

But Tape Backup is Also Problematic...

- Challenged by increasing data volume, proprietary backup software, unacceptable restore speeds, lack of operator skills
- While more cost effective than disk backup, especially at 100 TB or higher capacities, enterprise storage vendors have saturated the market with anti-tape advertisement
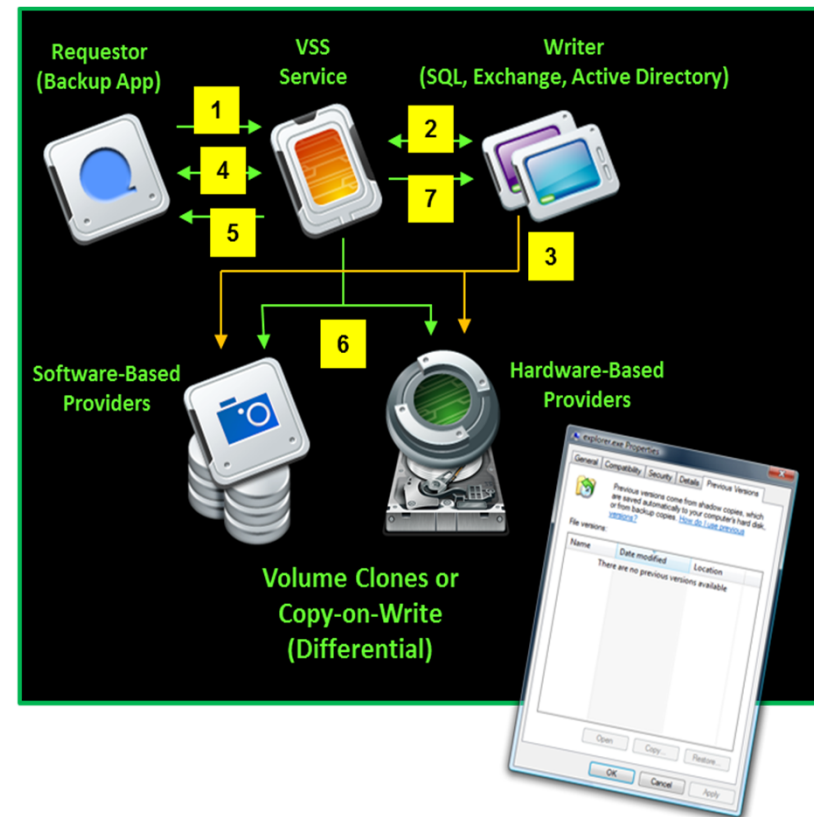- Reason is no longer the foundation of decision making in data protection

Maybe It's Time to Re-think Backup

- Instead of copying data into proprietary containers that must be restored to recover individual files...

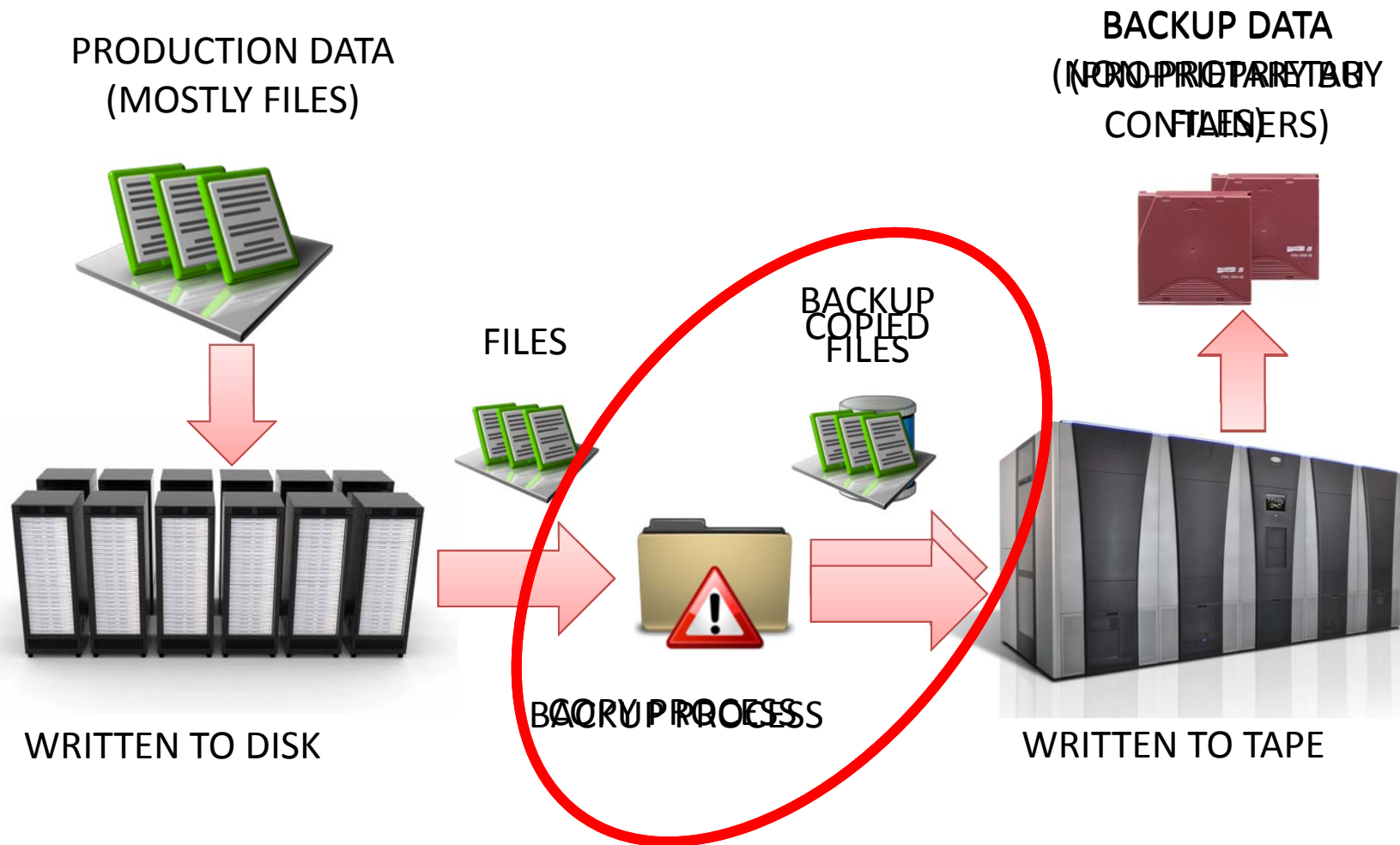- Why not simply copy files as files?
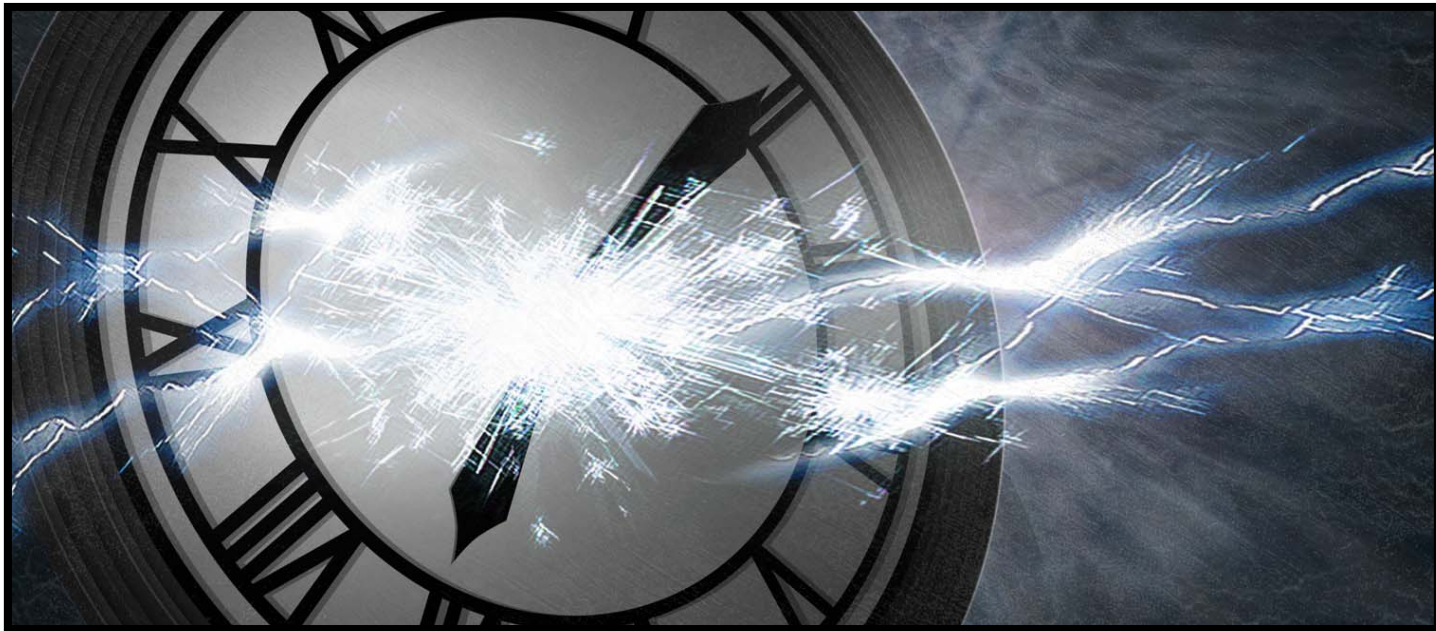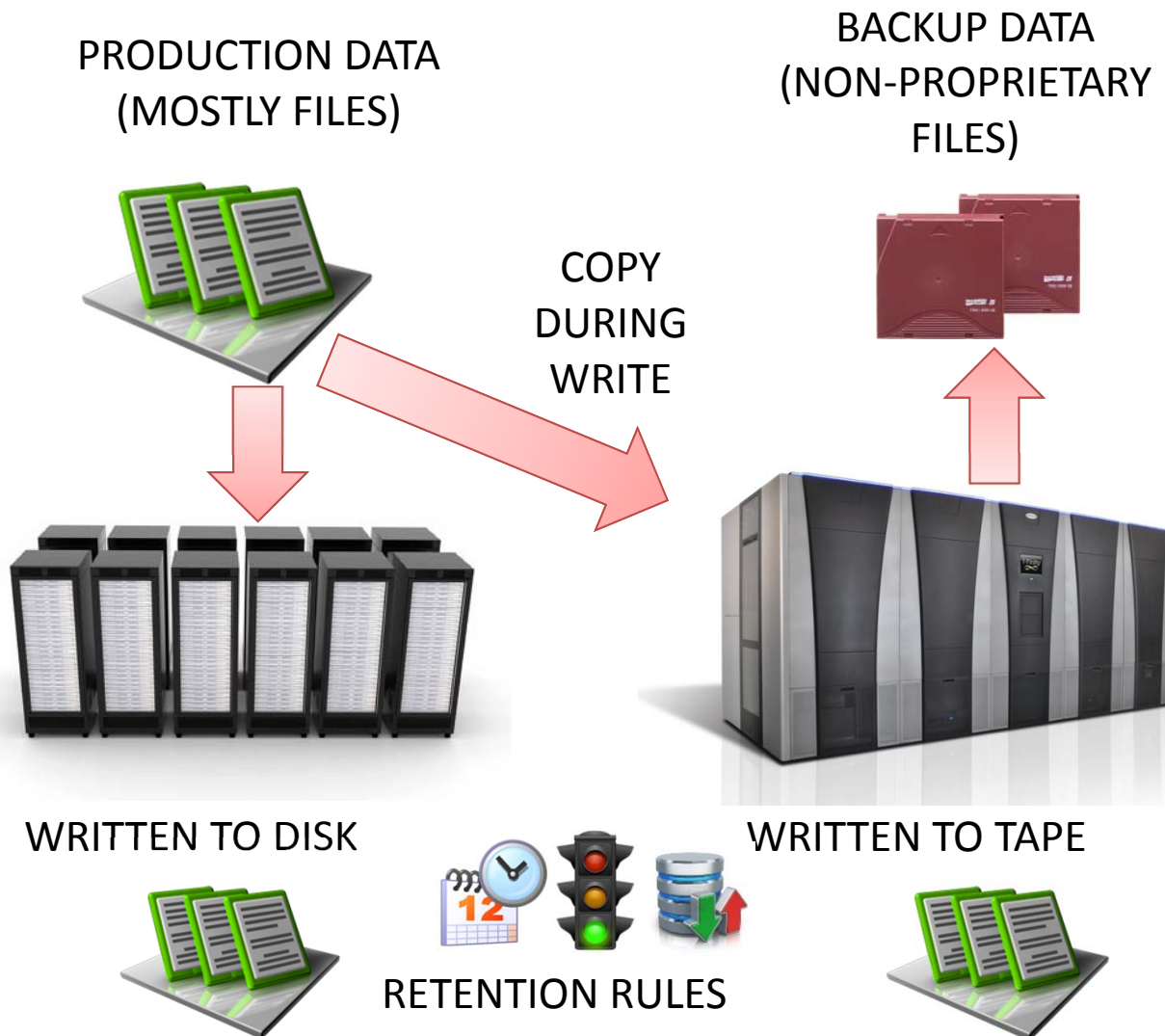
Microsoft is Going That Way…

- In case you haven't noticed, Volume Shadow Service (VSS) has disappeared in Windows 8 and Server 2012

- Microsoft pursuing Apple-like strategy of data protection via copy of change files to a network drive…

Simplify.

PRODUCTION DATA
(MOSTLY FILES)

BACKUP DATA
(NON-PROPRIETARY
CONTAINERS)

FILES

BACKUP
COPIED
FILES

WRITTEN TO DISK

BACKUP PROCESS

WRITTEN TO TAPE

Another Option (Some Call this "Active Archive")

PRODUCTION DATA
(MOSTLY FILES)

BACKUP DATA
(NON-PROPRIETARY
FILES)
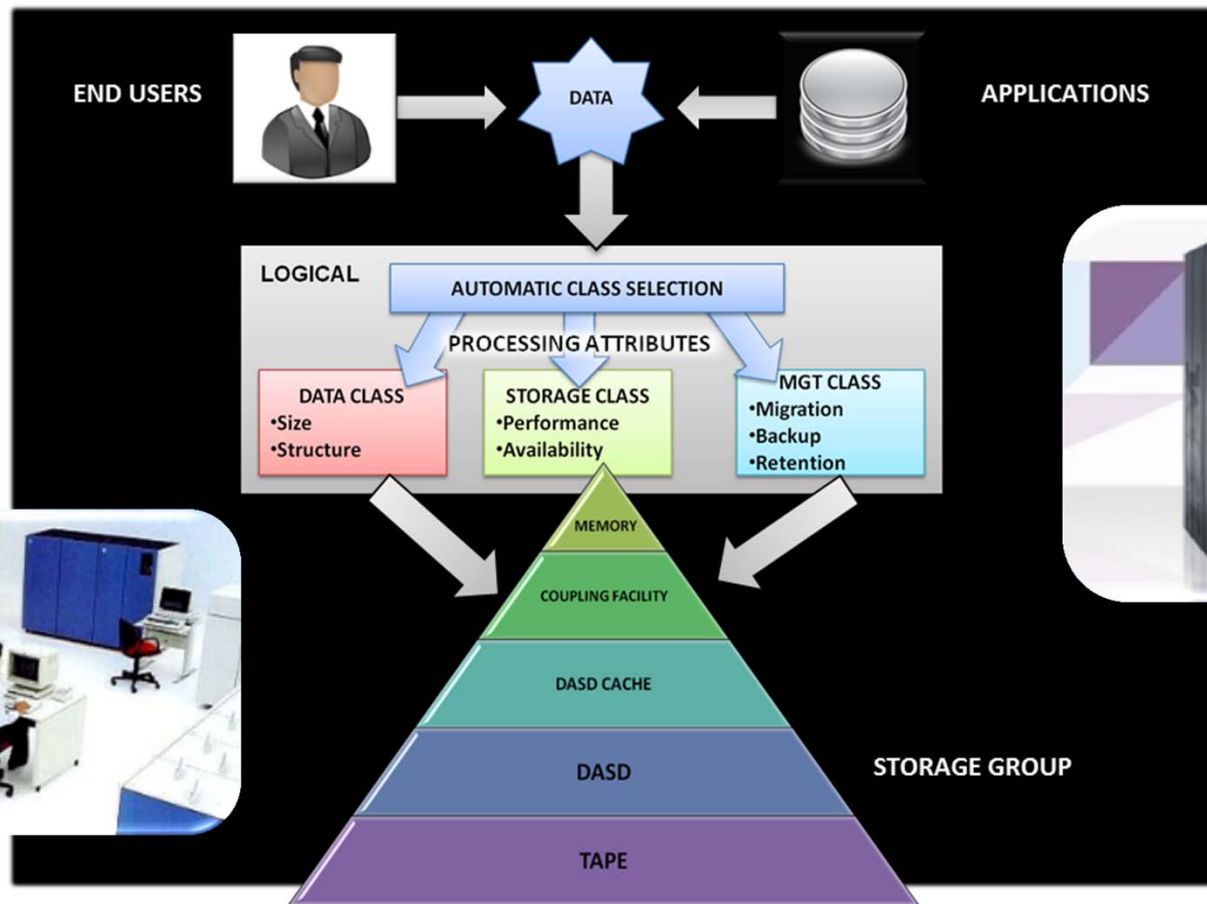
COPY
DURING
WRITE

WRITTEN TO DISK

WRITTEN TO TAPE

RETENTION RULES

I Call It Going Back to the Future…

- Tape traditionally played a production storage role, well suited to infrequently accessed data…

Key Enabler: Linear Tape File System (LTFS)

- Just when you thought that the technology vendors were running out of original ideas...

What's LTFS? Really?

- The Linear Tape File System was developed two years ago by IBM
- Yet another application of a file system structure and format to tape (there are several) but special because…
  - It is a unique approach leveraging tape partitioning
  - It is relatively easy to implement DIY or using a third party vendor appliance to front end the library
  - It is free (more or less) and soon to be developed as a standard by the SNIA

LTFS
LINEAR TAPE FILE SYSTEM

- LTFS Single Drive Edition (SDE)
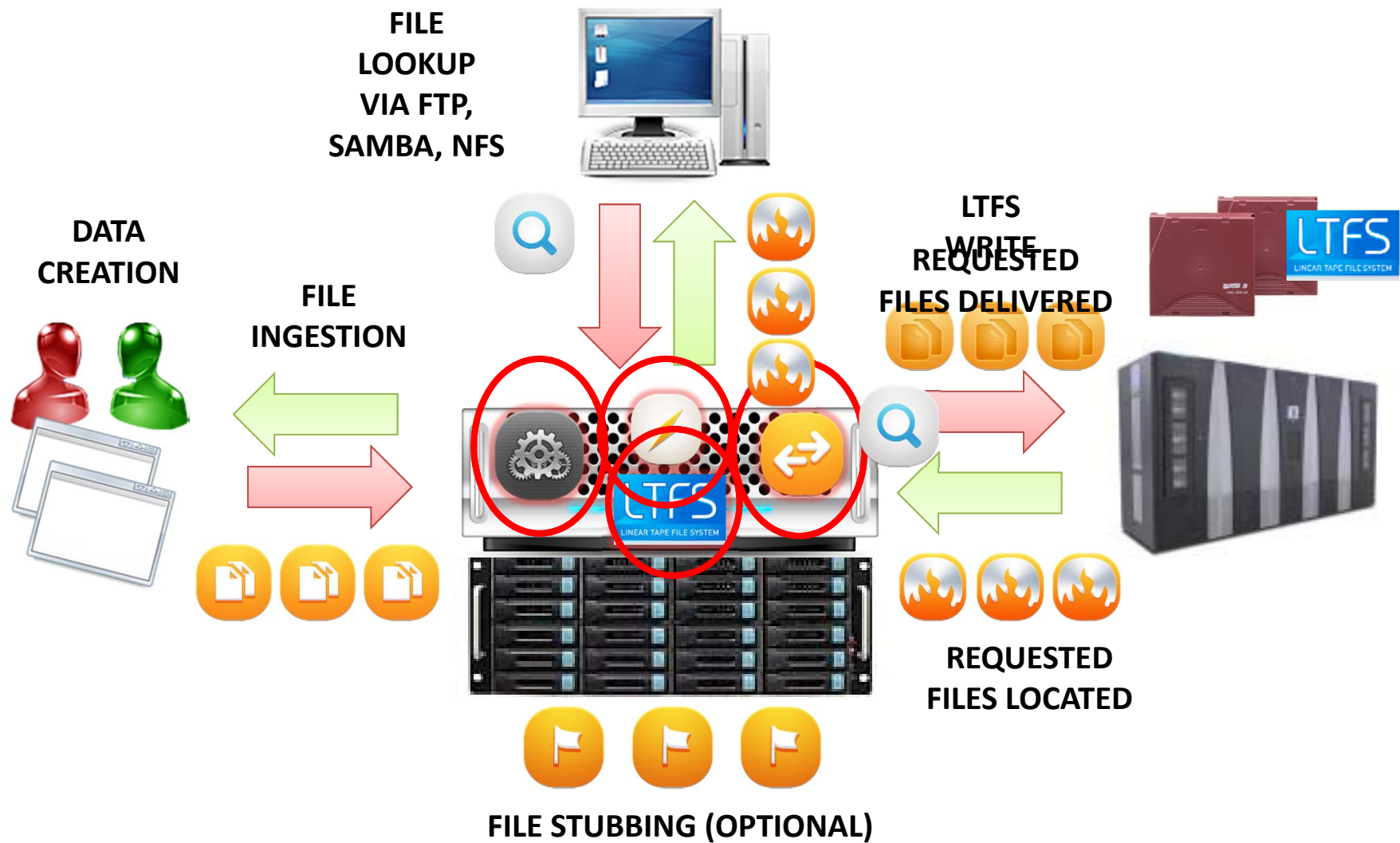- LTFS Library Edition (LE)
- LTFS Storage Manager (SM)

+

- Partition-capable Tape Media
- Partition-capable Tape Drives
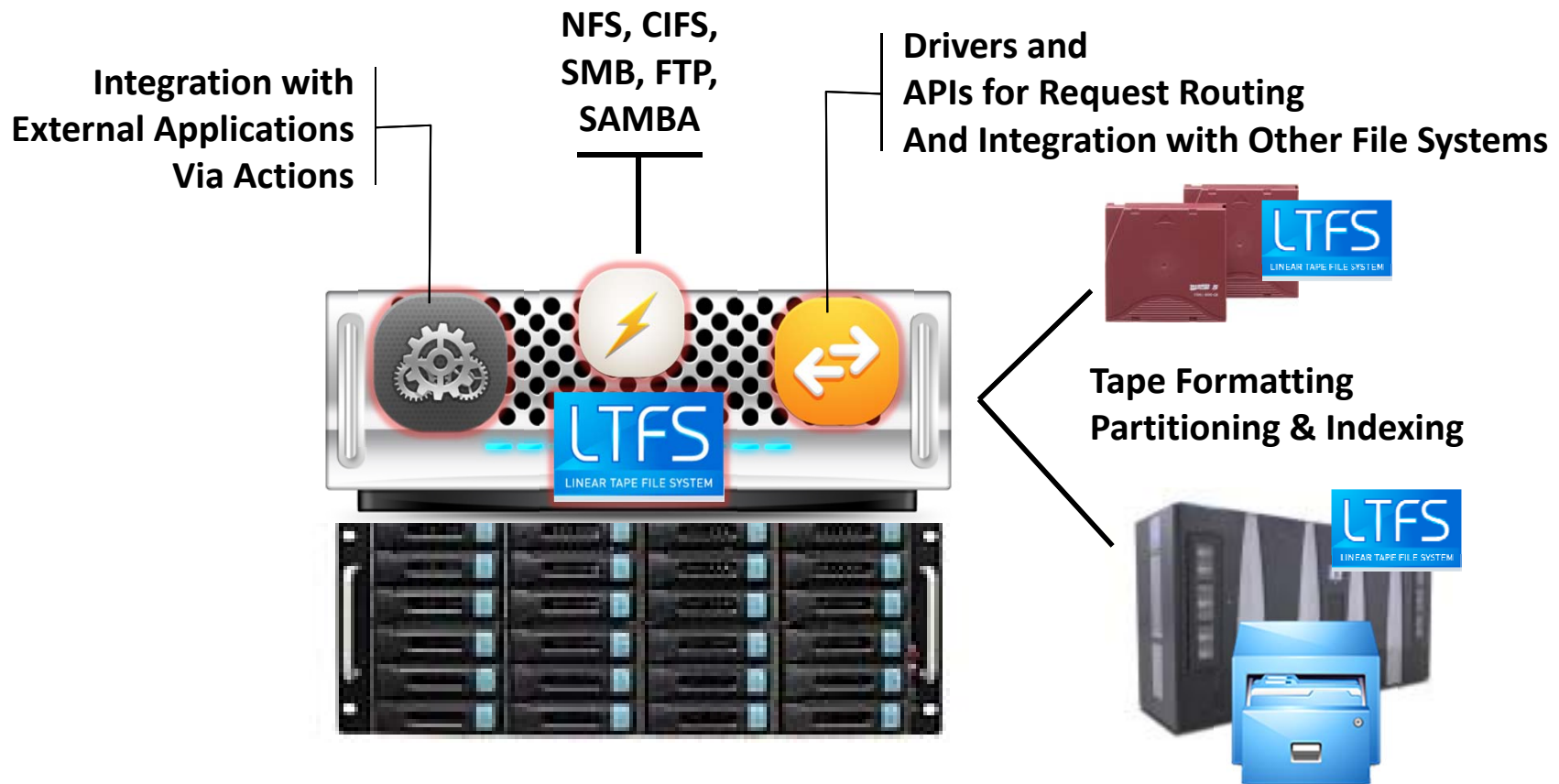- Server with Supported OS (Linux, Windows, Apple)

Basic Premise

FILE
LOOKUP
VIA FTP,
SAMBA, NFS

DATA
CREATION

FILE
INGESTION

LTFS
WRITE
REQUESTED
FILES DELIVERED

REQUESTED
FILES LOCATED

FILE STUBBING (OPTIONAL)

A Closer Look at the LTFS Front End Server

Integration with External Applications Via Actions

NFS, CIFS, SMB, FTP, SAMBA

Drivers and APIs for Request Routing And Integration with Other File Systems

Tape Formatting Partitioning & Indexing

For Your Inner Geek

- ## A simple LTFS volume

  - Index: a snapshot representation of the contents of the LTFS volume at a specific point in time…with generation numbers to indicate the latest changes to the tape volume…

  - Data partitions store the data in a series of extents following the volume label and LTFS label.
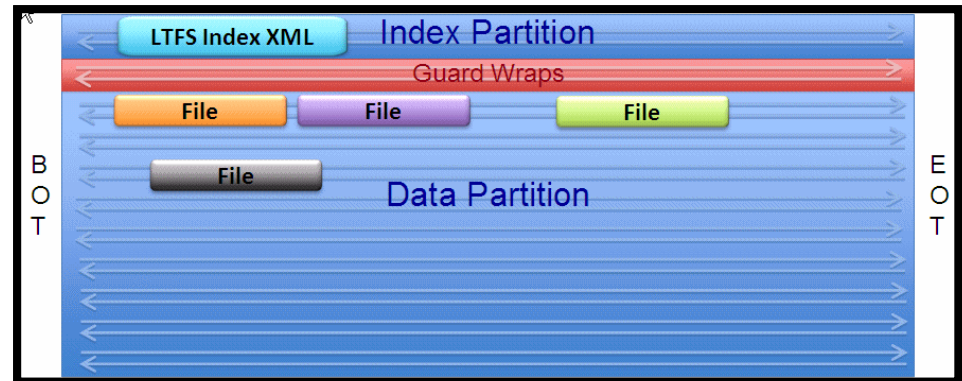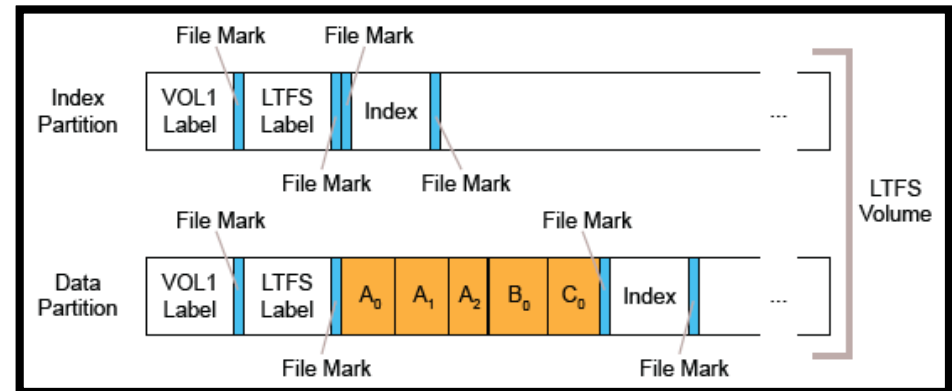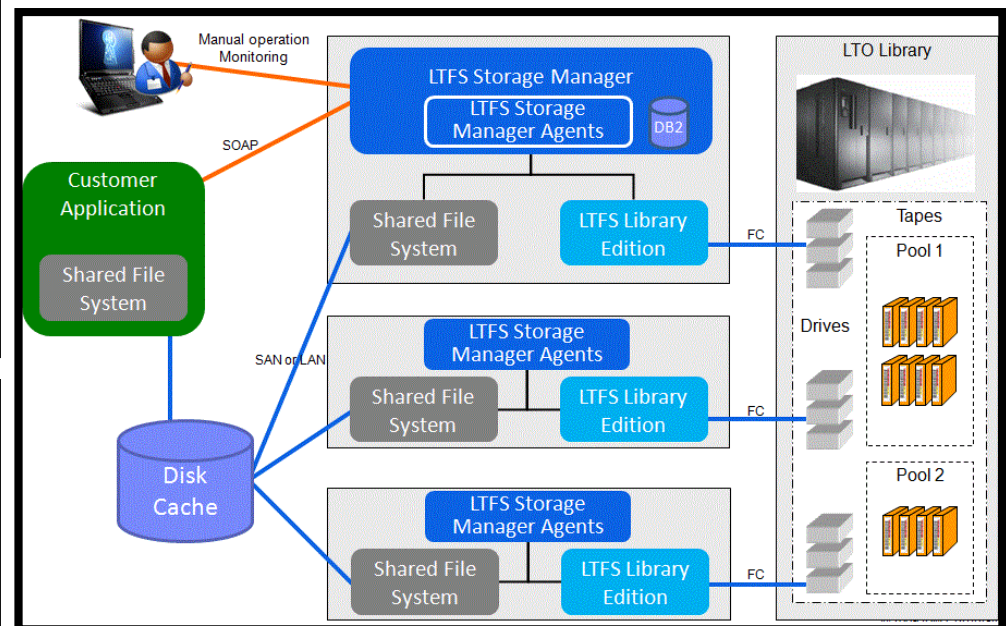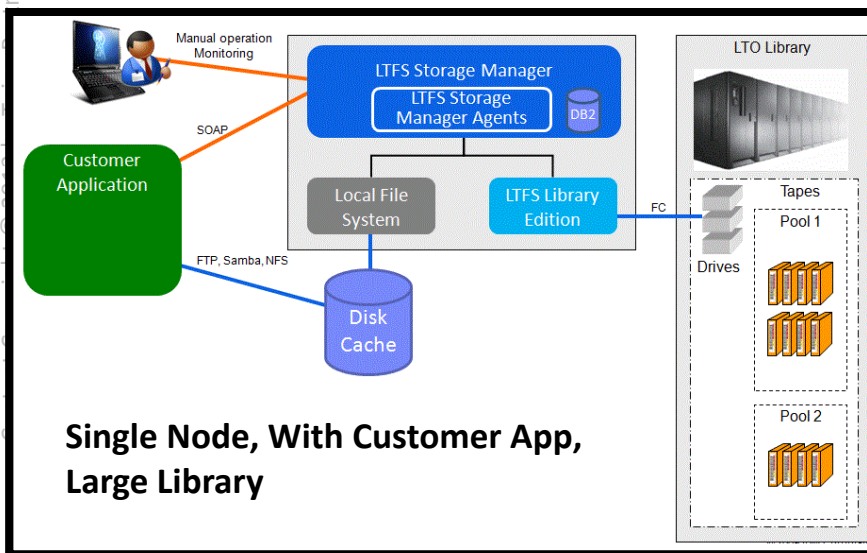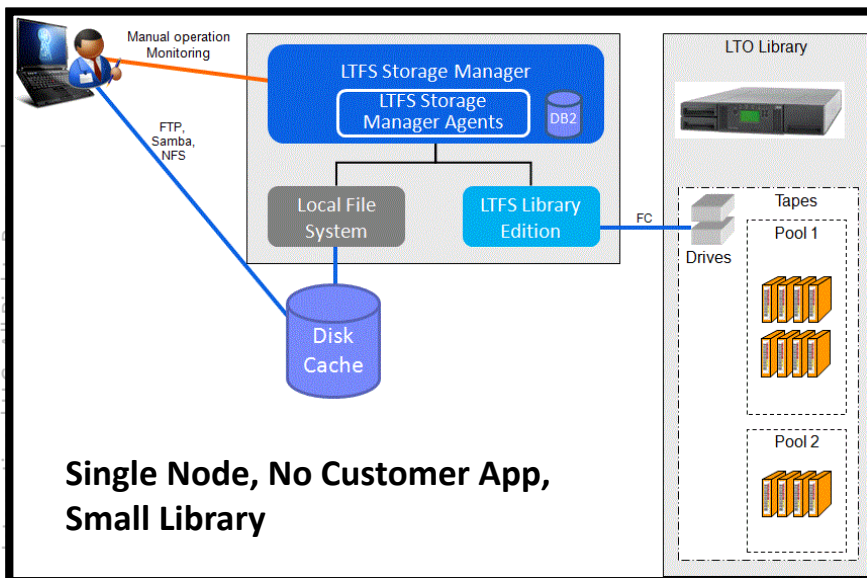

Illustrations from IBM LTFS Redbook sg248090.pdf

What the User Sees…

- Friendlier than a traditional tape content listing, but still…
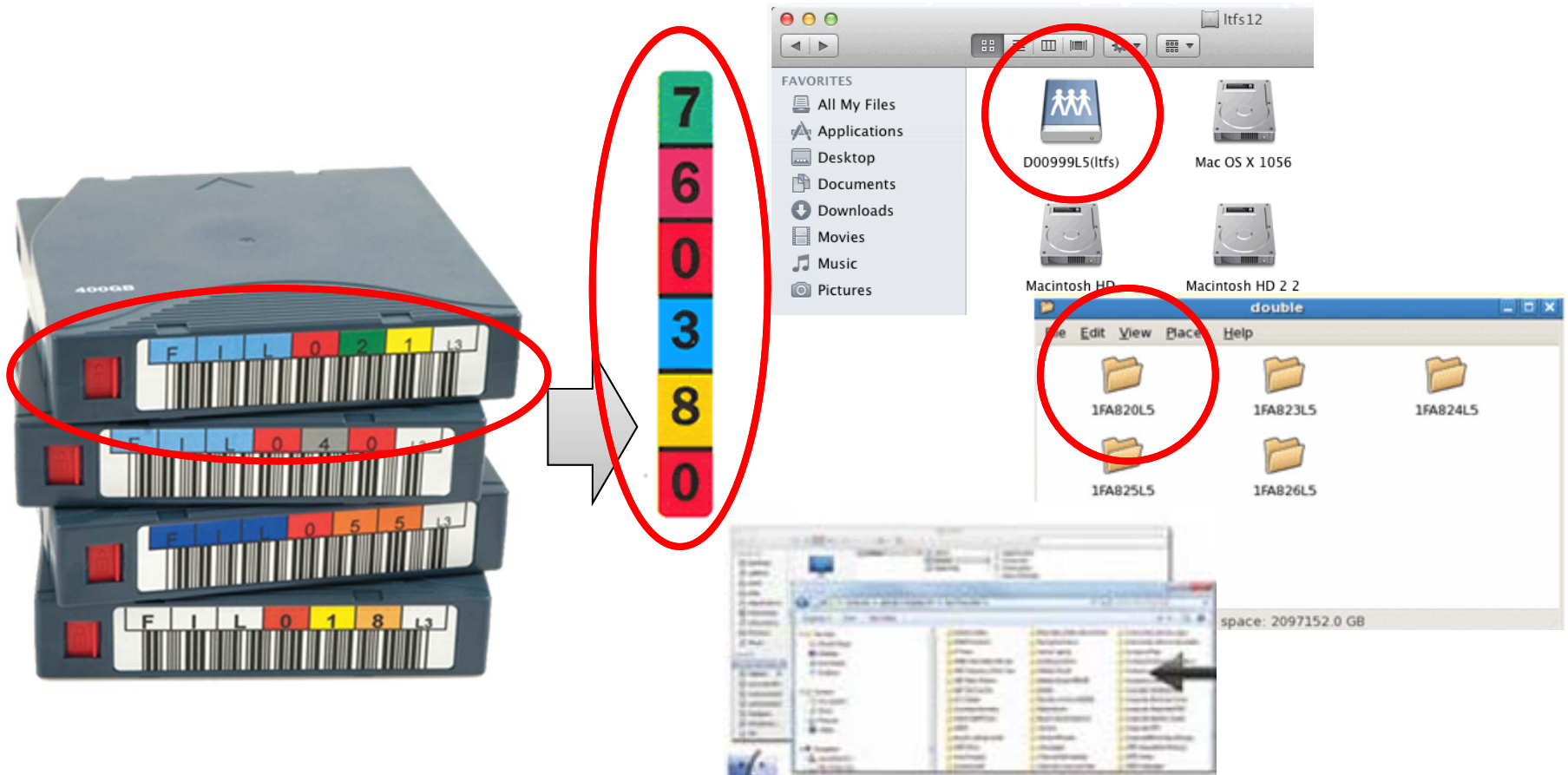
Implementation Examples from IBM

Single Node, No Customer App, Small Library

Single Node, With Customer App, Large Library

Multi Node, with Customer App, Scale-Out Library

But Sufficiently Open for Innovation and Enhancement

- Prefabricated LTFS Servers: Crossroads Systems StrongBox with user definable hard disk caching and optional media read/verify, library maintenance and management software

- Other vendors combining applications with LTFS ("Customer Application") such as archive software, media management software, picture archiving and communication systems, healthcare information management systems, etc.

- Still others writing data movers for automating the ingestion process

- *Unfortunately, though, some have tried to co-opt LTFS, forcing IBM to submit their code to SNIA for standardization...*

Why Do You Need File Caching on Disk?

- It takes time to
  - Find the file you are looking for
    - To locate the specific tape containing the directory and file
    - To grab the tape and place it into the drive
    - To advance the tape to the starting location of the file data and begin streaming it out
  - Between 20 seconds and two minutes (aka: the World Wide Wait)
  - That explains why the industry is positioning LTFS repositories as "archives"
- With file caching, user requests can be responded to immediately: cached data is sent in response while tape catches up…a bit of spoofing.

Ah hah!
See, that's something
you don't need to do
with disk!

Well, except for
all the memory you
buy for your favorite
NAS to make it perform
even reasonably well…

Slow-*ish* Find, Mount and Position Compensated for By Streaming Speed...

- Once data is in front of the read/write head, no storage streams data faster than tape...

LTO Ultrium 6

- Capacity: 2.5 TB native, 6.25 TB with 2.51 to 1 compression
- Average file access (excludes load/thread): 50 sec
- 8 Gbps FC, 6 Gbps SAS connectivity
- **Native sustained data rate (uncompressed) 160 MB/sec, 576 GB/hour**

Oracle T10000

- Capacity: T10000D data cartridge (BaFe): 8.5 TB, T10000 T2 sport cartridge: 1 TB (StorageTek T10000C)
- Average file access (excludes load/thread): 57 sec (17.5 sec for Sport Cartridge)
- **Native sustained data rate (uncompressed) 252 MB/sec, (compressed) 360 MB/sec**

TS1130 with IBM 3592 Media

- Recording technique: Linear Serpentine
- Number of tracks 1152
- Native capacity uncompressed: 1 TB (using JB/JX media), 640 GB (using JA/JW media), 128 GB (using JJ/JR media)
- **Native sustained data rate 160 MB/sec (uncompressed)**

Makes Tape Filestore Ideal for "Long Block Files.."

- Digital Video, for example – ask the media and entertainment businesses that have been the early adopters of LTFS

- But other opportunities exist, driven by data types:
  - Archive data for compliance and litigation
  - Medical patient data, files and images
  - Insurance claims and images, stock market transactions, banking records
  - Multi-media images, rich media, books, wikis, contracts
  - Security and digital surveillance video

IBM won an Emmy for LTFS

http://www.crossroads.com/pdf/sb/WhitePaper_LTFS_Takes_Tape_to_the_Next_Level%20_FredMoore.pdf

Tape Creates Additional Value as well

- ## As a tier in a storage hierarchy…

WHAT MOST FIRMS ARE DOING CURRENTLY

Per Fred Moore
Horison Information Strategies

Tier 3
$2.00
$0.20

Between 40-60% of data in tape libraries, off-site vaults: Archive, fixed content, compliance, reference data

Tier 2
$8.00
$1.00

B_____rrays fe_____using SAS and SATA:
_____a

Tier 1
$20.00
$7.00

Mission critical, OLTP, reven_____

Tier 0
$100.00
$50.00

Between 1-3% of data in SSD: I/O intensive, response time critical

http://www.horison.com/OracleTieredStorageTakesCenterStage.pdf
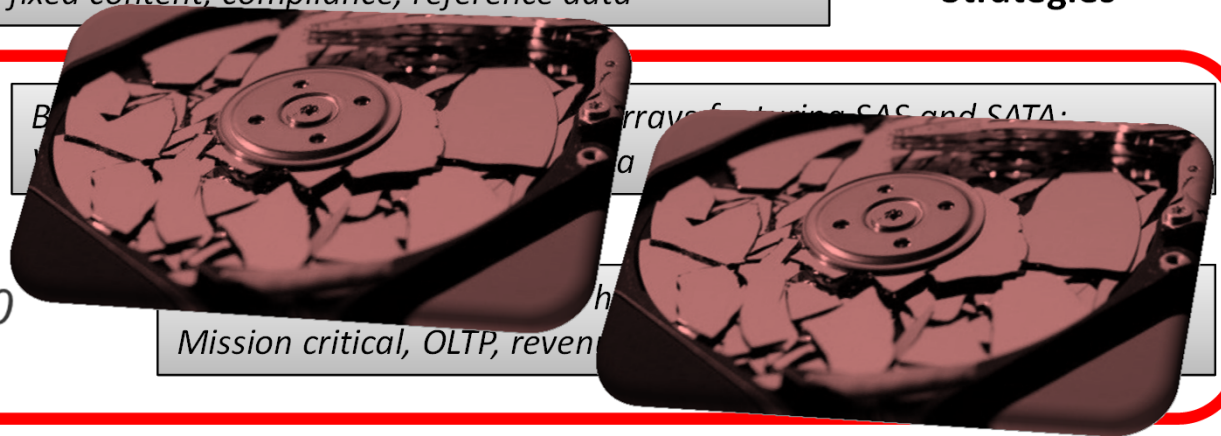
Doing the math…

- For a 100 TB storage complex…
  - Using only Tier 1 and 2: $765,000
  - Using Tiers 1 through 3: $359,250
  - Using Tiers 0 through 3: $482,250

  http://www.horison.com/OracleTieredStorageTakesCenterStage.pdf

**Per Fred Moore
Horison Information
Strategies**

Calculate for 212 Exabytes:

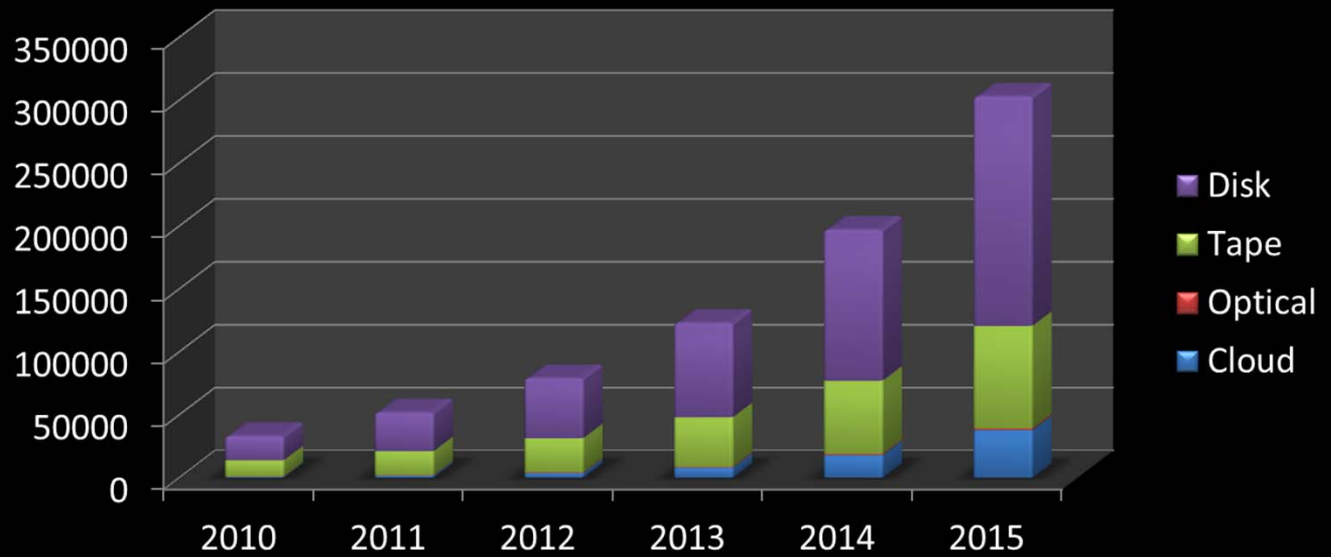**$1,022,370,000,000 (4 Tiers)
$1,621,800,000,000 (2 Tiers)**

*(Based on Gartner's projected 650% per annum
capacity demand growth rate –
WW external storage deployed
in shops with high rates of server virtualization)*

In any case, file data needs a Hosting Model that befits its access frequency

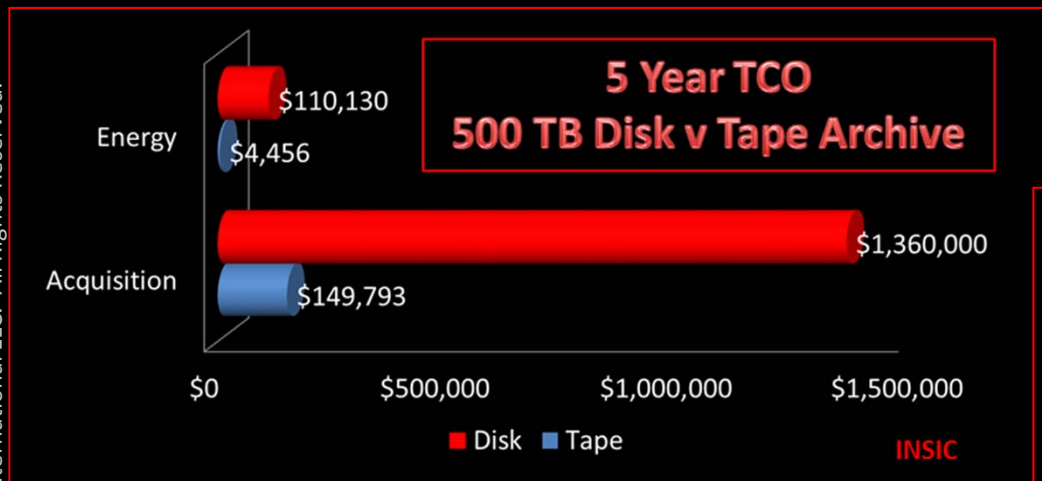- Archiving to disk is increasingly the norm, but makes little financial sense...

Enterprise Strategy Group

DEEP ARCHIVE

Cost of Ownership for Disk Archive is Nuts!

5 Year TCO
500 TB Disk v Tape Archive

Energy
- Disk: $110,130
- Tape: $4,456

Acquisition
- Disk: $1,360,000
- Tape: $149,793

Disk ■ Tape

INSIC

12-Year TCO Analysis

TCO = $67,193,181 (DISK)
TCO = $4,458.228 (TAPE)

Legend:
- ■ Energy
- ■ Floor Space
- ■ Equip/Media/Maint

Clipper Group

"The 12-year cost of energy for disk is greater than the entire TCO for tape over the same period: **Disk is 500 times more costly than tape, based on energy cost alone.***"*

http://www.insic.org/news/A&S%20Roadmap.pdf
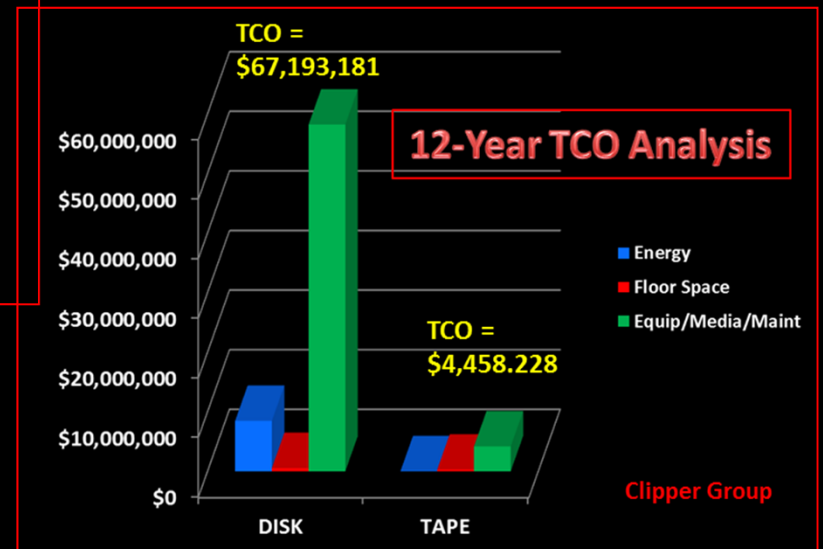
Think of the Density...

- ## And for those who care
  - Tape delivers greater density in a smaller space than any other media (except for DNA*)
  - Significant reduction in capacity per watt over disk and solid state

*2.2 Petabytes per gram of DNA achieved, per http://hardware.slashdot.org/story/13/01/23/1927221/ researchers-achieve-storage-density-of-22-petabytes-per-gram-of-dna

Given the Pace of Storage Capacity Expansion, with No Commensurate Budgetary Increases…

TAPE IS MAKING ITS WAY…

BACK TO THE FUTURE

Questions?

Thanks.

- ## Keep in Touch…

  - **jtoigo@toigopartners.com**

  - **@JonToigo**

  - **www.drunkendata.com**

  - **www.toigopartners.com**

  - **www.datainstitute.org**

Next time: Perhaps a Star Trek Theme?