The Evolving Roles of Tape and Disk in the High Performance-Data Protection Environment

Cindy Grossman
Vice President
IBM Tape and Archive Systems

Agenda

- Worlds Data Environment – Key Challenges
- Smarter Information Infrastructure
- IBM's Data Protection and Archive Business Strategy
- Key Product Initiatives
- Summary

Data Explodes - Our Planet is Getting More Complex

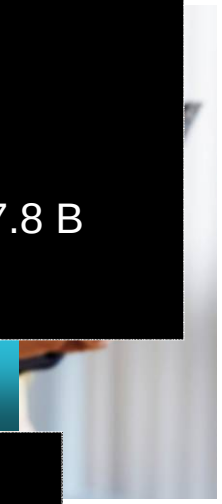
Information is being authored by billions of people and flowing from a trillion intelligent devices, sensors, and all manner of instrumented objects

Massive Data Growth

63 million US households to pay bills online by 2014

635 million broadband internet users in 40 countries by 2013

Online travel purchases in the US alone expected to jump to $157.8 B by 2013 about 37% increase over 2009



Daunting Compliance and Data Security Requirements

203 million Americans text

2.5 B texts sent each day in the USA

Worldwide mobile phone subscriptions reached 4B in 2008

Significant TCO and Energy Consumption Challenges

Managing Data Growth and Protecting Data…
….Key Challenges

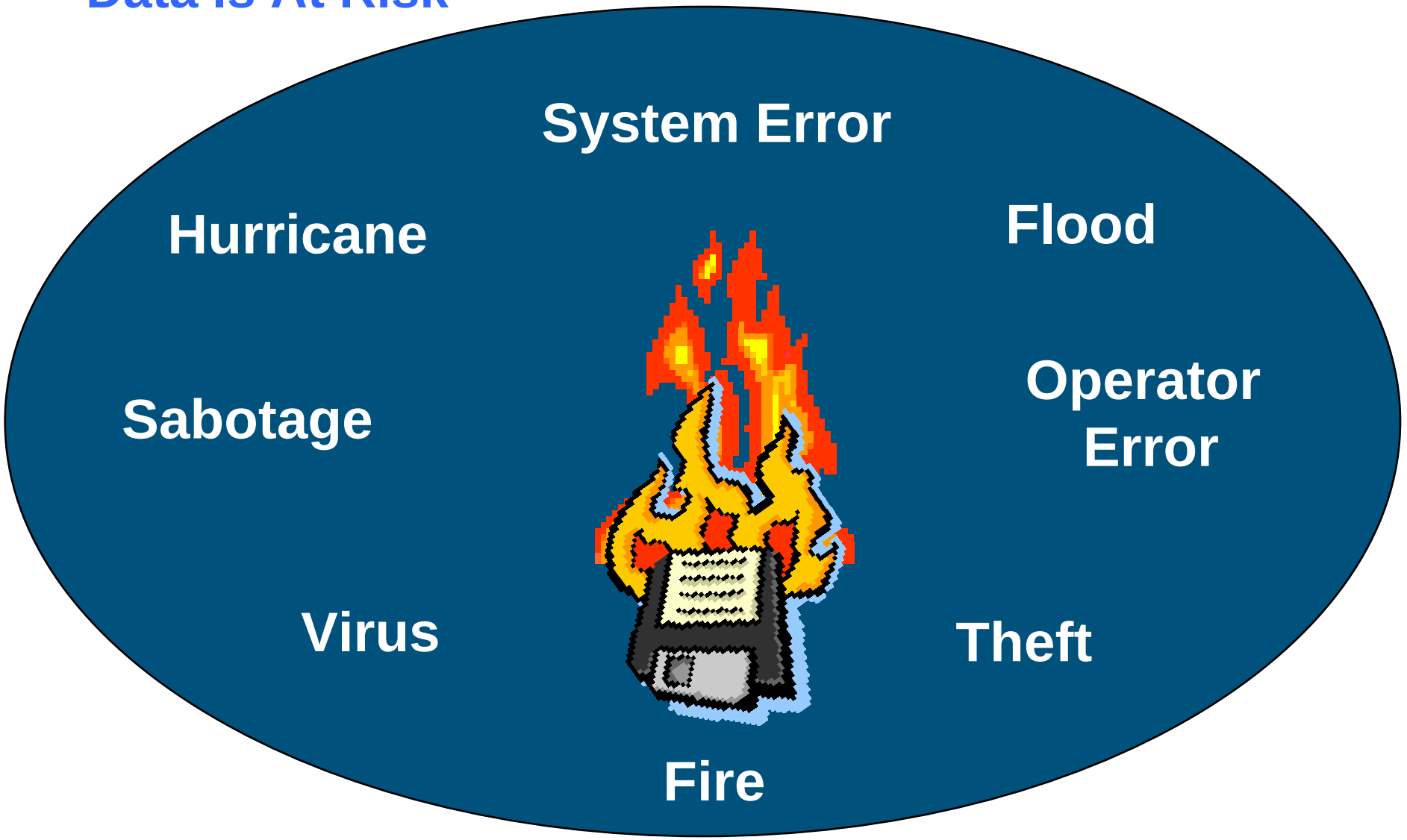
- In Fortune 500 companies
  - Average amount of electronic data on all platforms and media, other than tape records, grew from 86 TBs in 2003 to 490 TBs in 2008[1]
  - On current trend, it will reach close to 4.2 PBs by the end of 2013

- Managing this growth has become the primary source of pain for storage professionals[2]
  - Growing storage acquisition and management costs
  - Backup administration
  - Shrinking backup windows
  - Demanding service levels

[1]Value Proposition for IBM System z, International Technology Group, Oct 2009

[2] TheInfoPro™ Storage Wave 13 Study, Jan 2010

Data Is At Risk



Hard Lessons Learned

Blogging Service JournalSpace Company Data is Wiped Out

- Completely wiped out after the drives that housed their entire database were overwritten.
- Problem: Backups weren't actually backups
- Servers were set up with a mirrored RAID system
- If the primary drive should fail the secondary drive would be used to recover the primary.
- This alone is risky business: it only protects you from the failure of one drive.
- In this case, the drive didn't fail but instead due to system error …
  …the data was erased on one drive leading the other drive to follow suite and clear itself.
- A data recovery team was unable to retrieve the database.
- If you are replicating/mirroring and the data is corrupt then your replicated sites most likely have corrupt data

Source: http://lifehacker.com/5122848/hard-lessons-in-the-importance-of-backups-journalspace-wiped-out

Hackers Destroy Flight Simulation Site

Flight simulator site AVSIM has been "destroyed" by malicious hackers

- Launched in 1996, covered all aspects of flight simulation
- The attack took down the site's two servers
- Servers were backed up to each other every day, but..
- …An external backup system had not been established
- "*The hacker took out both servers, destroying our ability to use one or the other back-up to remedy the situation*." said Tom Allensworth, founder

Source: http://news.bbc.co.uk/1/hi/technology/8049780.stm

Don’t Gamble With Your Data

Backup and Archive Best Practices – Protect Data

- **Have multiple levels of protection:** at least 3 copies of data in different locations and one in a remote region for DR* - Use tape
- **Have technology diversification:** copies on different forms of media to avoid a media or system process disaster* - Use disk and tape
- **Have system isolation:** at least one copy offline to avoid intentional or unintentional corruption* - Use tape
- **Protect access to data** at rest and in transit – Use tape encryption

*Source: Debbie Beech, Sylvatica Consulting / David Hill, Mesabi Group

Large Truck Express Line Survives Hurricane

- Problem
  - **Hurricane Gastone flooded Data Center with 5 ft. of water**
  - **Total loss of hardware, networks, phone systems, generator, utility power**
  - **Good news: *We had a tape backup of 100% of our data made night before - off site!***
- Objective
  - **Protect assets and business resilience with comprehensive strategy**
  - **Control TCO with tiered storage strategy**

"You are out of your mind if you think you can live without tape."
Dick Cosby, Systems Admin.

- Solution
  - **Create flash copy for fast retrieval and window-less backup to tape**
  - **Backup 100% production data to LTO-4 tape and library**
  - **Global Mirror DR site with LTO tape library – Lights out!**
- Benefits
  - **No production System interruption**
  - **No save window-Set it & forget it**
  - **No production cycles, no operators, lights out operations**
  - **Multiple tiers of protection**
  - **Out of region protection**

Smarter Information Infrastructure…
…Leading the Transition to a Smarter Planet

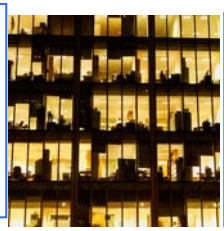
Today, IBM is helping our clients to literally change the way the world works – and to make the planet seem not just smaller and flatter but smarter.

Being smarter includes protecting data stores and achieving smarter archives which can help to attain efficient operations, simplified management, information availability and lower costs.

Mobility Infrastructure

+

Facilities Infrastructure



+

Production Infrastructure

+

Technology Infrastructure

+

Communications Infrastructure

Users are Managing Multiple Workloads

Business Applications

- High availability
- Selective disaster recovery
- Selective overwrite protection
- Block and file storage
- Ease of management
- Tier 1 and 2 data

Web, Collaboration and Infrastructure

- Cost Sensitive, high capacity
- Rapid scalability
- Non-disruptive expansion
- Ease of management
- Mostly file system storage
- Mostly Tier 2 data

Transaction Processing & Database

- High availability
- Disaster recovery with low RPO/RTO
- High throughput block storage
- Increased data protection
- Tier 1 data

Analytics & High Performance

- High throughput - contribute to overall system performance
- Very large file systems
- Mostly extractions of Tier 1 data

Tier 2 and Tier 3 Data Protection and Management

- Backup
- Archive
- Disaster Recovery
- Compliance Repository
- Test and Development
- Migration

Smarter Storage Optimization Occurs… …when Disk and Tape Work Together

- **Disk for backup target for high recall applications**
  - Fast backup & retrieval for high performance SLAs
- **Tape for low recall backups, data retention, data security**
  - Best practices
    - Offline for system isolation-protection
    - Portable for disaster protection
  - Lowest TCO and energy conservation
  - High reliability
  - Encryption for data security
  - WORM capability for compliance
- **What IT Managers Want**
  - Blended tiered storage
  - Flexibility and rich portfolio of options
  - Offerings that are proven and work together

5 yr. TCO study of SATA Disk vs. LTO Tape Library System

- Disk 23X more costly, VTL Dedupe 5X more
- Disk 290X more energy consumed

Source: Clipper Group 2008

Directions in Data Protection

- **Primacy of disk in the data protection domain**
  - Disk in data protection to continue with storage of increased quantities of data
  - Tape as backup for low frequency apps, archive, off-line data protection
- **Deduplication in the Enterprise**
  - In 2009, Deduplication is still relatively early in the adoption curve
  - Maturation in 1-2 years; vendors with the best technologies will remain in the market
- **Further extension and innovation of VTLs**
  - VTLs evolved from simple disk-based target, to advanced platforms that will continue to deliver innovative features and functionality
- **Emergence of electronic data transport within the Data Protection domain, and the blending of DP with DR**
  - Customers deploying remote replication of some of their data protection domain
  - This is enabled by deduplication with lower bandwidth replication
- **Continued innovation and integration with the backup applications**
  - Backup apps to remain the primary management paradigm for the data protection realm over the next five years
  - Expect enhancements to backup API programs allowing increasingly tighter management between the backup app, VTLs and physical tape
  - Tape to capture new content and content not currently in digital form for: archive, compliance, encryption, seismic, video/digital storage
  - Tape technology will widen it's $/GB advantage over disk in years to come

Data Protection and Archive Strategy

That's Smart!

- ***Provide workload optimized hierarchical solutions for Data Protection & Archive that place information on the lowest cost device that meets service level objectives.***

Information Compliance

Integrate tape with compliance solutions &
Expand WORM capabilities to include information in disk cache

To reduce reputation risks and audit deficiencies

Information Availability

Continue to enhance tape systems availability characteristics &
Expand VT remote replication capabilities for multi-site operations

To deliver continuous, reliable access to information

Information Retention

Extend physical tape to new use cases
Expand and enhance virtual tape retention and residency options

To support information retention policies

Information Security

Expand security controls and integrate with LDAP solutions &
Integrate with IBM's end to end security solution portfolio

To protect and enable secure sharing of information

IBM Data Protection and Archive Initiatives

Improve Total Cost of Ownership

IBM Tape Density Achievement – 35TB Tape

•Jan. 2010: IBM Research - Zurich recorded data onto an advanced prototype tape developed by Fujifilm Corporation of Japan

•Density of 29.5 billion bits per square inch -- about 39 times the data density of today's popular standard tape

•Could produce future tape cartridge holding up to 35TB

•New critical IBM technologies were developed:
- •Dramatically improved precision of controlling the position of the read-write heads
- •More than 25-fold increase in the number of tracks
- •New detection methods to improve the accuracy of reading magnetic bits
- •New low-friction read-write head

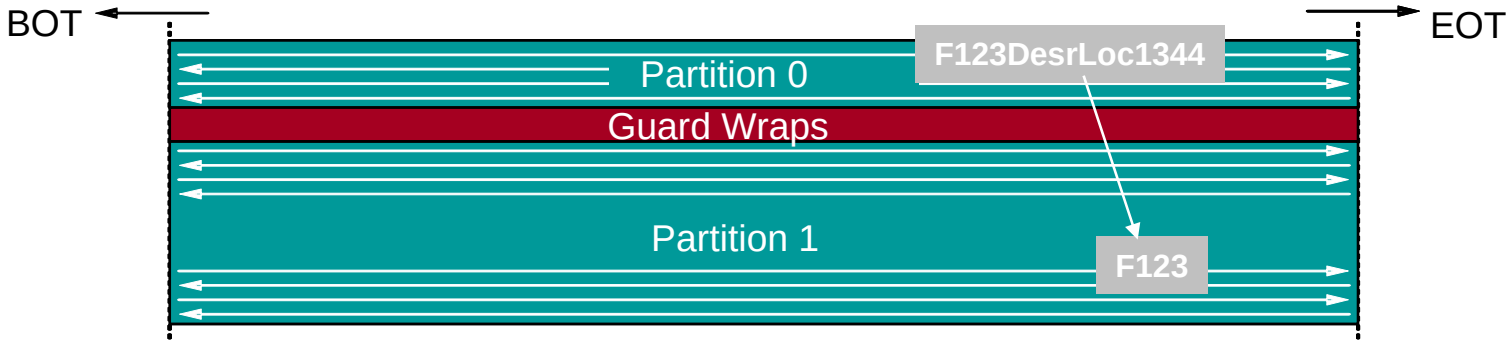
•Represents a step towards developing technologies to achieve tape areal recording densities of 100 billion bits per square inch and beyond

Tape's Future is Strong

Enhance Tape File Access and Management

Vision - IBM Long Term File System - The Smarter Tape

- Efficient: can utilize new dual-partition planned for LTO Gen-5 (next TS1100)
- Index partition and data partition: Mount a tape as if it was a hard drive
- Self describing tape enabling hierarchical directory structure, file names, file properties, metadata files, fast search indexes, domain-specific information



Potential Applications:

- *Video archive and individual scene access*
- *Design and drawing content for manufacturing, architecture, and more*
- *Medical / Health industry images*
- *e-discovery*

Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Media & Entertainment Industry

- Key Requirements
  - Very large archives and long term archive requirements: 50-100 years
  - Film/video tape should be self-contained open tape format
    - *Files stored in application-independent fashion*
    - *Ability to distribute and interchange files on tape*
    - *Files readable from information stored to tape without requiring application's DB*
    - *Ability to recover from corruption or loss of application's DB*
  - Lower cost than current methods: *LTO tape can address*
- LTO-5 Partitioning and IBM Long Term File System Potential Business Value:
  - Can reduce cost of video storage media and equipment
  - Supports file-based workflow, increases data mobility and can unify across organizations to a single storage media for video formats
  - Can reduce data ingest time
  - Can reduce archive storage space by 10x (compared to video-tape)
  - Supports green initiatives: can reduce energy consumption by up to 100x (compared to HDD)

Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Increase Tape Storage Utilization – Protect More

IBM Information Infrastructure for Data Archiving

IBM System Storage TS3500 Tape Library High Density Storage Only Frames

- **Leading high density slot technology**
  - Offers up to 3 times the capacity of existing frames in the same footprint
  - Allows non-disruptive addition of capacity on demand
  - TS3500 can store over 15K 3592 carts or over 20K LTO carts or a combination
- **Client Value**
  - Lower price per slot and lower TCO
  - Add significant capacity in a small space
  - Energy efficient option for long term archive
  - Add capacity as required and pay as you grow
  - Reduce downtime during capacity upgrades
  - Archive data at rest while minimizing power and cooling costs

Store over 3PB of data in 10 square feet

Information Retention

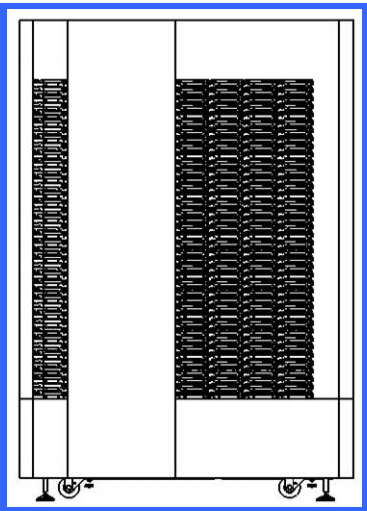
Lowest cost storage for long term data retention

ibm.com/systems/storage/tape/ts3500

TS3500 Model S24 and S54 High Density Frames

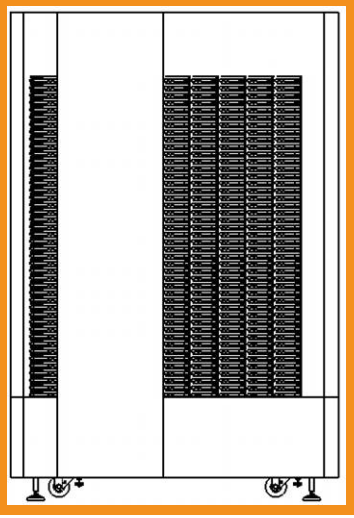
- **S24 - 3592 HD Frame**
  - Max. 4 cartridge tiers in HD slot
  - 1000 slots 3592 media (compare to D23: 400 slots 3592 media)
  - Storage capacity increase: from 6,200 slots to over 15,200

- **New Frame S24 provides:**
  - The capacity of 2.5 Model D23s
  - for the price of 2 Model D23s
  - in the footprint of 1 Model D23.



- **S54 - LTO HD Frame**
  - Max. 5 cartridge tiers in HD slot
  - 1320 slots LTO media (compare to D53: 440 slots LTO media)
  - Storage Capacity increase: from 6,800 slots to over 20,000*

- **New Frame S54 provides:**
  - The capacity of 3 Model D53s
  - for the price of 2.3 Model D53s
  - in the footprint of 1 Model D53.



Improve Access to Data and Business Continuity

IBM Tape Virtualization Offerings

TS7650G – ProtecTIER De-Duplication Gateway for Open Systems

- Enterprise strength de-duplication with native replication
- High Availability Cluster supporting a wide range of disk storage
- High Performance – write data rate up to 1GB/second (clustered)

TS7650 – ProtecTIER De-Duplication Appliance for Open Systems

- 4 Pre-configured appliances for rapid deployment into backup environments
- Scalable performance up to 500MB/sec inline data deduplication
- High Availability Cluster and native replication

TS7680 – ProtecTIER De-Duplication Gateway for the Mainframe

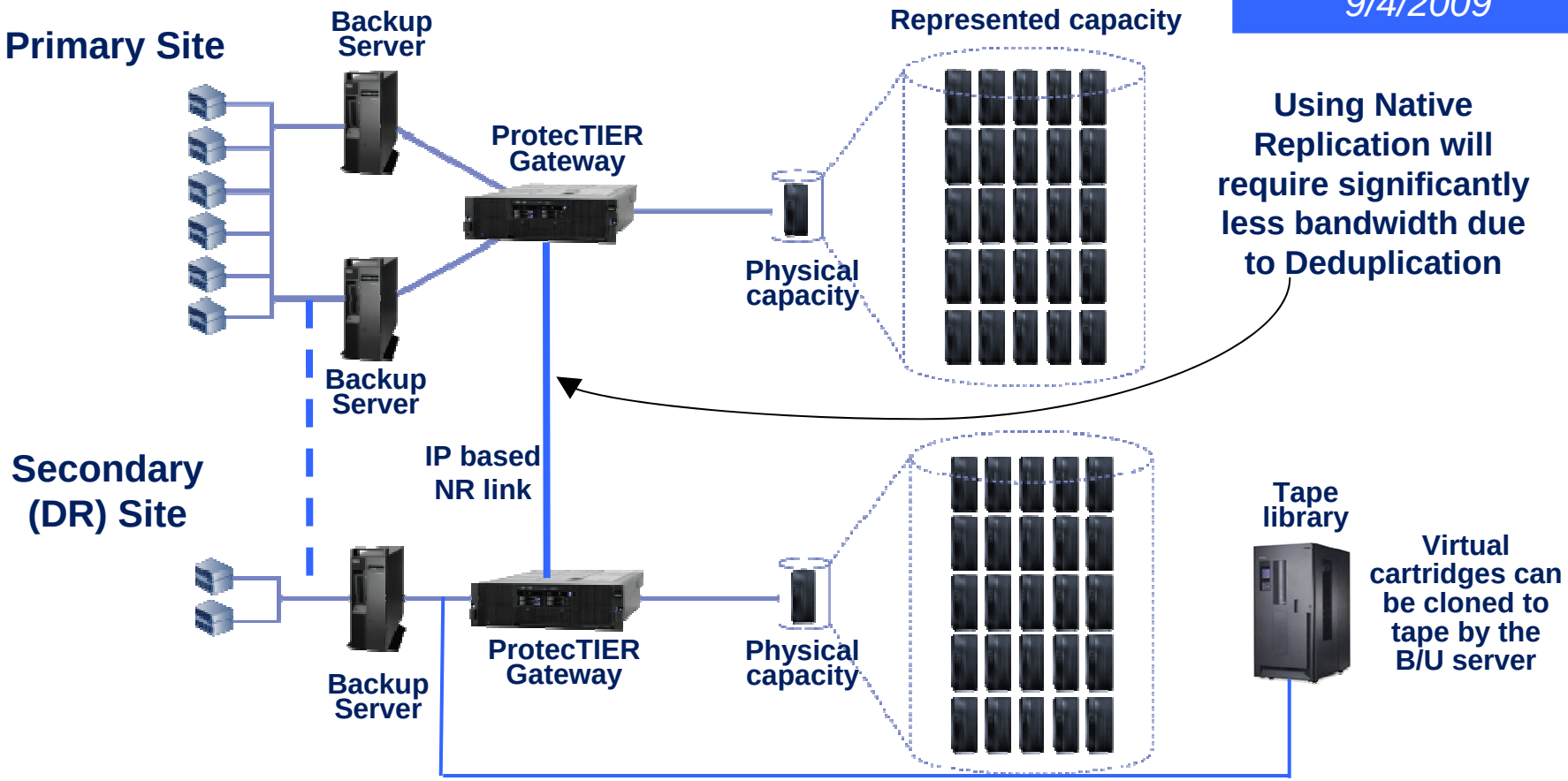
- High performance in-line data deduplication for System z data centers
- Supports a high-availability configuration and Retains DFSMS™ (SMSTAPE) functionality
- Can improve tape application performance while reducing infrastructure costs

TS7700 – Virtualization Engines for the Mainframe

- TS7740 helps manage backups/recalls & resources with a tiered hierarchy of disk and tape
- TS7720 provides a cache-centric solution for frequently accessed data
- Supports business continuity with 7740 & 7720 in 4-way GRID with automated replication

ProtecTIER Native IP-Based Replication

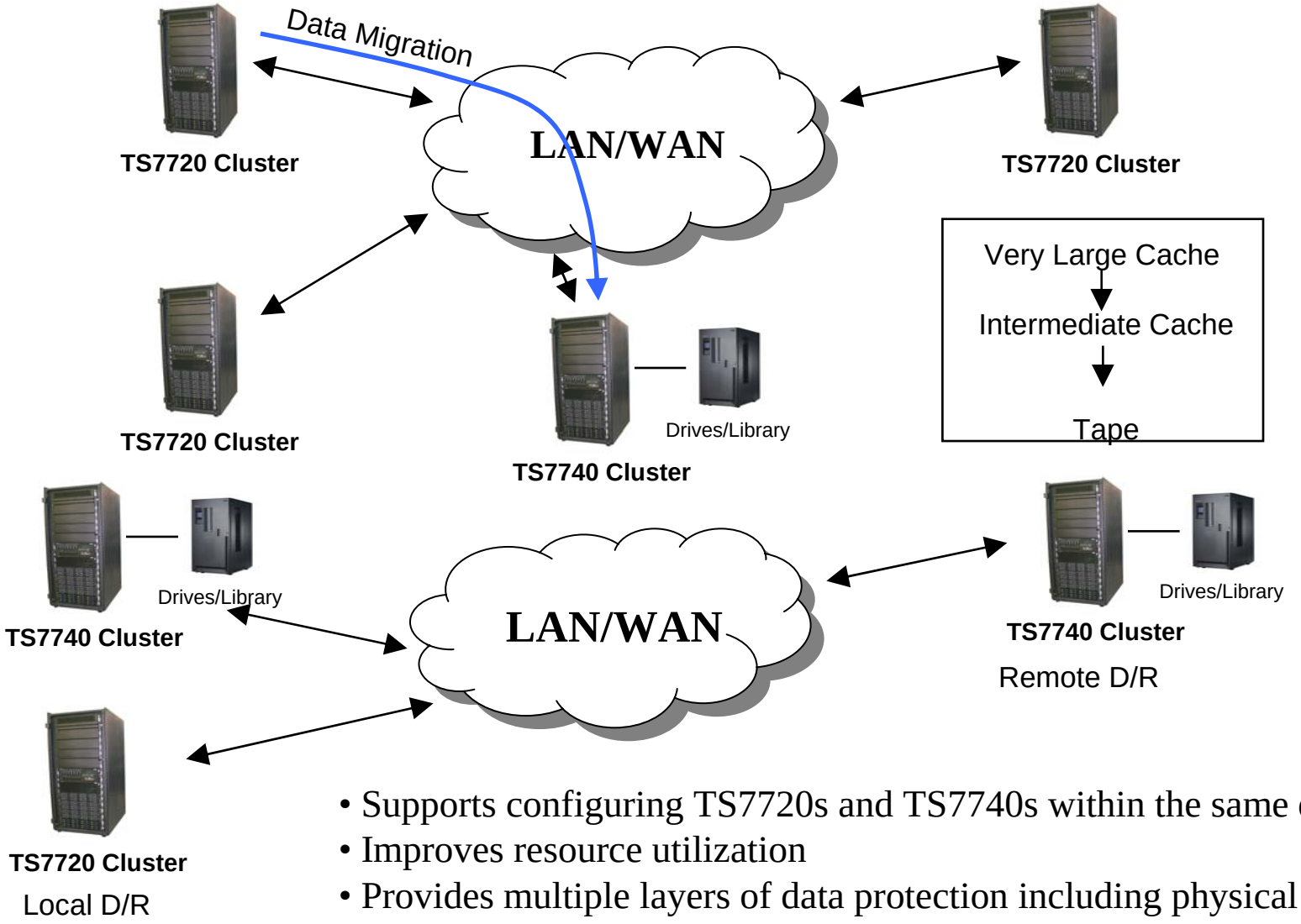
Available
9/4/2009



Statements of IBM's future plans and direction are provided for informational purposes only. Plans and directions are subject to change without notice.

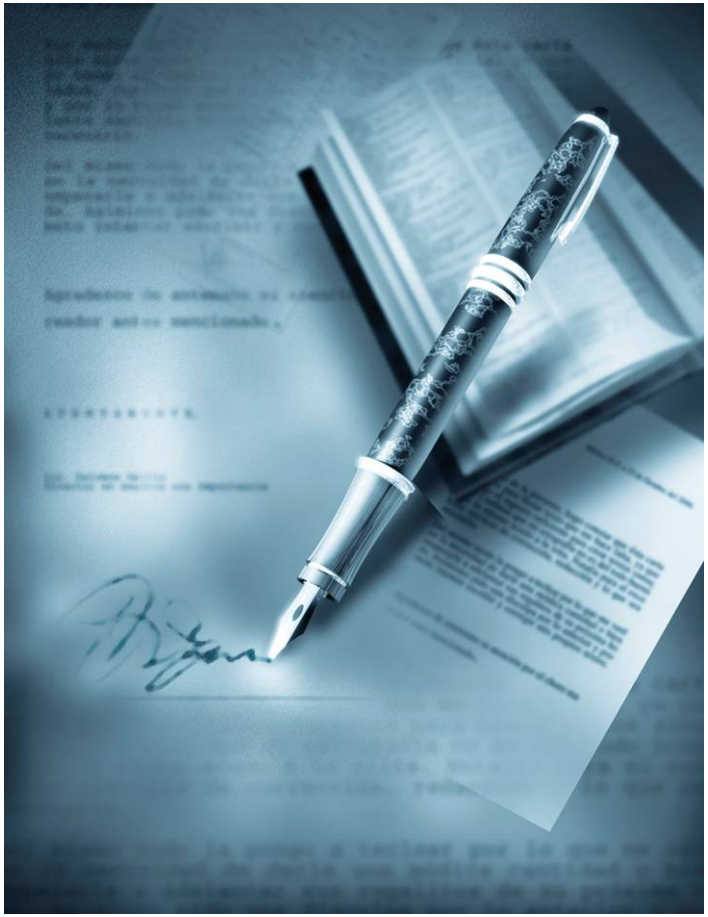
- IP Based Native Replication
- 1 to 1 replication from primary to 2nd site
- Can clone cartridges to tape at the 2nd site
- Replication concurrent with backup
- Cartridge-level replication policy
- Can import/export cartridges via backup application
- Direction: Many to 1 replication

TS7720/TS7740 Hybrid Grid Options



- Supports configuring TS7720s and TS7740s within the same domain
- Improves resource utilization
- Provides multiple layers of data protection including physical tape

Address Compliance and Archive Initiatives



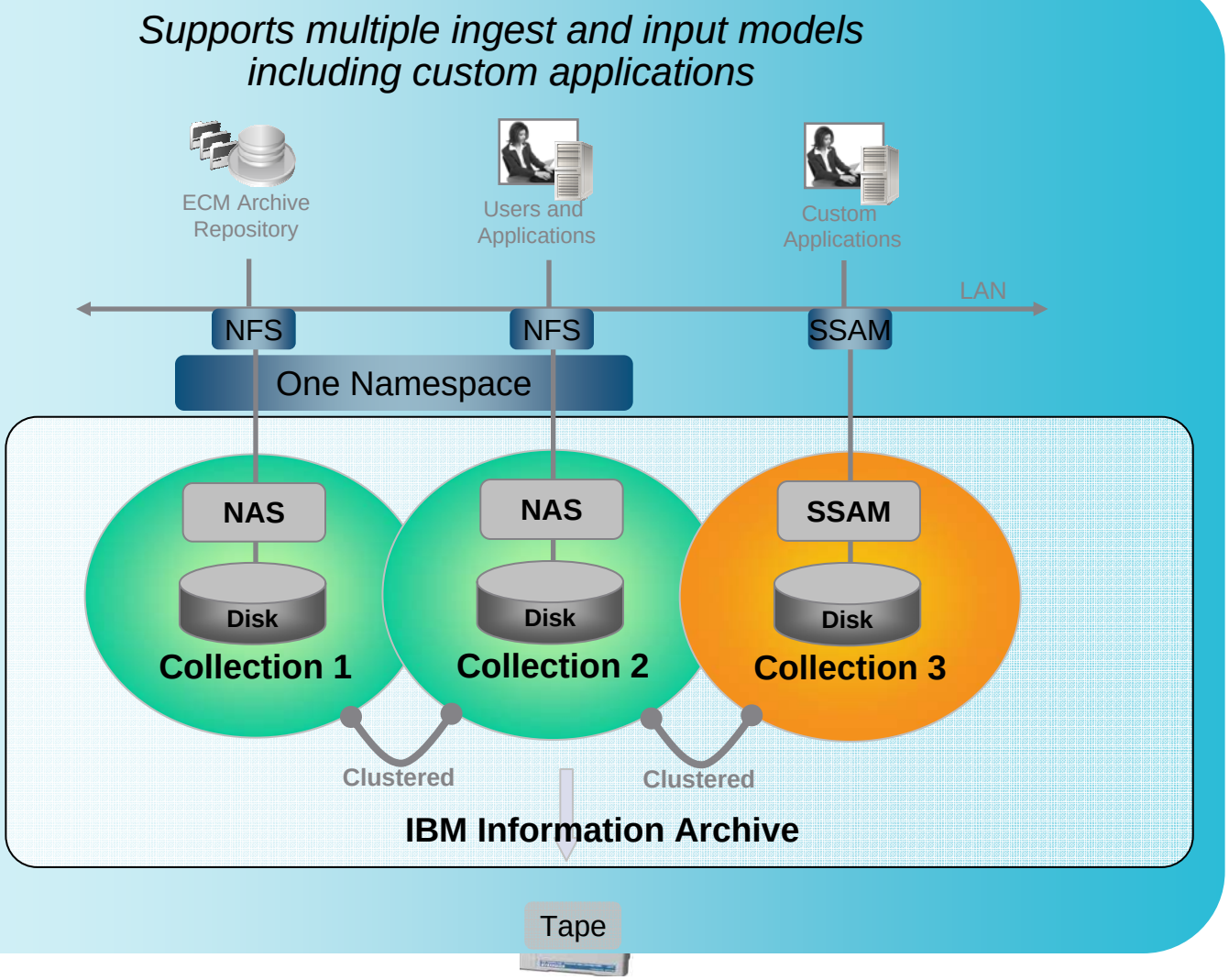
Introducing.. IBM Information Archive

- The Next Generation Information Retention Solution
- A universal storage repository for all types of content
  - Structured (database) and unstructured (files, email, images, docs)
- A robust, scalable, secure information retention solution.
- A fully integrated Hardware, Software and Services solution
- Addresses complete information retention needs of mid size and enterprise clients
  - Business, Legal, or Regulatory

Introducing “Collections” for flexibility

Advanced 3 x 3 Collection Architecture

- A single Information Archive can be partitioned into 3 archive collections (Collections are like virtual IAs)
- Leverage collections by data value or by different usage (ie. one collection for compliant data storage, another for test data storage OR collections could be used by different departments in a customer environment.)
- Each collection can be customized to support different protection levels



Information Archive Collections

- **Collection = Set of archived documents managed under the same policy domain**
- **Types of collections: NAS, SSAM**
- **Multiple collections per IA appliance (up to 3 in R1) with separation of data**

Interface Policy

- Interface protocol (NAS or TSM API)
- Object commit method (NAS only)
  - XML file, timeout, NetApp Snaplock™

Retention Policy

- Controlled internally or by external application
- Time-based (internal) and Event-based (external) retention
- Automatic (internal) or manual (external) deletion after expiration

Storage Policy

- Disk replication
- Tape storage tier
- Encryption (tape only)
- Deduplication
- Compression
- Shredding (SSAM only in R1)

Compliance Policy

Mode	Delete before expire?	Retention Period	
		Shorten?	Lengthen?
Basic	Yes	Yes	Yes
Intermediate	No	Yes	Yes
Maximum*	No	No	Yes

*SSAM Collections are always Maximum compliance

Tape and Archive Systems - Roadmap

- **Short-term**
  - LTO Tape: Gen-5 Tape Drive and Automation Support
  - TS7680: New ProtecTIER Data Deduplication VTL Gateway for mainframe
  - TS7650 Dedupe VTL: Many to one replication, disk refresh, LUN masking
  - TS7700 Mainframe VTL: new disk cache, enhanced cache removal policies, selective write protect for DR testing
  - Information Archive: Clustered NFS v3, HA failover, device replication, 2TB drives
- **Later in 2010**
  - Tape Libraries: SMB High Density Libraries
  - TS7650 Dedupe VTL: SMB appliance, OST support, dedupe off mode
  - TS7700 Mainframe VTL: new server, 5-6 way clusters, 2M logical vols, 10Ge links
  - TS7680 Mainframe Dedupe VTL: logical replication
  - IA: Index/Search, GPFS Currency, Internal storage refresh, Logical Replication for DR
- **Directional**
  - Enterprise Tape: Next Gen TS1100 tape drive; TS3500 Library Shuttle Complex
  - TS7650 Dedupe VTL: File System Interface, Deduplication enhancements
  - TS7680 Mainframe Dedupe VTL: Replication enhancements, DR testing enhancements
  - TS7700 Mainframe VTL: Data Deduplication, Larger logical volumes, 8Gb FICON

Addressing the IT Environment with Solutions for a Smarter Planet

- **Massive Data Growth**

  -IBM Data Growth Solutions

  - Tape Density Demo – 35TB Tape
  - Long Term File System
  - High capacity TS1100 and LTO tape
  - High density library frames
  - TS7700 VTL cache enhancements
  - TS7650G data de-dupe scalability
  - TS7680 ProtecTIER Gateway for System z

- **Daunting Compliance and Data Security Requirements**
  - IBM Data Security Solutions
    - System z security
    - Tape drive encryption
    - Disk encryption
    - IBM Information Archive
    - IBM WORM Tape

- **Significant TCO and Energy Consumption Challenges**
  - Tape is lowest energy consuming storage
  - High density frame enhances energy & floor space usage efficiency
  - Data de-duplication reduces space, power requirements, bandwidth

धन्यवाद
Hindi

谢谢
Simplified Chinese

תודה רבה
Hebrew

Спасибо
Russian

Gracias
Spanish

Thank You
English

شكرا
Arabic

Obrigado
Brazilian Portuguese

ありがとう
Japanese

Danke
German

Grazie
Italian

감사합니다
Korean

Merci
French

நன்றி
Tamil

謝謝
Traditional Chinese

ขอบคุณ
Thai