

Storage Growth, Technologies, and Trends

- ◆ Growth
- ◆ Tape vs. Disk
- ◆ VTL
- ◆ WORM
- ◆ MAID
- ◆ JBOD
- ◆ Storage Farm
- ◆ Compression
- ◆ De-duplication
- ◆ Replication
- ◆ Virtualization
- ◆ Archiving

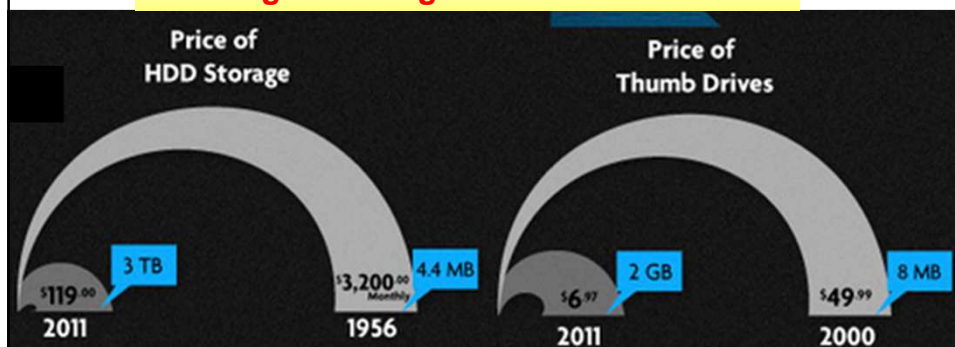
- ☐ Faster data access
- ☐ Improved energy efficiency
- ☐ Reduced backup and restore times
- ☐ Reduced costs and storage capacity

Page 1

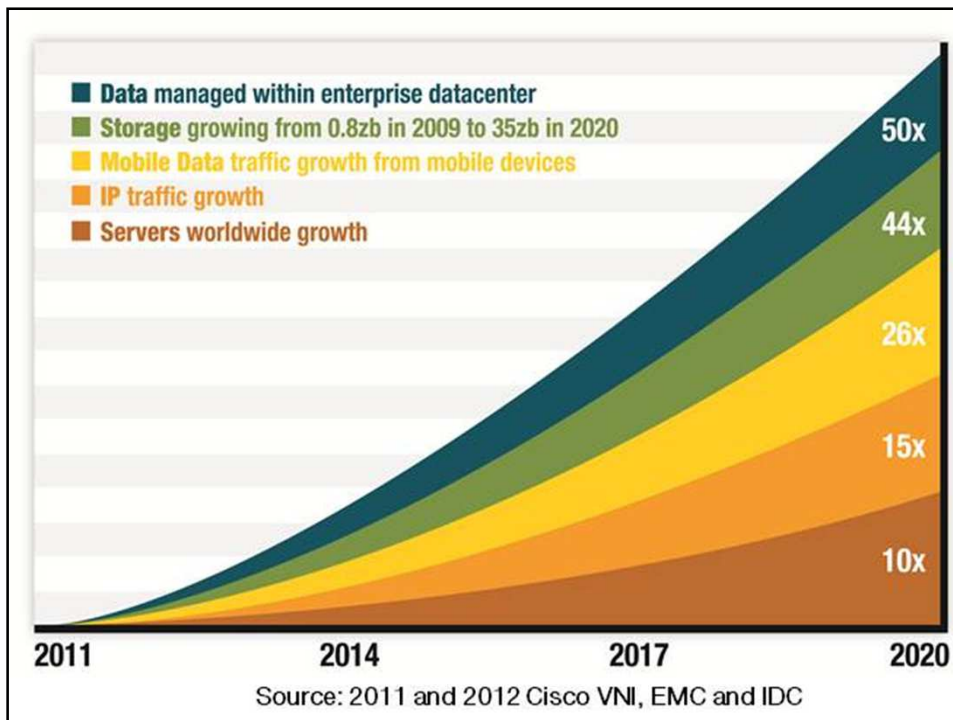
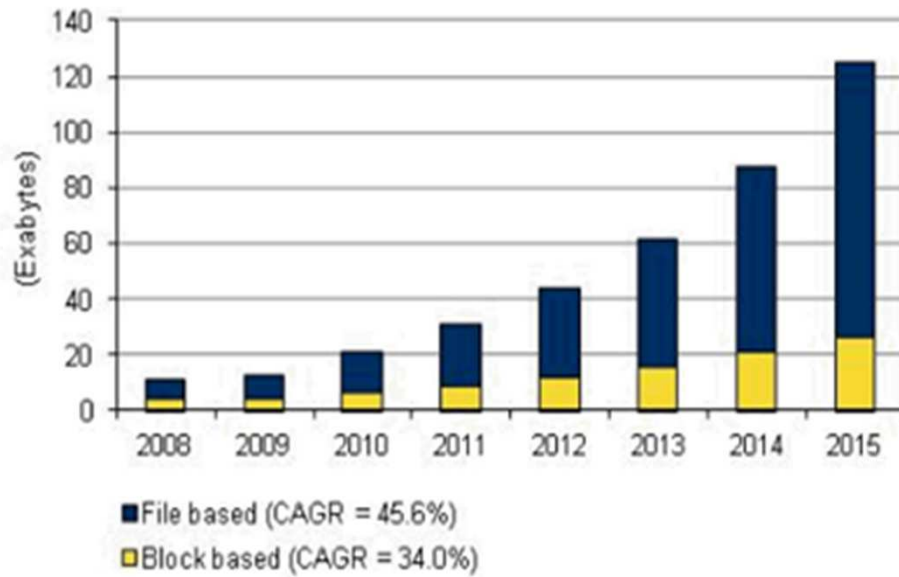
Storage Trends and Growth

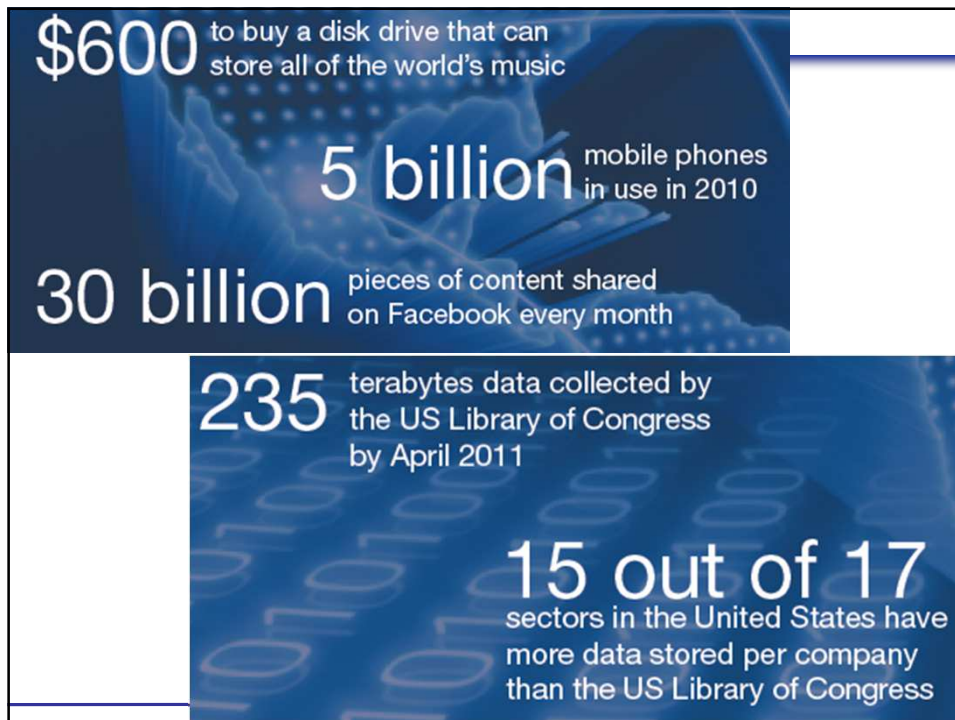
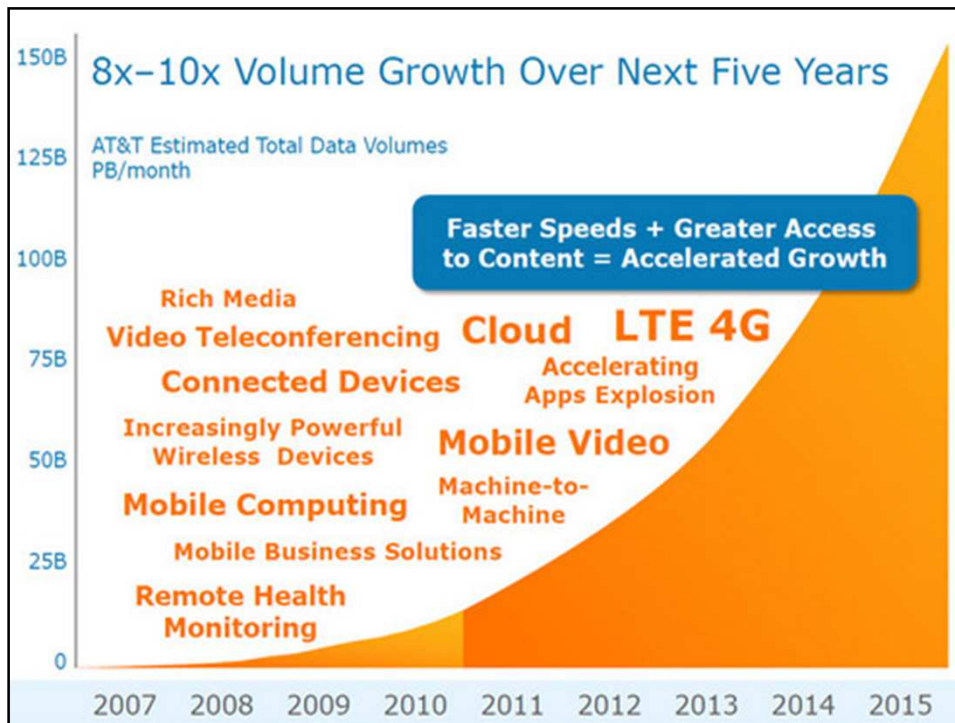
- | | |
|--|---|
| <ul style="list-style-type: none"> ◆ Technology <ul style="list-style-type: none"> ■ Faster ■ Cheaper ■ Better | <ul style="list-style-type: none"> ◆ Business/User Experience <ul style="list-style-type: none"> ■ Easier ■ More choices <ul style="list-style-type: none"> ◆ Physical ◆ Virtual ◆ Cloud |
|--|---|

Challenges: Management and Protection



Worldwide File-Based Versus Block-Based Storage Capacity Shipments, 2008–2015





What Do You Store?

Everyone's needs are different, but basically we can break down storage based on types of media consumed on a regular basis.

- ◆ Photos
- ◆ Audio
- ◆ Video
- ◆ Documents or raw data
- ◆ Software programs such as applications, games, etc...



Page 7

Putting it into Perspective (1 TB)

- ◆ **Photos**
 - We can store more than 1 million 2.5 Megapixel high-resolution photos
- ◆ **Audio**
 - We can store roughly 1,500 hours of CD-quality Audio
 - With MP3 or MP4 compression that can translate to around 5,000 hours.
- ◆ **Video**
 - Similarly that's around 500 hours of DVD quality video or about 250 DVD movies.
- ◆ **Software**
 - We can also store hundreds of computer programs and games.
- ◆ **Documents**
 - You can potentially store several million documents with this amount of space. Considering that documents vary greatly in size – depending of course on how much data you store in each document – we can still safely assume that if the medial person stores no more than a few megabytes of data per document on average this gives us plenty of room to store an archive of documents for an entire lifetime

Page 8

How Many Songs?

Approximately how many songs have been produced in the US and UK since the music industry was born?

- ◆ On Average, there are 50 albums released every week in the United States of America and the United Kingdom.
- ◆ Thomas Edison invented the phonograph in 1878 but it was *always* a very poor seller. The first artist to sell serious numbers of records was [Frank Sinatra](#) in the 1940 (250,000,000 records in his lifetime), so lets start there!
- ◆ Lets also assume that there are 12 songs per album.
 $50 \text{ albums per week} * 52 \text{ weeks} = \underline{2,500 \text{ per year}}$
 $2,500 \text{ per year} * 71 \text{ years} = \underline{177,500 \text{ unique albums}}$
 $12 \text{ songs} * 177,500 \text{ albums} = \underline{2,130,000 \text{ songs}}$
- ◆ Appropriately 5 MB per song (.mp3) → **1.065 TB**



\$? To buy a disk drive that can store all of the world's music?

COSTCO WHOLESALE [Click here for warehouse](#)

Order by Item Number Order Sta


SHOP ALL DEPARTMENTS SEARCH: What do you want to find?

Home Computers & Printers Hard Drives & Memory External Hard Drives Compare up to 4 p

4.0TB

DESKTOP

USB



HGST Touro™ Desk Pro 4 TB External Drive by HGST

☐ Compare product

Rated ★★★★★ (out of 5 reviews)

Share this Product: [f](#) [t](#) [p](#) [e](#)

Item #696489

Features:
4 TB, USB 3.0, 7200 RPM, For PC & Mac.
Two Levels of Protection for Your Most Valued Information

Your Price **\$189.99**

Shipping & Handling included *

QUANTITY

[Add to Cart](#)

The estimated delivery time will be approximately 3 - 5 business days from the time of

□ Disk vs. Tape

- ◆ Tape is not as cheap (*and disk is not as expensive*) as you might think
 - Must include price of drives, slots and tapes.
 - Midrange tape libraries range from \$4/GB to \$11/GB
 - Disk prices range from \$3/GB to \$11/GB
- ◆ You can do neat things with disk
 - Incremental backups are one of the greatest backup performance challenges
 - Use as a target for all incremental backups. (Full, too, if you can afford it.)
- ◆ For offsite storage, duplicate all disk-based backups to tape
- ◆ *Leave disk-based backups on disk if possible*



Page 11

Trade-Offs Between Tape and Disk

◆ Disk based recovery

- Is an on-line recovery
- Instant access
- Allows for hitting much tighter RTOs and RPOs



◆ Tape based recovery

- Is an off-line recovery
- Must load, configure and read the data – before it can be accessed
- Lengthy process

◆ Through-put rates

- Read from disk vs. read from tape

◆ Removable media

- A tape can be taken out of the library and taken to the recovery site
- A tape can be dropped, broken and lost – and you may not know it

◆ Ability to store vs. ability to access

- Tape capacities increased 10 fold, but access rate increased 3 fold

◆ Consider the “Value of Your Data”

Page 12

Tape Rotation: Grandfather-Father-Son

- ◆ The daily and weekly Full backups reflect the end of each period.
- ◆ Business Requirements:
 - The Sales department needs to restore the database back to the end of the previous month.
 - The Sales department needs to restore the database back to any given week ending within the current month.
 - The Sales department also wants the flexibility to restore the database back to any daily ending of current week.
 - What is the **minimum number of tapes** do you need to budget for 2 months (assume 4 weeks per month)?

	MON	TUE	WED	THU	FRI	SAT	SUN
8:00 PM							
9:00 PM							
10:00 PM							
11:00 PM	FULL	FULL	FULL	FULL	FULL	FULL	FULL

Page 13

How many tapes do I need for 2 months?

	MON	TUE	WED	THU	FRI	SAT	SUN
Week 1	M	T	W	T	F	S	WK1
Week 2	M	T	W	T	F	S	WK2
Week 3	M	T	W	T	F	S	WK3
Week 4	M	T	W	T	F	S	Month 1

	MON	TUE	WED	THU	FRI	SAT	SUN
Week 5	M	T	W	T	F	S	WK1
Week 6	M	T	W	T	F	S	WK2
Week 7	M	T	W	T	F	S	WK3
Week 8	M	T	W	T	F	S	Month 2

11 Tapes: Month1, Month2, WK1, WK2, WK3, M, T, W, TH, F, S

Page 14

Bank of America, Time Warner, Ameritrade, and Citigroup;

What do they have in common?

Page 15

COMPUTERWORLD An IDG company Quicklink Search Computerw

[Home](#) [News](#) [Topics](#) [Print Edition](#) [Services](#) [Subscribe](#) [Events](#) [In Depth](#) [Blogs](#) [XML](#) [Fe](#)

[Management](#) [Careers](#) [Security](#) [Hardware](#) [Software](#) [Data Mgmt](#) [Networking](#) [Government](#) [Mobile](#) [Development](#) [Industr](#)

Home > Browse Topics > Data Management

Bank loses credit card info of 1.2M federal workers

Tapes disappeared while being shipped to a backup center

News Story by Joanne Morrison

FEBRUARY 26, 2005 (REUTERS) - WASHINGTON -- Computer tapes containing credit card records of U.S. senators and more than a million U.S. government employees are missing. **Bank of America Corp.** said yesterday, putting the customers at increased risk of identity theft.

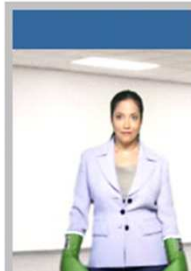
The security breach, which included data on a third of the Pentagon's staff, angered lawmakers already concerned after criminals gained access to thousands of consumer profiles in a database maintained by a data profiling company, ChoicePoint Inc. ([see story](#))

Bank of America Corp. didn't release details of how the tapes were lost, but Sen. Charles Schumer (D-N.Y.) said he had been informed by the Senate Rules Committee that the data tapes were likely stolen off a commercial plane by baggage handlers.

[Print-friendly](#) [E-n](#)

Related to this topic

- > [Explainer: Data mining](#)
- > [Apache group releases upg](#)
- > [Oracle looks to ease conter](#)
- > [Archive and backup: What's](#)
- > [Start-up to index e-mail, oth](#)



COMPUTERWORLD
An IDG company

QuickLink Search Computerworld

[Home](#) [News](#) [Topics](#) [Print Edition](#) [Services](#) [Subscribe](#) [Events](#) [In Depth](#) [Blogs](#) [XML Feeds](#)

[Management](#) [Careers](#) [Security](#) [Hardware](#) [Software](#) [Data Mgmt](#) [Networking](#) [Government](#) [Mobile](#) [Development](#) [Industry](#)

Home > Browse Topics > Security > Privacy

Ameritrade warns 200,000 clients about potential data breach

A backup tape with account information is missing

News Story by [Todd R. Weiss](#)

APRIL 20, 2005 (COMPUTERWORLD) - A computer backup tape containing account information of more than 200,000 Ameritrade Inc. clients was apparently lost or accidentally destroyed while being shipped, prompting the online investment brokerage to notify the clients of a potential breach.


Donna Kush, a spokeswoman for the Omaha-based company, today confirmed that a package of data backup tapes was damaged in transit in late February by a shipping company that isn't being named. Four of the tapes in the package disappeared after the package was damaged but three were later found by the shipper during a search of its facility, she said.

The fourth tape is still missing and is presumed to still be lost in the facility or to have been destroyed accidentally.

Print-friendly E-mail this

Related to this topic

- > Antispyware firm warns of massive ID theft ring
- > Court orders CardSystems to retain breach inform
- > Gartner: Bank card fraud too easy for phishers
- > Life beyond passwords
- > 'Car Whisperer' puts hackers in the driver's seat



Free E-Newsletters

Keep up on technology news and trends with our free e-mail newsletters! Select from daily and weekly updates -- including alerts and roundups by topic.

→ Sign up today!

Knowledge Centers

- Security
- Storage
- Mobile & Wireless
- Hardware
- Networking
- Software

More Knowledge Centers:

Select one...

Partner Zones

Blade-Ready Data Center
Making Storage Simple

Features

- Latest Headlines
- This Week's Issue

COMPUTERWORLD
An IDG company

QuickLink Search Computerworld

[Home](#) [News](#) [Topics](#) [Print Edition](#) [Services](#) [Subscribe](#) [Events](#) [In Depth](#) [Blogs](#) [XML Feeds](#)

[Management](#) [Careers](#) [Security](#) [Hardware](#) [Software](#) [Data Mgmt](#) [Networking](#) [Government](#) [Mobile](#) [Development](#) [Industry](#)

Home > Browse Topics > Data Management

Missing Backup Tapes Spur Encryption at Time Warner

Data security boost follows loss of info on 600,000 employees

News Story by [Lucas Mearian](#)

MAY 09, 2005 (COMPUTERWORLD) - Time Warner Inc. last week said it will "quickly" begin encrypting all data saved to backup tapes, after 40 tapes with personal information on about 600,000 current and former employees were lost in transit to a storage facility.

The incident is among the biggest in a string of recent data-security mishaps that have also affected companies such as ChoicePoint Inc., Bank of America Corp. and Reed Elsevier Group PLC's LexisNexis Group unit.

A shipping container that held the 40 data tapes was lost on March 22, Time Warner spokeswoman Kathy McKiernan said. The tapes went missing during a routine shipment to an off-site facility by records management and storage firm Iron Mountain Inc. McKiernan wouldn't provide more details.

However, McKiernan did say Time Warner is trying to convince officials at Boston-based Iron Mountain to change some of their handling procedures. She declined to expand

Print-friendly E-mail this

Related to this topic

- > Explainer: Data mining
- > Apache group releases upgrade to De
- > Oracle looks to ease content manager
- > Archive and backup: What's the differ
- > Start-up to index e-mail, other unstruc



Free E-Newsletters

Keep up on technology news and trends with our free e-mail newsletters! Select from daily and weekly updates -- including alerts and roundups by topic.

→ Sign up today!

Knowledge Centers

- Security
- Storage
- Mobile & Wireless
- Hardware
- Networking
- Software

More Knowledge Centers:

Select one...

Partner Zones

Blade-Ready Data Center
Making Storage Simple

Features

- Latest Headlines
- This Week's Issue
- Shark Tank
- Research
- Webcasts
- White Papers

COMPUTERWORLD
An IDG company

QuickLink


[Home](#) | [News](#) | [Topics](#) | [Print Edition](#) | [Services](#) | [Subscribe](#) | [Events](#)

[Management](#) | [Careers](#) | [Security](#) | [Hardware](#) | [Software](#) | [Data Mgmt](#) | [Networking](#) | [Governance](#)

Home > Browse Topics > Hardware

Citigroup says UPS lost its data

Information on 3.9M customers lost in transit

News Story by Jonathan Stempel

JUNE 06, 2005 [\(REUTERS\)](#) - NEW YORK -- Citigroup Inc., the world's largest bank, today said account and payment history data on 3.9 million of its customers was lost in transit by United Parcel Service Inc.

The disappearance is the latest in a series of data breaches involving U.S. banks, including the second-largest, Bank of America Corp. [\(see story\)](#).

New York-based Citigroup said the data was stored on computer tapes and lost while UPS, the world's biggest package carrier, was shipping them to an Experian credit bureau in Texas.

The tapes, which also contained Social Security numbers, covered CitiFinancial Branch Network customers and about 50,000 customers with closed accounts from CitiFinancial Retail Services. Customers of CitiFinancial Auto and CitiFinancial Mortgage are unaffected.

Free E-Newsletters
Keep up on technology news and trends with our free e-mail newsletters! Select from daily and weekly updates -- including alerts and roundups by topic.
→ Sign up today!

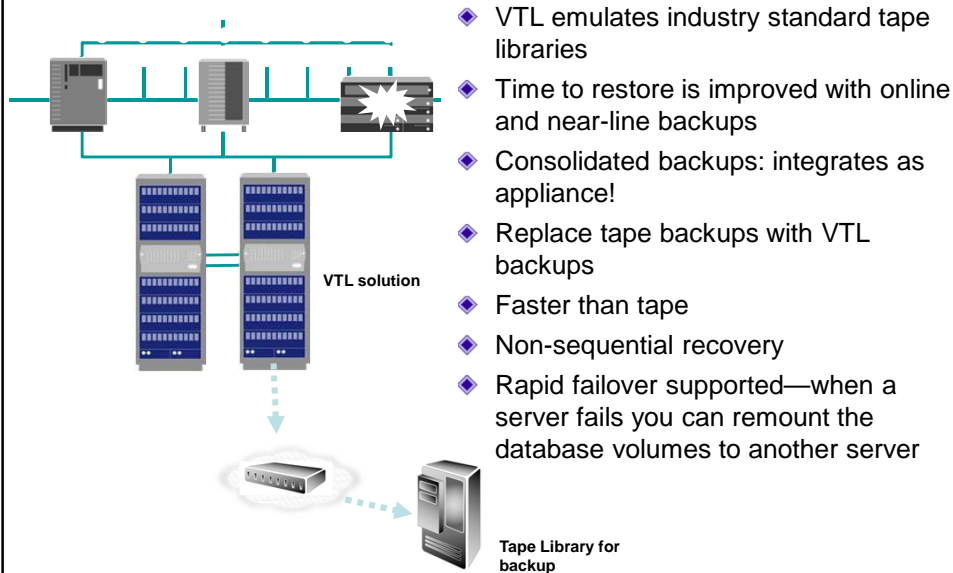
Knowledge Centers
Security
Storage
Mobile & Wireless
Hardware
Networking



10 rules for better backups

1. Mirroring does **not** replace backups.
2. The most common use of restores isn't after a catastrophe.
3. Regularly test your ability to restore.
4. Keep those tape heads clean.
5. Beware of dirty tapes.
6. Tapes decompose over time.
7. Make sure you can still read old media.
8. Make two copies of critical tapes.
9. Vault one copy of your backup off-site.
10. Encrypt sensitive data on backup tapes.

❑ VTL: Efficient Backup and Restore



Page 21

❑ WORM

- ◆ **Write Once, Read Many** (alternatively **Write One, Read Multiple** or **Write Once, Read Mostly** or **WORM**) refers to computer data storage systems, data storage devices, and data storage media that can be written to once, but read from multiple times.
- ◆ **WORM is useful for immutable versioning and secure logging, where you wish to record information that an attacker can not modify.**
- ◆ One common WORM media is a **CD-R** or **DVD-R**, which can not be rewritten, but CD-Rs are slow and do not allow data to be appended.
- ◆ **WORM tapes** may be appended to, but do not allow fast random access.
- ◆ **WORM disk drives**, are disk drives with modified device drivers or firmware that does not permit data to be re-written. The advantage of WORM disk drives is that random reads and appends are both fast operations.
- ◆ A preferred solution to meet Archive and Compliance requirements



Page 22

SanDisk WORM Write-Once SD Cards Can't Be Altered, Last 100 Years



SanDisk has created the first write-once SD memory card after over a year of **talking about it**. The WORM (Write Once Read Many) cards cannot be altered or deleted and are designed for information that must be kept intact, such as electronic voting records and police work. They are only 128MB for now, but bigger sizes will likely show by the end of the year. SanDisk claims that the stored information will last 100 years, so if McCain is elected president, you'll have something to send your great grandchildren in Iraq. Pricing available upon request—yikes!—press release down below.

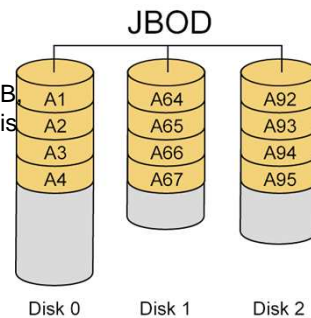
MAID

- ◆ In computing, a **massive array of idle disks** (more commonly known as a **MAID**) is a system using hundreds to thousands of hard drives for near-line data storage.
- ◆ MAID is designed for **Write Once, Read Occasionally (WORO)** applications.
- ◆ In a MAID each drive is only spun up on demand as needed to access the data stored on that drive.
- ◆ Compared to RAID technology a MAID has increased storage density, and decreased cost, electrical power, and cooling requirements.
- ◆ These advantages comes at the cost of much increased latency, significantly lower throughput, and no or lower redundancy.
- ◆ Most large hard drives are designed for near-continuous spinning; their reliability will suffer if spun up repeatedly to save power.
- ◆ MAID can supplement or replace tape libraries in hierarchical storage management.



❑ JBOD

- ◆ JBOD stands for **Just a Bunch of Disks**.
- ◆ The controller treats each drive as a stand-alone disk, therefore each drive is an independent logical drive.
- ◆ JBOD does not provide data redundancy.
- ◆ **Concatenation** or **Spanning** of disks is not one of the numbered RAID levels, but it is a popular method for combining multiple physical disk drives into a single virtual disk. It provides no data redundancy.
- ◆ As the name implies, disks are merely concatenated together, end to beginning, so they appear to be a single large disk.
- ◆ Concatenation is sometimes used to turn several odd-sized drives into one larger useful drive, which cannot be done with RAID 0.
- ◆ For example, JBOD could combine 3 GB, 15 GB, 5.5 GB and 12 GB drives into a logical drive at 35.5 GB, which is often more useful than the individual drives separately.



Page 25

❑ The Storage Farm

- ◆ The Storage Farm (disk space) is a centrally managed pool of over a large amount of space shared among all storage farm servers - *The storage is always growing.*
- ◆ Storage in the Storage Farm is typically provided by SAN or NAS technology.
- ◆ Organizations achieve economies of scale by purchasing storage in large volume.
- ◆ Systems Support can offer more fully featured storage space with a wider range of availability and performance characteristics.

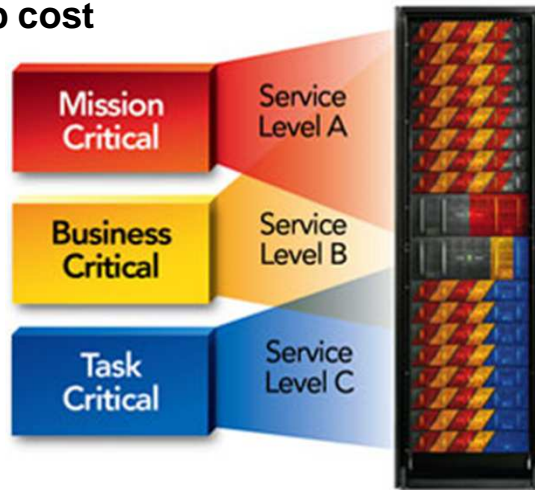
Benefits of the Storage Farm

- ◆ **Professional Storage Management:** The Storage Farm staff performs all required firmware updates and maintenance procedures, monitors availability and performance, and assists with procurement and life-cycle upgrades.
- ◆ **Smart Purchase:** Using the Storage Farm means you can purchase the amount of space you need now and expand if and when your needs change. You don't need to purchase extra space until you're sure you need it.
- ◆ **Risk Reduction:** Storage Farm equipment has redundant components such as power supplies and boot drives.
- ◆ **Easy Migration:** If your storage needs change, you can move your data to a different storage tier. For example, if you add a critical application to your server, you can upgrade to Tier 1 storage.

Page 26

Tiered Storage Offering

- ◆ Increase utilization rates
- ◆ Reduce administrative time
- ◆ Lower ownership cost
- ◆ Save energy
- ◆ Virtualization



What about storage



Shared Storage Offerings - Service Characteristics by Offering

Service Characteristics	Mission Critical Data	Business Critical Data	Corporate Data	Reference Data	Backup & Archive Data
Availability	99.999%	99.99%	99.95%	99.9%	Offline
Performance	☆☆☆☆☆	☆☆☆☆	☆☆☆	☆☆	☆
Compliance Capability					
Recovery Point Objective	< 1 hour	< 24 hours	< 24 hours	< 7 days	N/A
Recovery Time Objective	< 1 hour	< 24 hours	24 - 72 hours	72+ hours	N/A
Continuity Management Recovery Classifications	Immediate or Near Immediate	Intermediate	Intermediate	Gradual	N/A
Cost	\$\$\$\$\$\$	\$\$\$\$	\$\$\$	\$	\$
Examples	Sales	E-Mail SAP	Development & Test systems	Image, Workflow, Workspace	Sales History

Corporate Compliance

Regulatory Compliance

Page 28

Example

Storage Farm Tiers and Classes

Tier Definitions and Rates:

Three Tiers of SAN-based storage are offered, as follows:

Tier 1: Highest performance and availability, while low cost is less important. Examples of this would include high-usage operational databases and applications that are critical to the daily operation of the University. This is the only tier providing connectivity to the IBM mainframe.

Tier 3: Lowest cost of online storage, while performance and high availability are less important. Examples of this include development & test systems, and low-activity departmental file servers.

Tier 2: Mixtures of the above, where performance, cost and availability all have some level of importance. Examples of this would include lower usage databases and applications, or applications that are important, but not critical, to the operation of the University.

Rates (Monthly fees):

The rates below are effective beginning July 1, 2007:

Storage Management Charge (per system):	\$64.00
Fibre Channel Connection Charge (per connection):	\$52.00
Tier 1 (\$/GB-mo):	\$3.43
Tier 2 (\$/GB-mo):	\$0.49
Tier 3 (\$/GB-mo):	\$0.29

Page 29

Green Storage

- ◆ **What it is:** The use of storage technologies to reduce the amount of power consumed by enterprises.
- ◆ **Why it's important:** Companies are going green in their servers and in their data centers, so why not in their storage subsystems?



Page 30

Adoption of SSD

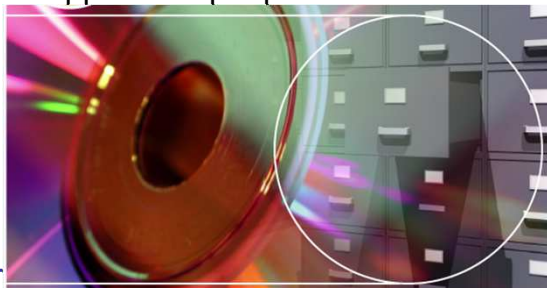
- ◆ **What it is:** SSD, or solid-state disk, refers to memory used to store data, be it flash or DD RAM. The technology responds faster and consumes less power than standard hard drives.
- ◆ **Why it's important:** Better performance and lower power consumption than disk.
- ◆ **The catch:** Cost. SSDs are more expensive than most hard-drive technologies right now.



Page 31

Application-Focused Storage

- ◆ **What it is:** Storage architectures managed by the business or infrastructure applications, not the SAN.
- ◆ **Why it's important:** Application-focused storage shows three major benefits:
 - Better storage SLA design
 - Improved likelihood of tiering across storage systems
 - The ability to empower application groups to manage storage resources.
- ◆ **The catch:** Storage professionals could be seeing their critical resources taken over by the application people within the organization.



Data Deduplication

- ◆ **What it is:** Data deduplication technologies eliminate redundant data at the subfile level from backups.
- ◆ **Why it's important:** Data deduplication can reduce the amount of storage consumed by as much as 20- or 30-to-1.
- ◆ **The catch:** For one, regulatory restrictions require that data be stored unaltered; data deduplication technically violates these mandates by replacing the duplicated data with pointers to the original data. Also, data deduplication may not be effective in all scenarios.



Data deduplication

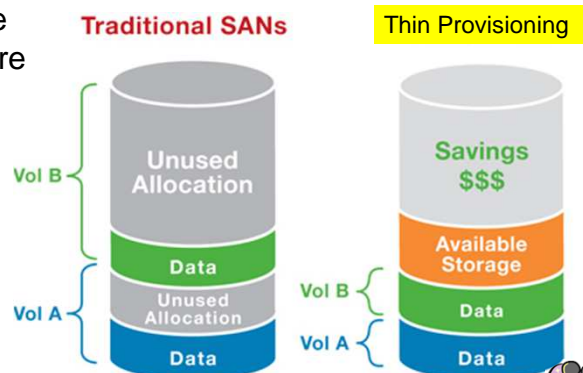
- ◆ Data deduplication (often called "intelligent compression" or "single-instance storage") is a method of reducing storage needs by eliminating redundant data.
- ◆ Only one unique instance of the data is actually retained on storage media. Redundant data is replaced with a pointer to the unique data copy.
- ◆ For example, a typical email system might contain 100 instances of the same one MB file attachment. If the email platform is backed up or archived, all 100 instances are saved, requiring 100 MB storage space.
- ◆ With data deduplication, only one instance of the attachment is actually stored; each subsequent instance is just referenced back to the one saved copy. A 100 MB storage demand could be reduced to only one MB.
- ◆ Data deduplication offers other benefits:
 - Lower storage space requirements will save money on disk expenditures.
 - The more efficient use of disk space also allows for longer disk retention periods, which provides better recovery time objectives (RTO) for a longer time and reduces the need for tape backups.
 - Data deduplication also reduces the data that must be sent across a WAN for remote backups, replication, and disaster recovery.

<https://www.youtube.com/watch?v=flJrKt-AQYtE>



Thin Provisioning

- ◆ **What it is:** A storage virtualization technique that provides enough disk capacity for applications to run efficiently, pooling the rest among applications.
- ◆ **Why it's important:** Thin provisioning saves on power and disk usage.
- ◆ **The catch:** Thin provisioning works fine in environments that are predictable, but where there are unexpected data surges, provisioning systems may falter.



<https://www.youtube.com/watch?v=bpZKieK0uTQ>

Data Compression

- ◆ Storing data in a format that requires **less space** than usual.
- ◆ Compression is useful because it helps reduce resources usage, such as data storage space or transmission capacity. Because compressed data must be decompressed to use, this extra processing imposes computational or other costs through decompression.
- ◆ Data compression is particularly useful in communications because it enables devices to transmit or store the same amount of data in fewer bits. There are a variety of data compression techniques and formats, such as **ARC** and **ZIP**.
- ◆ Data compression is also widely used in backup utilities, spreadsheet applications, and database management systems.
- ◆ Compression is also used in the multimedia area, such as audio/video. <http://www.youtube.com/watch?v=0E5kCRsr4gQ>

<https://www.youtube.com/watch?v=dImkc8tacd0>

Disaster Recovery

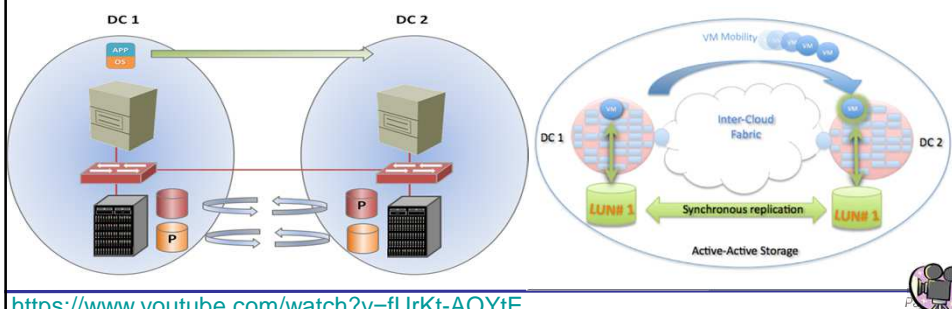
- ◆ **What it is:** Traditionally, organizations deployed disaster recovery programs to only a subset of technologies within the organization. Now, disaster recovery products use continuous data protection to deliver affordable DR for everyone.
- ◆ **Why it's important:** New DR products are less expensive than traditional offerings.
- ◆ **The catch:** Still requires a sizable investment.



Page 37

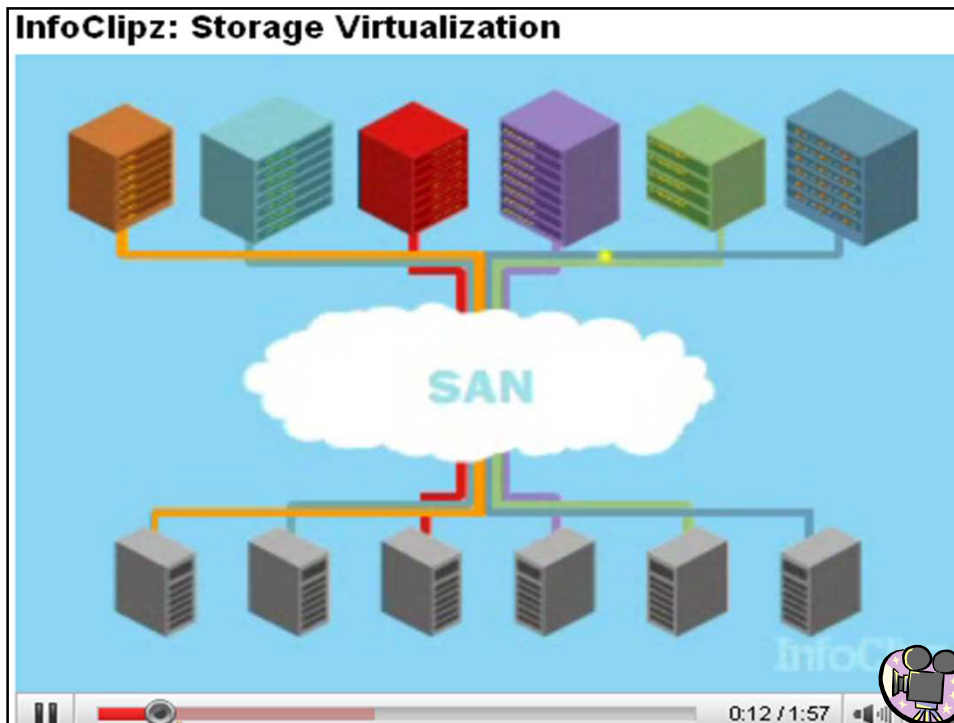
Replication

- ◆ Replication in computing involves sharing **information** so as to ensure consistency between redundant resources, such as software or hardware components, to improve reliability, fault-tolerance, or accessibility.
- ◆ High availability, disaster recovery, business continuity, etc. are all key concerns of any data center design. They all describe separate components of the big concept: **"When something does go wrong how do I keep doing business."**



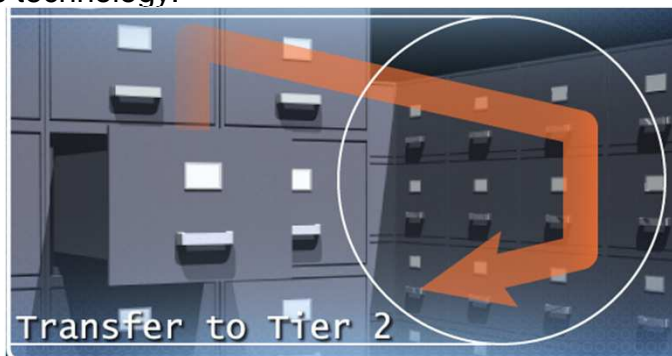
<https://www.youtube.com/watch?v=flUrKt-AQYtE>





Hierarchical Storage Management in the Array

- ◆ **What it is:** Storage arrays that migrate data from Tier 1 storage to Tier 2 and Tier 3 storage based on a variety of criteria, including last access.
- ◆ **Why it's important:** HSM has been around since mainframe days as a means of optimizing storage investment. Now, vendors are delivering HSM in single arrays, making it simpler to implement the technology.



Page 40

Archiving

- ◆ Data archiving is the process of moving data that is no longer actively used to a separate data storage device for long-term retention.
- ◆ Data archives consist of older data that is still important and necessary for future reference, as well as data that must be retained for regulatory compliance.
- ◆ Data archives are indexed and have search capabilities so that files and parts of files can be easily located and retrieved.
- ◆ Data archives are often confused with data backups, which are copies of data.
 - Data backups are used to restore data in case it is corrupted or destroyed.
 - Data archives protect older information that is not needed for everyday operations but may occasionally need to be accessed.
- ◆ Should be implemented with **Records Management** and **ILM**

Page 41

For DBA ...

- ◆ One of the easiest ways to improve the lifetime performance of a database is proper setup of the physical and logical drives.
- ◆ What RAID level is being used for each database file?
- ◆ What workload is being generated?
 - Read Only vs. Update
 - OLTP vs. OLAP
- ◆ What bottlenecks does my storage administrator see?
- ◆ What can we (DBA & Storage Admin.) do to improve the configuration?



Storage Solutions for DBA Challenges

Backup issues

- ◆ **Scalability:** it can take a long time to physically copy 200 TB of data to tape or disk
- ◆ **Cost:** it costs a lot to purchase 200 TB of storage so you can perform physical image backups of 200 TB databases. It is costly to even purchase hardware to test backups.
- ◆ **Performance:** keeping large databases in hot backup mode negatively affects the performance of high-transaction systems (inserts, updates, deletes)
- ◆ **Complexity of systems:** multiple databases, interlinked systems, different database versions, RMAN/non-RMAN, RAC, ASM, etc.
- ◆ **Manageability:** it is often difficult to setup, manage, and test backups

Recovery issues

- ◆ **Manageability:** human errors, lost data, inconsistent data, physical failures, corruption can require restores. It can be a manual and scary process to recover to a consistent point-in-time
- ◆ **Performance:** how in the world do you recover a 200 TB database in 15 minutes?

Page 43

Storage Solutions for DBA Challenges

Data growth

- ◆ **Cost:** DBAs tend to put storage on single class of storage without archiving or tier considerations
- ◆ **Performance:** system response time is 5 seconds now. What happens when my data triples in size?
- ◆ **Manageability:** it is often a political process to get additional storage from non-DBA groups

Development and testing

- ◆ **Cost:** it costs a lot to purchase 200 TB of storage to get an image copy of production to test and development!
- ◆ **Scalability:** it is too much to provide 5, 10 or more copies to development and testing teams
- ◆ **Manageability:** it can take 25% or more of a DBA's time to manage the cloning process

Page 44