IBM Mainframe TS7700 Grid as Cloud Like Storage

Ralph Beeston

IBM

08/09/2012

Session Number
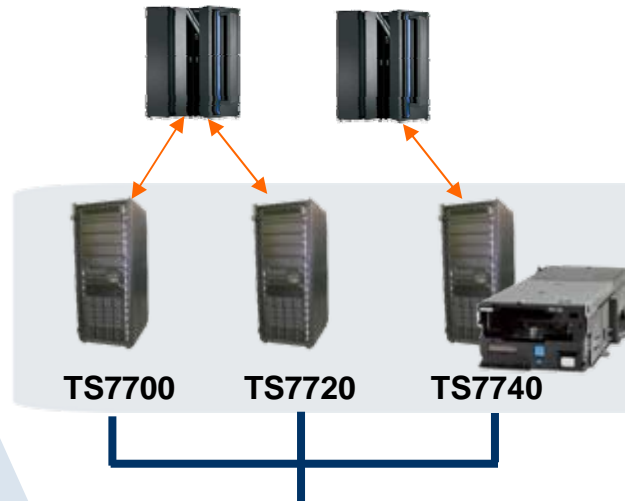
12025

IBM TS7700 System Z Virtual Tape Engine
Leveraging Grid as Cloud Storage for System Z Tape

Grid Cloud

- Disk Speeds
- TS1100 Tape Integration
- Transparent Business Continuance
- Encryption

TS7700   **TS7720**   **TS7740**

Grid Cloud

Replication

Cumulative FICON throughput of over 6GB/s

System z hosts view up to 1536 equivalent devices

Grid access to all data independent of where it exists

TS7700 Virtualization Engine Solutions

- ## TS7720 Virtualization Engine
  - Access-centric applications (image data, report servers, critical backups, HSM ML2)
  - Up to 580TB disk cache
    - 1.7 PB with 3:1 compression

  Hosts

  TS7720 Cluster

- ## TS7740 Virtualization Engine
  - Cost-efficient applications (HSM, general backups, aged data, archive)
  - Up to 28TB disk cache
    - 84TB with 3:1 compression
  - Over 21PB backend physical tape storage
    - 63PB with 3:1 compression

  Hosts
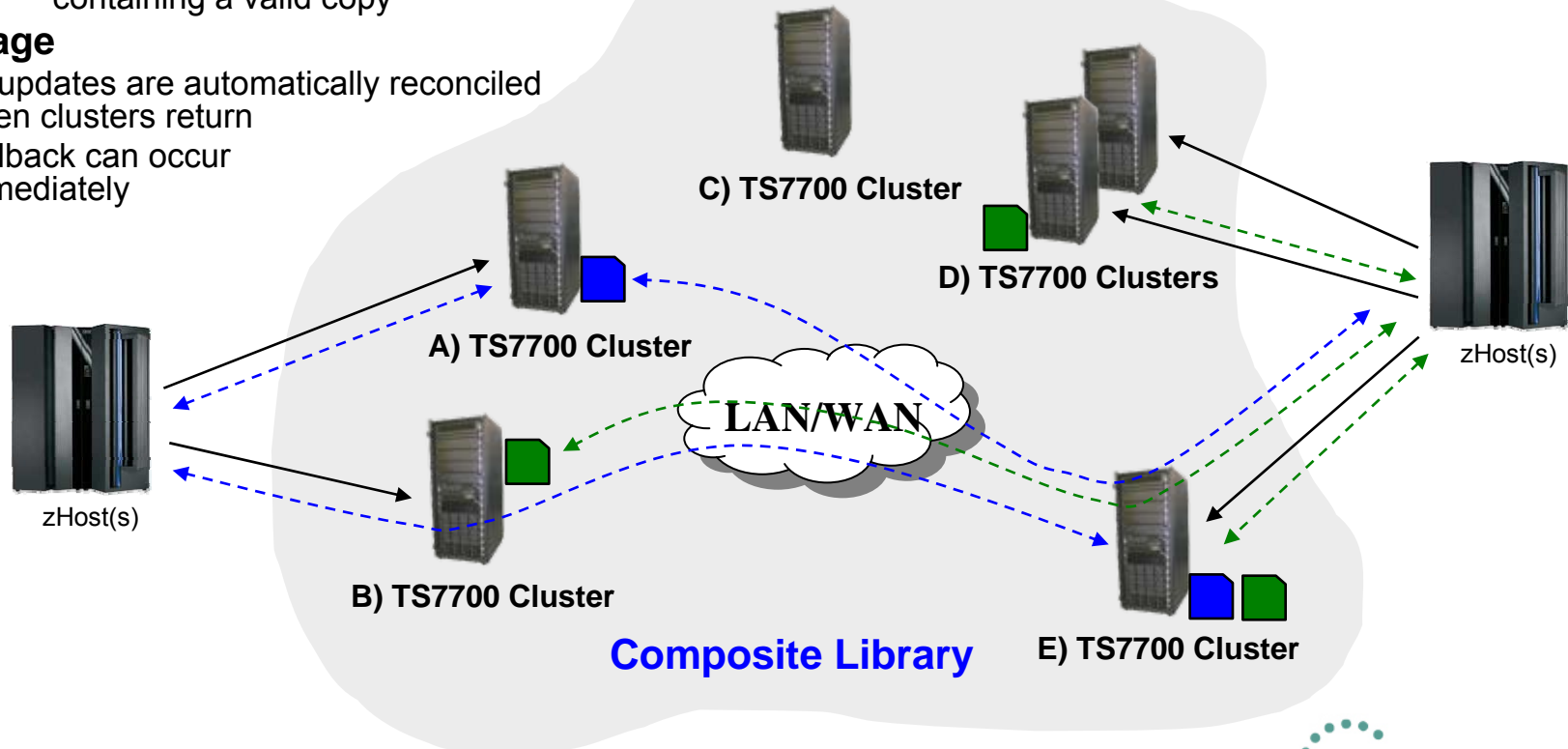
  Drives/Library

  TS7740 Cluster

TS7700 Grid Solutions – **More than just replication**

- **All clusters are equal players – Cloud availability, before it was a popular!**
  - No concept of primary, secondary, or standby nodes
  - Each cluster's devices within an entire grid always have access to all volumes
  - System z hosts views the entire grid as one large composite library with up 1,536 common devices
  - Volume data is accessible from any cluster's devices independent of where copies exist
    - User intervention or host knowledge of where data exists is not required
    - If a local copy isn't present, IP is used as a channel extender to clusters containing a valid copy

- **Post outage**
  - All updates are automatically reconciled when clusters return
  - Failback can occur immediately

C) TS7700 Cluster

D) TS7700 Clusters

A) TS7700 Cluster

zHost(s)

LAN/WAN

zHost(s)

B) TS7700 Cluster

Composite Library

E) TS7700 Cluster

How available is your entire solution?

- **Single Points of Failure don't just apply to internal components**
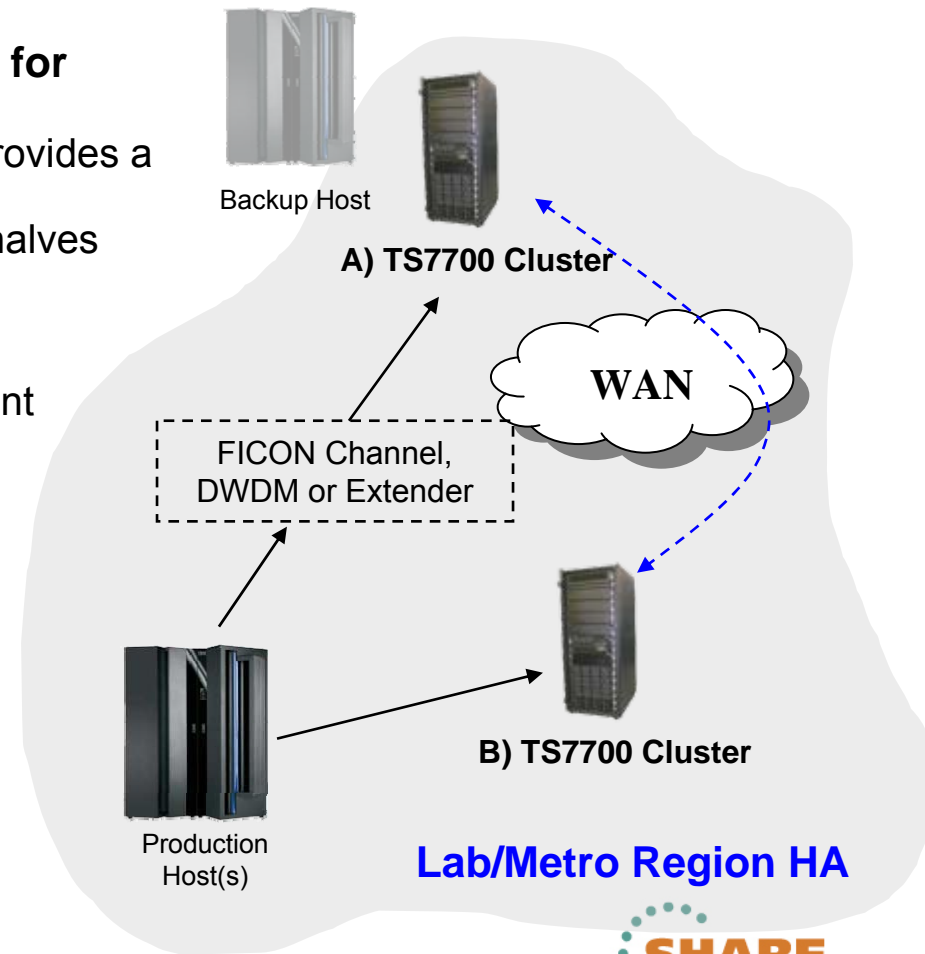 - Power outage
  - Double failures. Even continuously available components have bad days.
  - Regional disaster
- **Do you want to rely on manual DR recovery for your production continuation?**
  - Only a complete redundant configuration provides a true HA solution.
  - The greater the distance between the two halves the more available the solution
    - Power outages
    - Lab failures
  - If the outage is temporary, do you really want to bring up your DR environment to keep running?
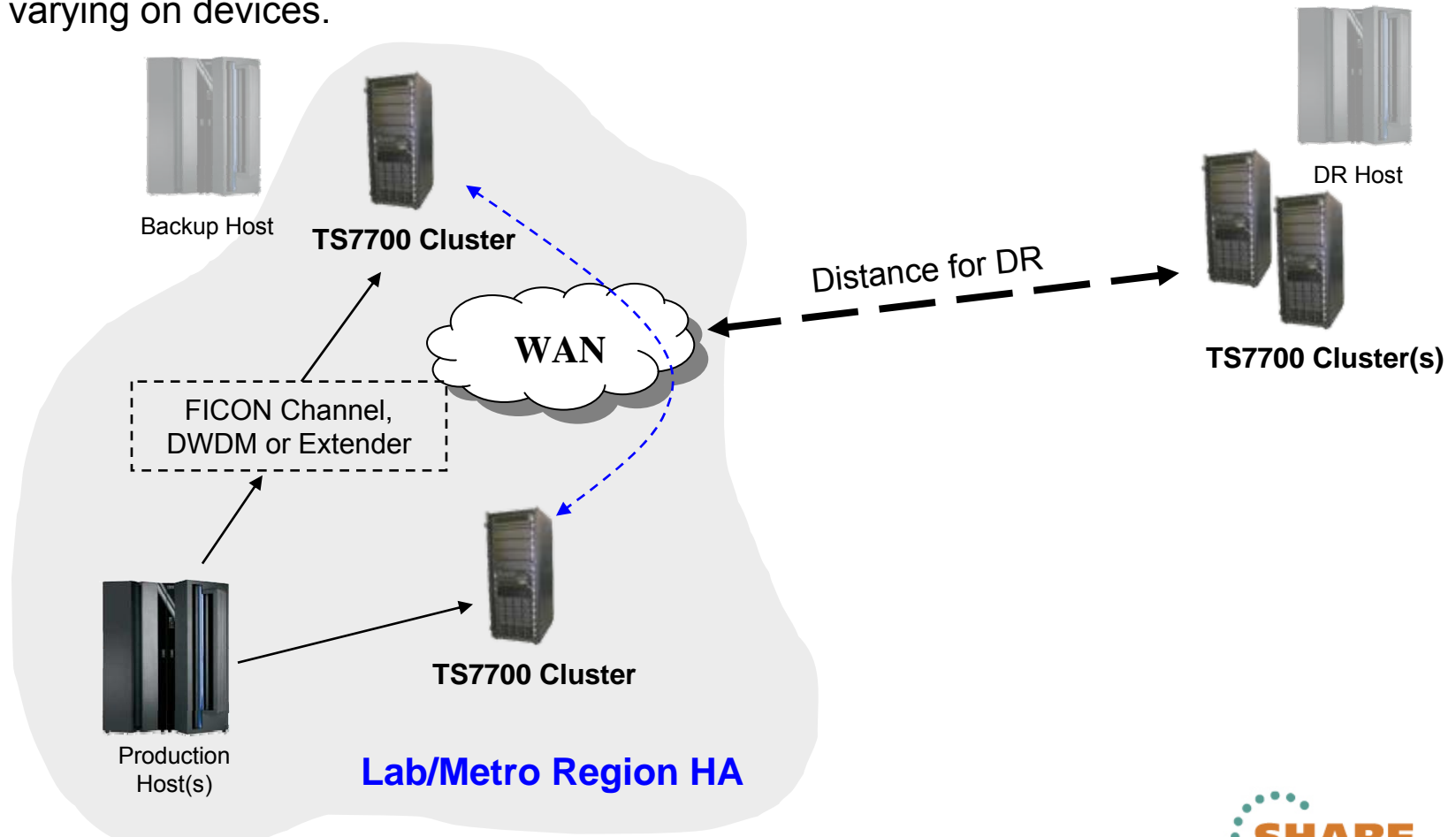- **A true HA solution allows continuous operations without failing over to a DR component**
  - The TS7700 provides true active-active support.
  - Outages can occur, but future operations can still continue transparently and nearly immediately.
    - For both new allocations and READS!

Backup Host

A) TS7700 Cluster

WAN

FICON Channel, DWDM or Extender

B) TS7700 Cluster

Production Host(s)

Lab/Metro Region HA

Is distance for DR also a requirement?

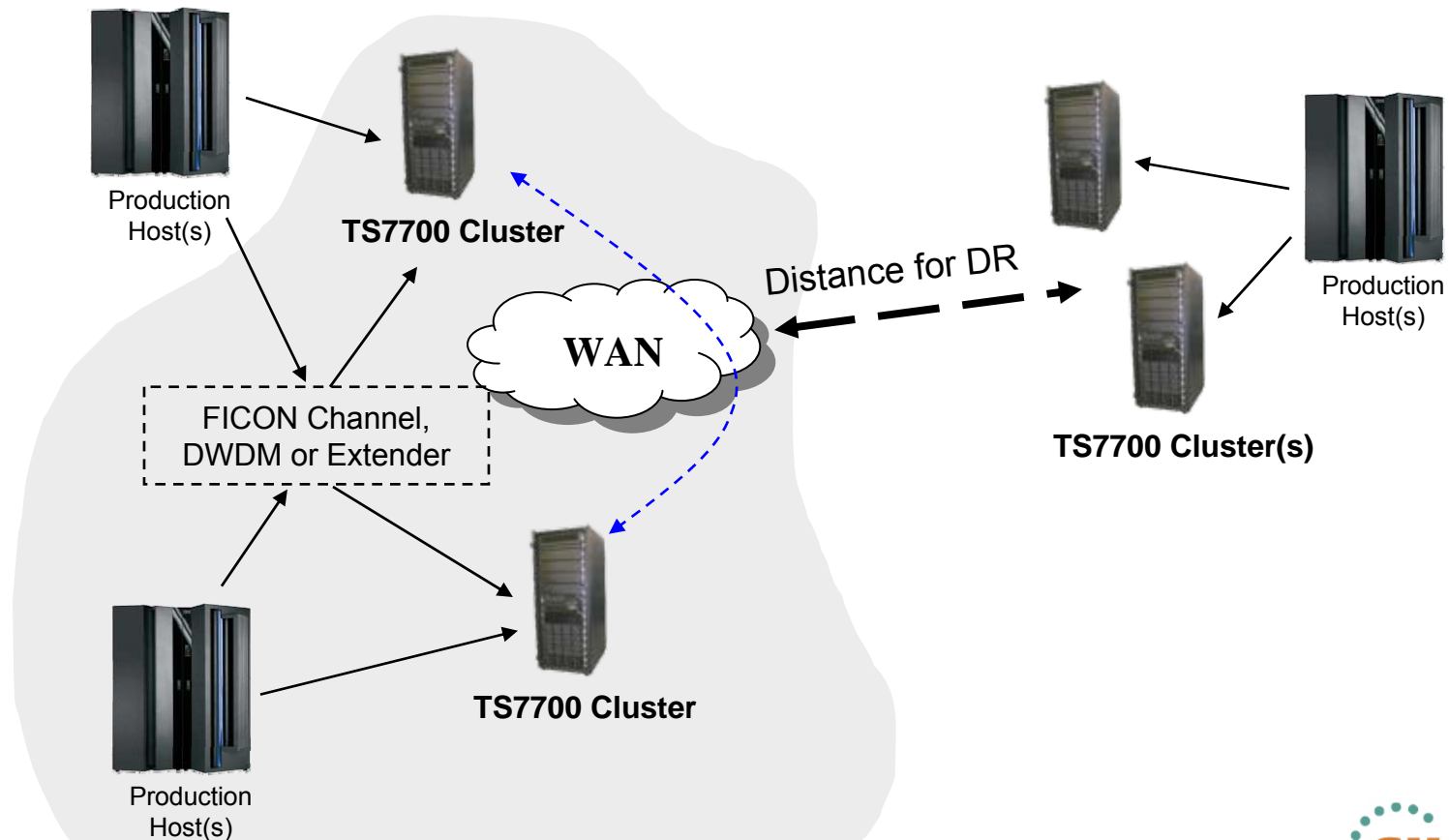- **Introduce 3rd distant site for DR**
    - In the event the entire production environment is lost, recovery can occur at a remote DR location
    - Once the zSeries environment is recovered, continuing operation is as easy as varying on devices.

Backup Host

TS7700 Cluster

DR Host

FICON Channel, DWDM or Extender

WAN

Distance for DR

TS7700 Cluster(s)

TS7700 Cluster

Production Host(s)

Lab/Metro Region HA

SHARE
in Anaheim

2012

Does each site play a role in production?

- **How about using each site for dual purpose? No problem!**
    - Though a site may be labeled as an HA pair or DR site, the technology views each site as an equal player in a grid. By design, any site can concurrently play the roll of all of the above.
    - Run production to all sites at the same time and use adjacent/remote peers for HA and/or DR.
    - The same volume range(s) and scratch pool(s) can even be used at all locations!

Production
Host(s)

TS7700 Cluster

Production
Host(s)

Distance for DR

WAN

FICON Channel,
DWDM or Extender

TS7700 Cluster(s)

TS7700 Cluster

Production
Host(s)

Complete your sessions evaluation online at SHARE.org/AnaheimEval
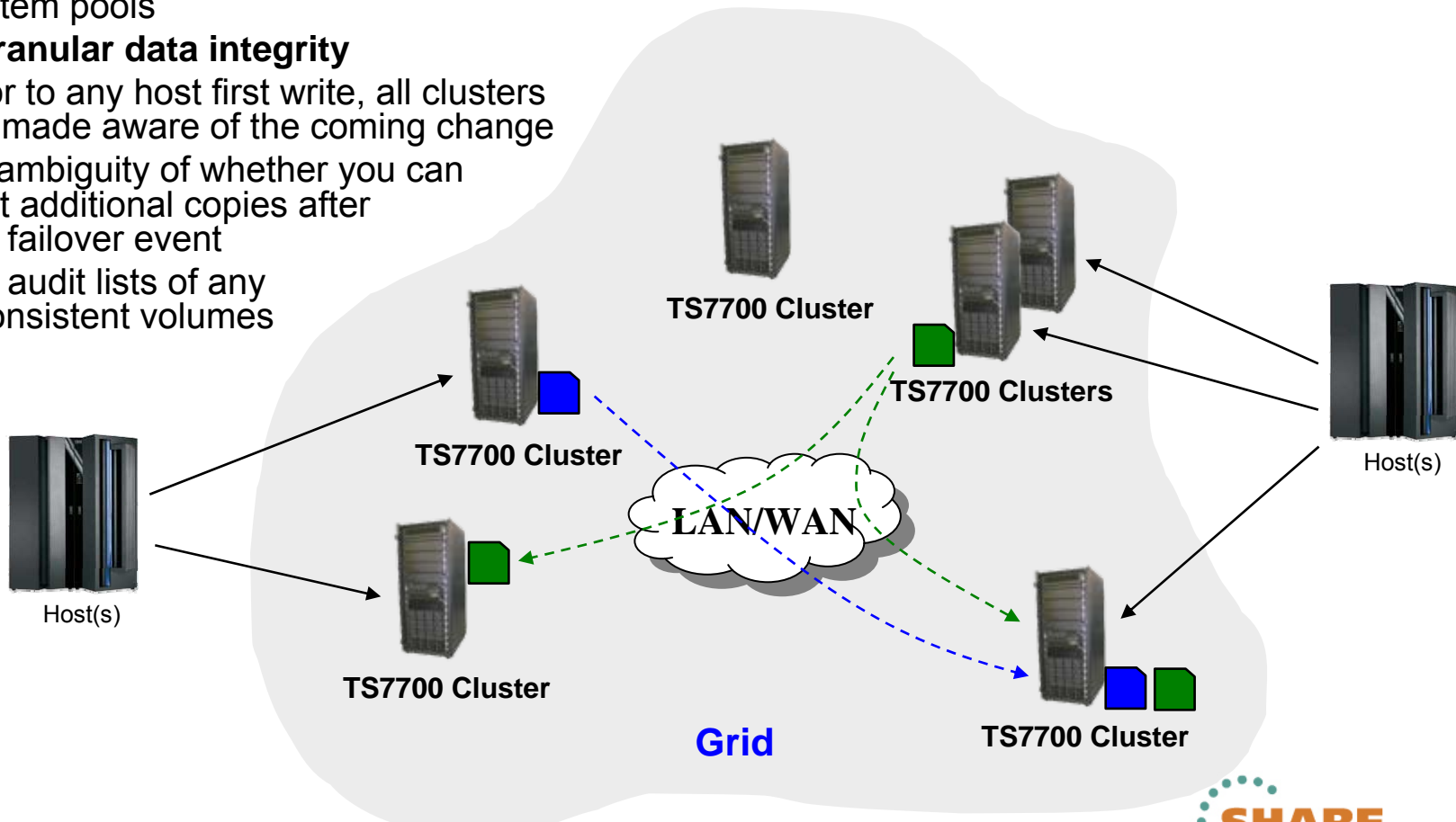
Volume Granular Replication

- **DFSMS policy managed replication**
  - Each volume, independent of where it is created, can have one to six copies within a grid
  - Replication can occur synchronously, immediately or asynchronously or mixtures of the three
  - Simply use your SMS ACS routines to determine which method is used at volume granularity
 - No dependency on Tape Management System pools

- **Volume granular data integrity**
  - Prior to any host first write, all clusters are made aware of the coming change
  - No ambiguity of whether you can trust additional copies after any failover event
  - Full audit lists of any inconsistent volumes

TS7700 Cluster

TS7700 Clusters

Host(s)

TS7700 Cluster

LAN/WAN

Host(s)

TS7700 Cluster

TS7700 Cluster

Grid

SHARE in Anaheim
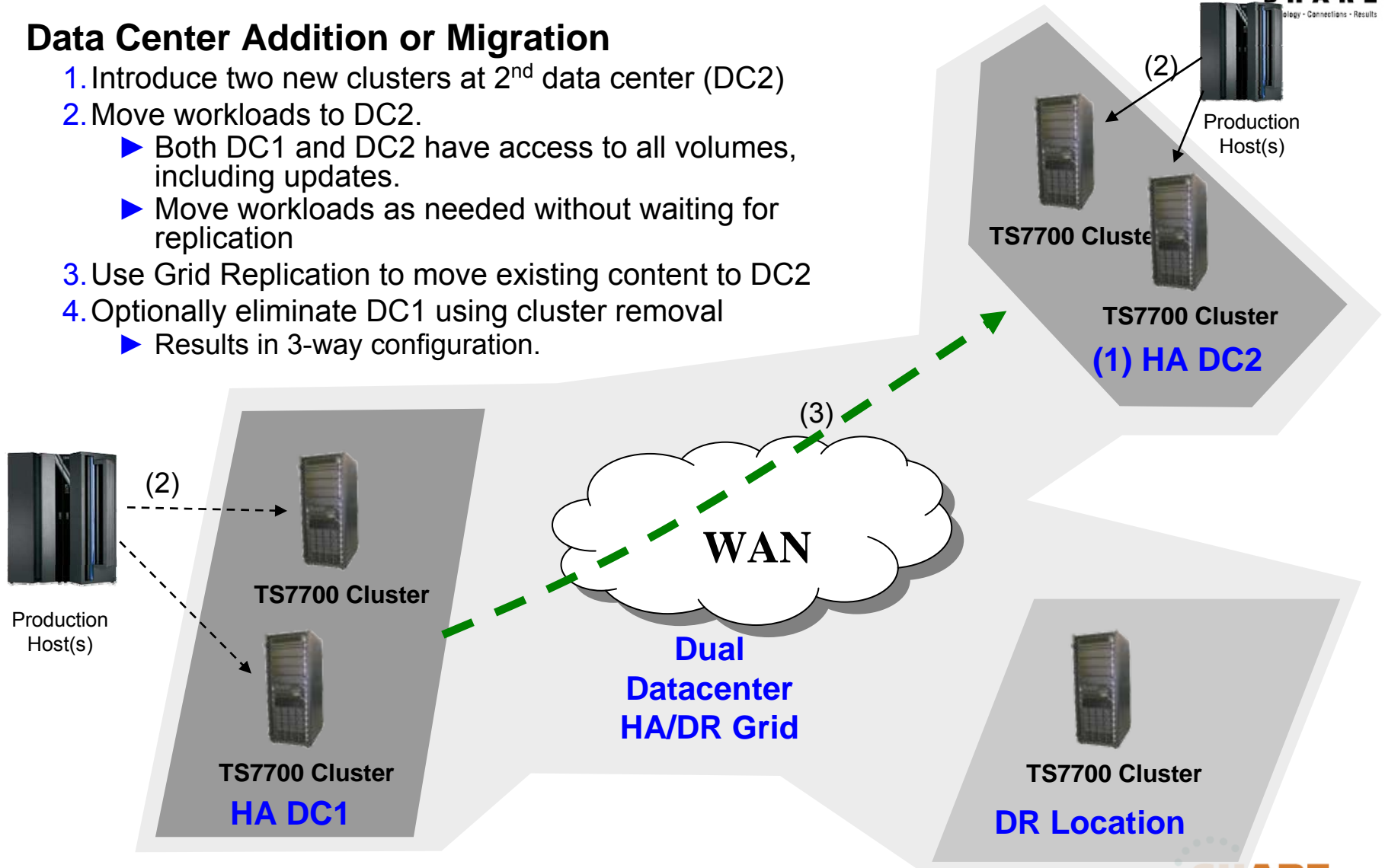2012

Tapes First Mainframe Synchronous Copy

- **Volume End Guaranteed/Immediate mode copy does not provide a Zero RPO**
  - Applications which stack datasets to tape such as DFSMShsm ML2 migrations can remove previously stacked datasets from DASD prior to end of volume processing
  - Content not yet replicated up to that point on tape is exposed to a single datacenter copy only on tape
- **TS7700 provides the first sync point granular Zero RPO synchronous copy method**
  - Up to two sites will be kept consistent after each implicit or explicit tape SYNC operation
  - Policy managed at volume granularity. Only volumes requiring sync are sync'd.
  - Provides applications, such as DFSMShsm, dataset level replication (Zero RPO!)
  - When duplexing fails, downgrade to sync-deferred or fail job based on policy management.
  - Additional Deferred and/or Immediate copies can occur once volume is closed
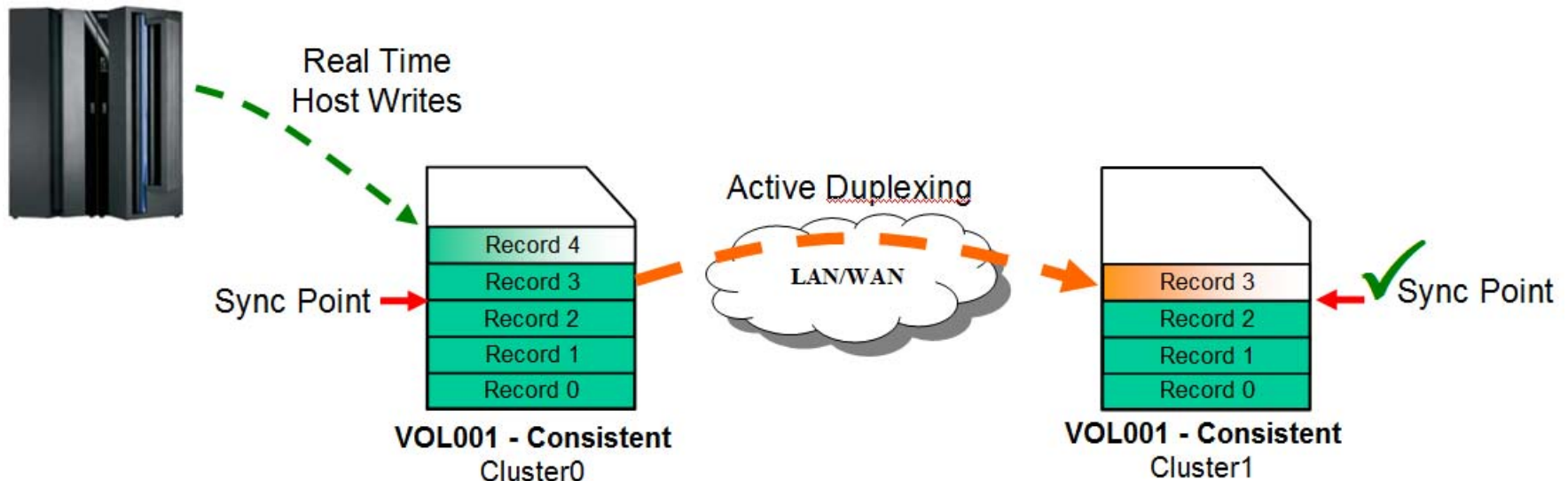
Data Center Migrations Made Easy
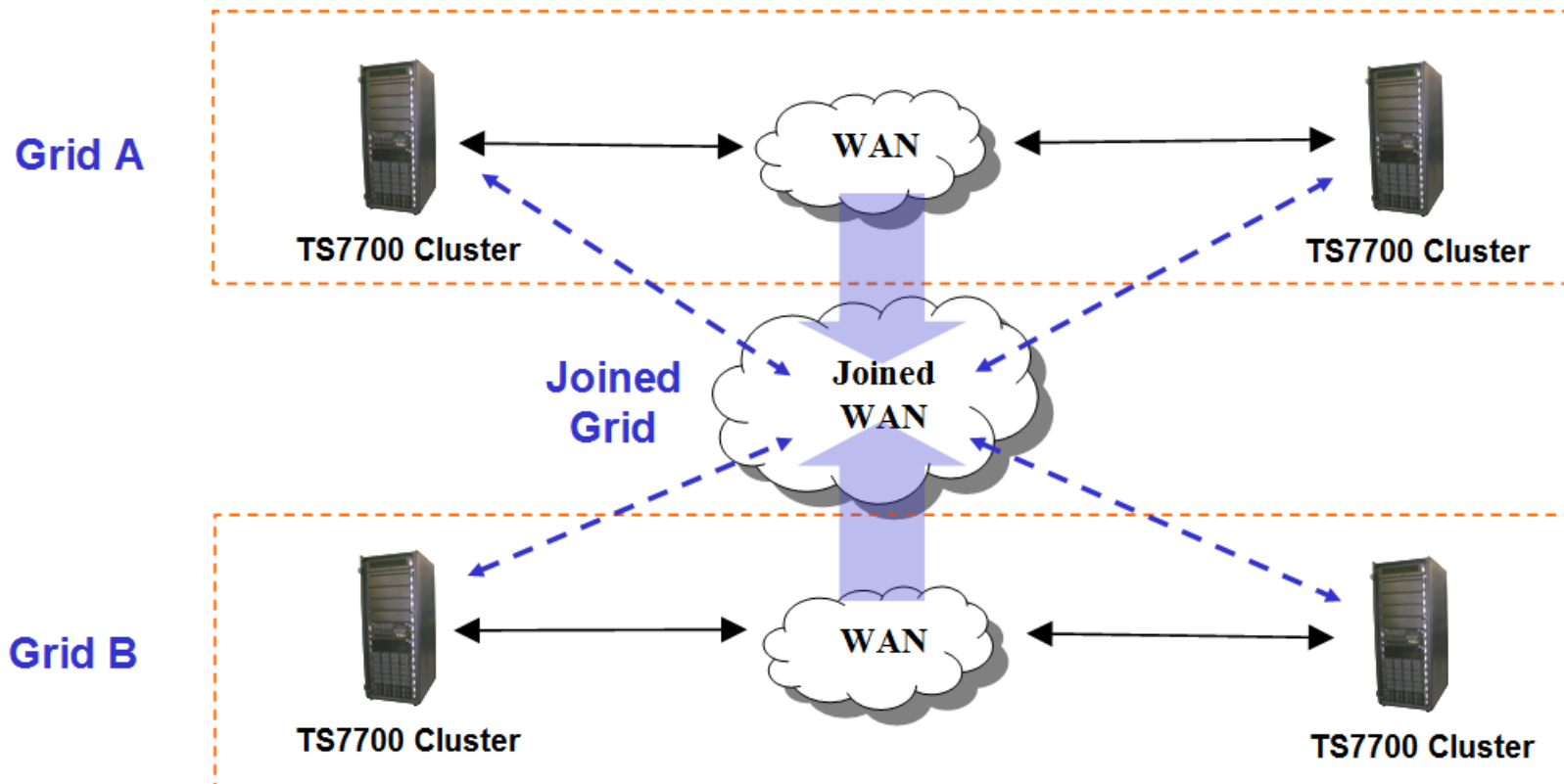
- **Data Center Addition or Migration**
    1. Introduce two new clusters at 2nd data center (DC2)
    2. Move workloads to DC2.
        - ► Both DC1 and DC2 have access to all volumes, including updates.
        - ► Move workloads as needed without waiting for replication
    3. Use Grid Replication to move existing content to DC2
    4. Optionally eliminate DC1 using cluster removal
        - ► Results in 3-way configuration.

(2)

Production Host(s)

TS7700 Cluster

TS7700 Cluster

(1) HA DC2

(3)

(2)

TS7700 Cluster

Production Host(s)

WAN

Dual Datacenter HA/DR Grid

TS7700 Cluster

HA DC1

TS7700 Cluster

DR Location

SHARE in Anaheim
2012

Grid Merge

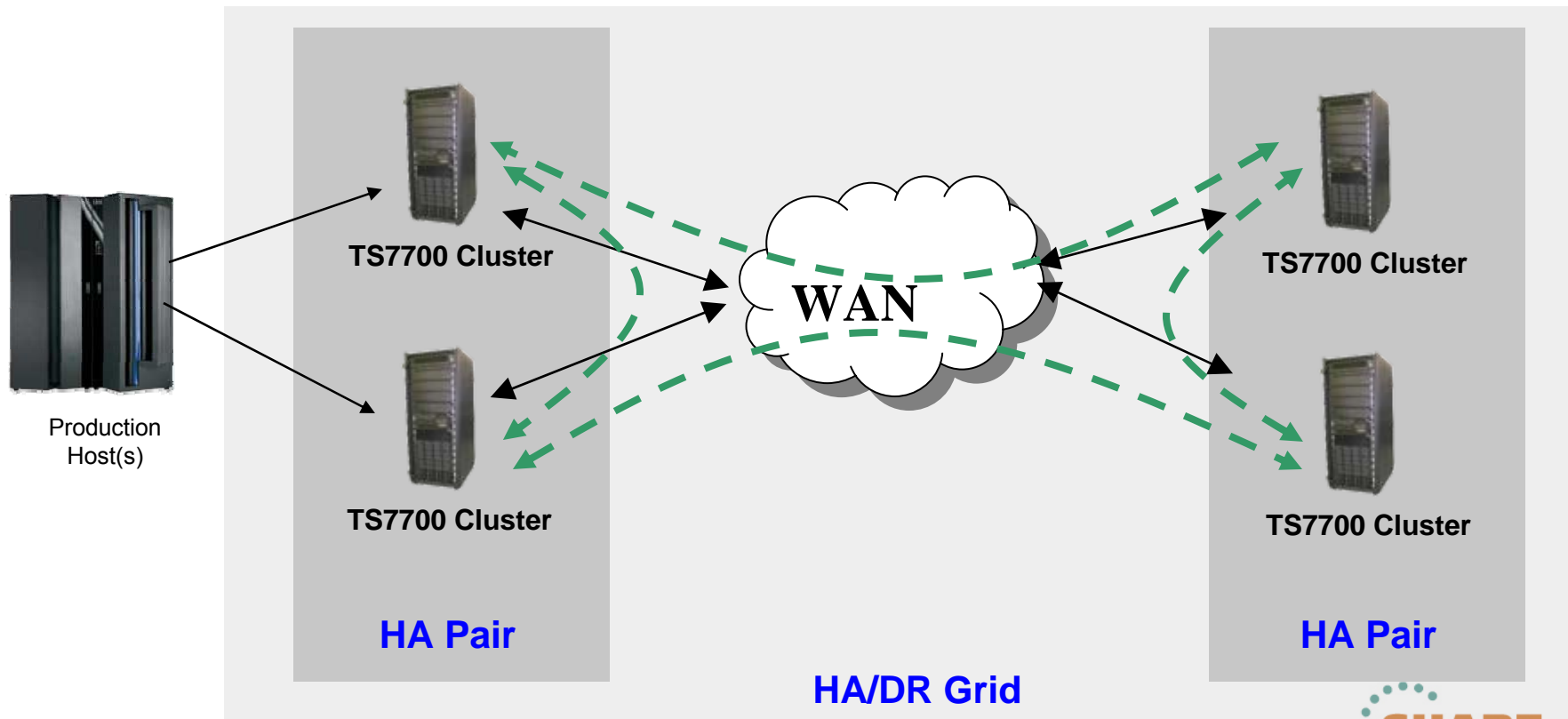- Merge Two Existing Grids into a single <u>Joined Grid</u>

3-Way HA/DR Grid Configurations

- Two production TS7700 clusters for HA both replicating into a more remote TS7700
  - Each production TS7700 can optionally replicate to its adjacent cluster
  - HA production clusters may be tightly coupled within the same location or spread apart at metro distances
  - Remote cluster contains all DR relevant data
    - Used for Disaster Recovery Testing
    - Used as a Disaster Recovery Location
    - Can also be used for production
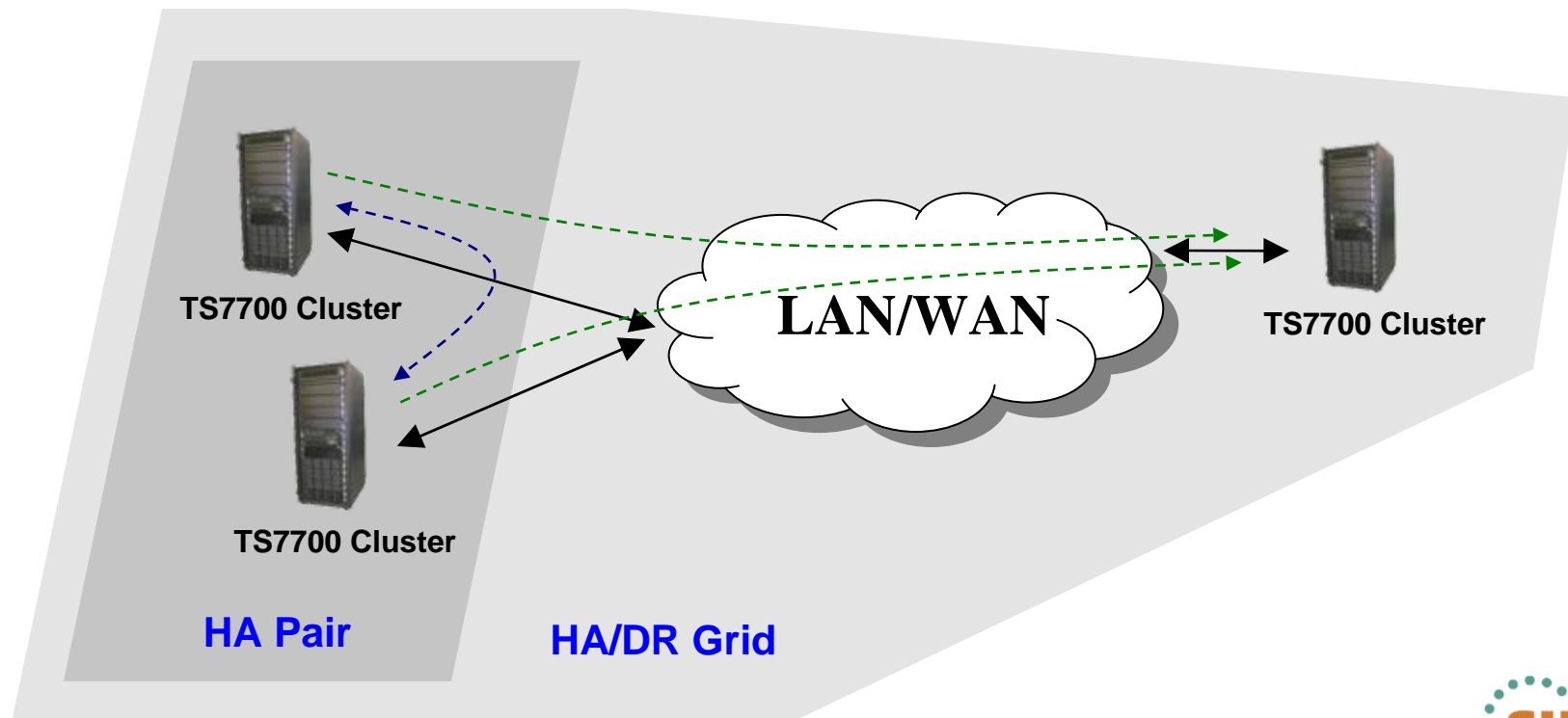
4-way HA/DR Configuration

- **Two production clusters at metro distances**
  - Host workload balanced across both clusters
  - Content replicates between local clusters and remote clusters
  - During an outage, all reads are local
  - HA pair clusters can be of mixed types (Hybrid)
- **Two remote DR clusters at metro distances**
  - Each DR cluster receives a copy of the production data

Production Host(s)

TS7700 Cluster

TS7700 Cluster

WAN

TS7700 Cluster

TS7700 Cluster

HA Pair

HA Pair
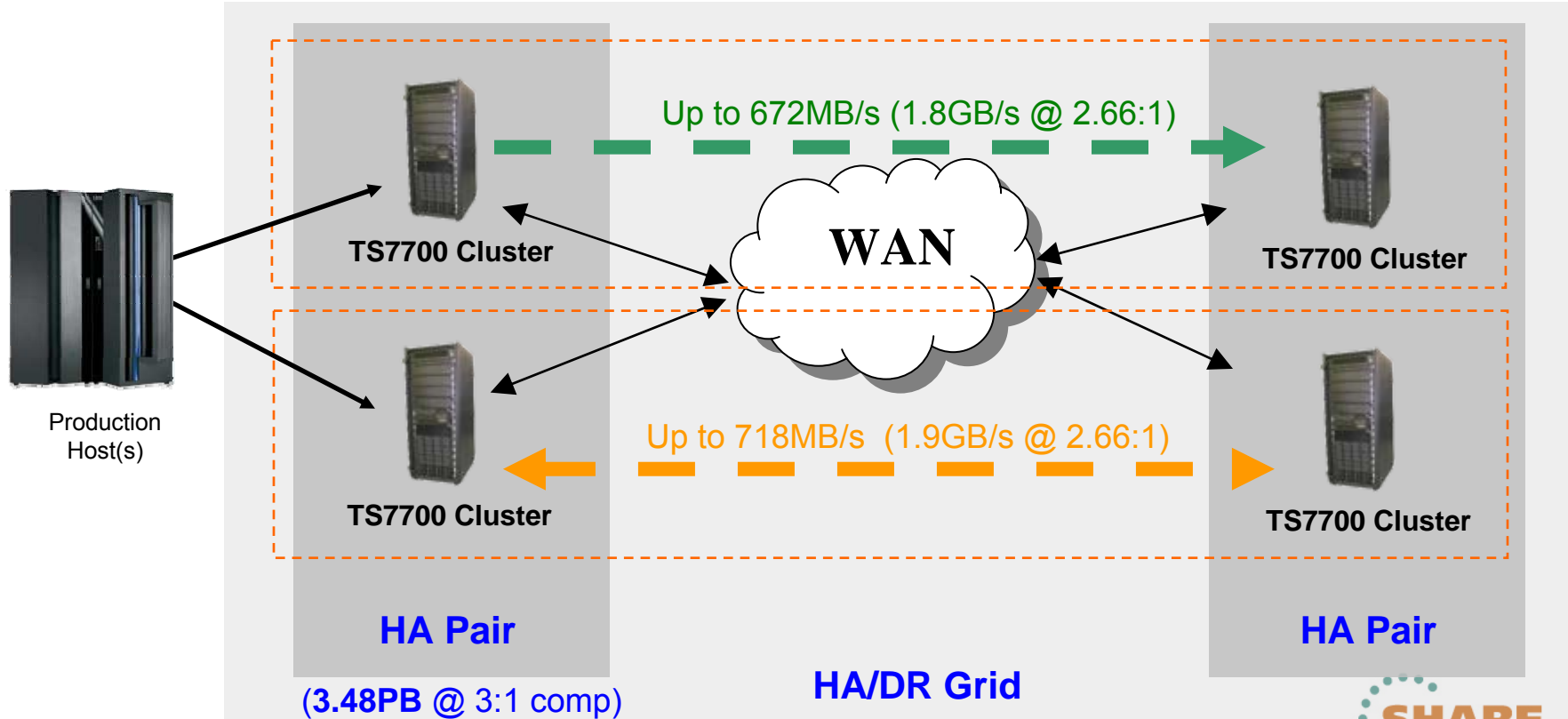
**HA/DR Grid**

SHARE in Anaheim

2012

4-way HA/DR Partitioned Configuration

- **Two production clusters at local or metro distances**
  - Host workload balanced across both clusters
  - Content written to particular mounting cluster is only replicated to one remote cluster
- **Two remote clusters at metro or DR distances**
  - Each remote cluster replicates content from one of the two production clusters
- **High availability at both production and secondary location without four copies of the data**
  - Same capacity as two 2-way configurations, with the high availability of a 4-way configuration
  - A true high available and disaster recoverable business continuous configuration !

Production Host(s)

TS7700 Cluster

Up to 672MB/s (1.8GB/s @ 2.66:1)

WAN

TS7700 Cluster

TS7700 Cluster

Up to 718MB/s (1.9GB/s @ 2.66:1)

TS7700 Cluster

HA Pair

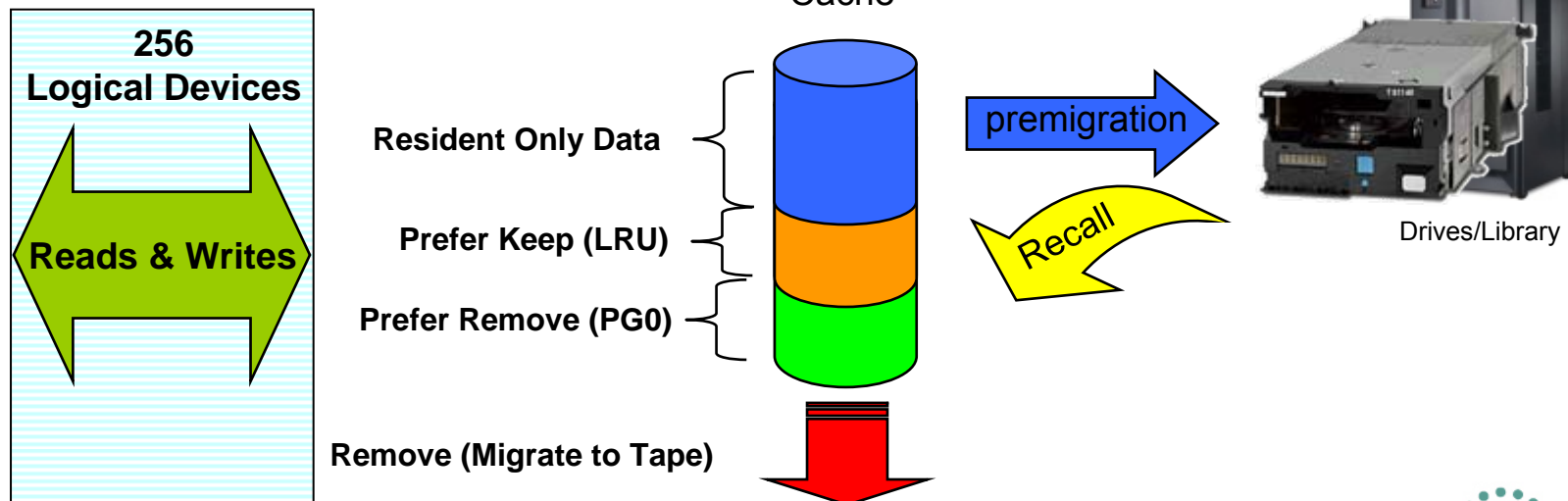(**3.48PB** @ 3:1 comp)

HA/DR Grid

HA Pair

4-way HA/DR Partitioned Configuration

- **When cluster outage occurs**
  - New jobs only utilize one half of configuration
  - Jobs for read can access content in both partitions

SHARE
in Anaheim
2012
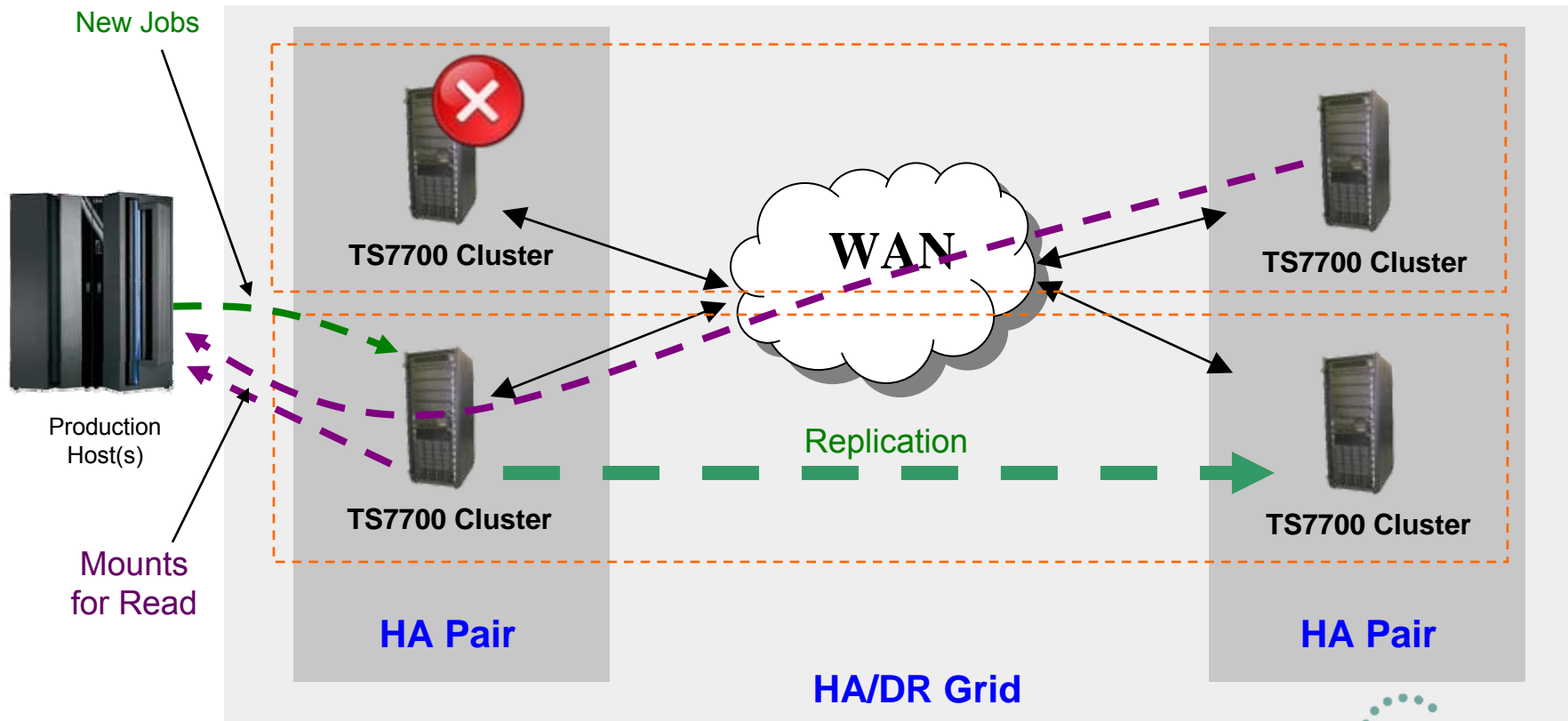
TS7740 Intelligent Tape Integration

- **Automatic Tier Management Between Disk Cache and Physical Tape**
  - All host reads and writes are through disk cache at disk speeds
    - The host is only aware of logical devices as the access point to all volumes
  - New or modified data is stacked/copied to physical tape within the TS3500 library
  - Data remains in disk cache using policy management
    - Data can be preferred to be flushed from cache (PG0)
    - Data can be kept in cache (PG1) which relies on an LRU algorithm
  - Aged or PG0 volumes are removed from disk cache automatically leaving a copy on tape
  - Just a few terabytes of disk cache provides the benefits of disk without giving up the benefits of physical tape

256 Logical Devices

Reads & Writes

Disk Cache

Resident Only Data

Prefer Keep (LRU)

Prefer Remove (PG0)

premigration

Recall

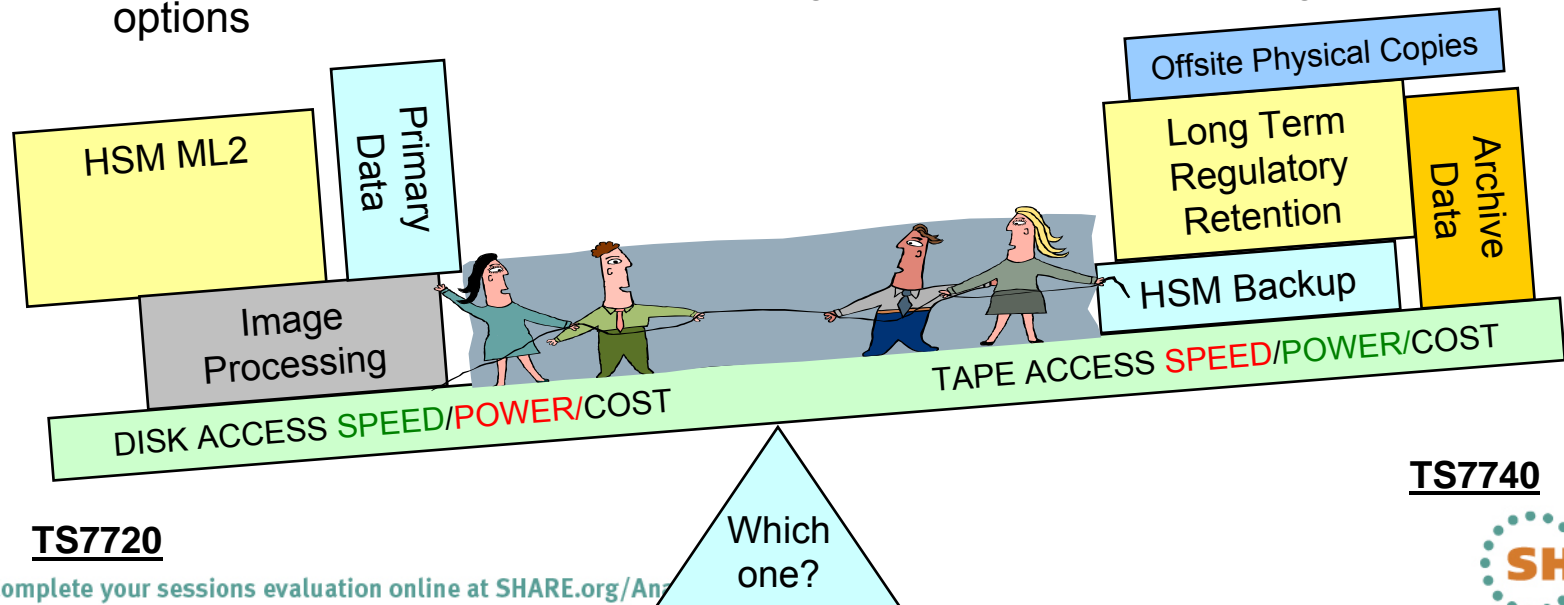Drives/Library

Remove (Migrate to Tape)

Hybrid Grid Configurations

- **Not one size fits all**
  - Disk only solutions are optimal for primary data applications or applications which need fast access times
  - Solutions tightly integrated with physical tape are best for archive and backup data
  - Total cost of ownership also factors into a customer's choice which doesn't always agree with usage case
- **Hybrid is the best of both technologies**
  - Using Grid architecture, intermix disk only TS7720 solutions with disk/tape TS7740 solutions
  - Allows customer to configure how their data is managed in order to fully benefit from an intermixed configuration
  - Allows IBM to be flexible in its offerings while also differentiating itself with smarter options

Offsite Physical Copies

HSM ML2 | Primary Data

Long Term Regulatory Retention | Archive Data

Image Processing | HSM Backup

DISK ACCESS SPEED/POWER/COST

TAPE ACCESS SPEED/POWER/COST

TS7720

TS7740

Which one?

SHARE in Anaheim

2012

Hybrid Intelligent Tape Integration

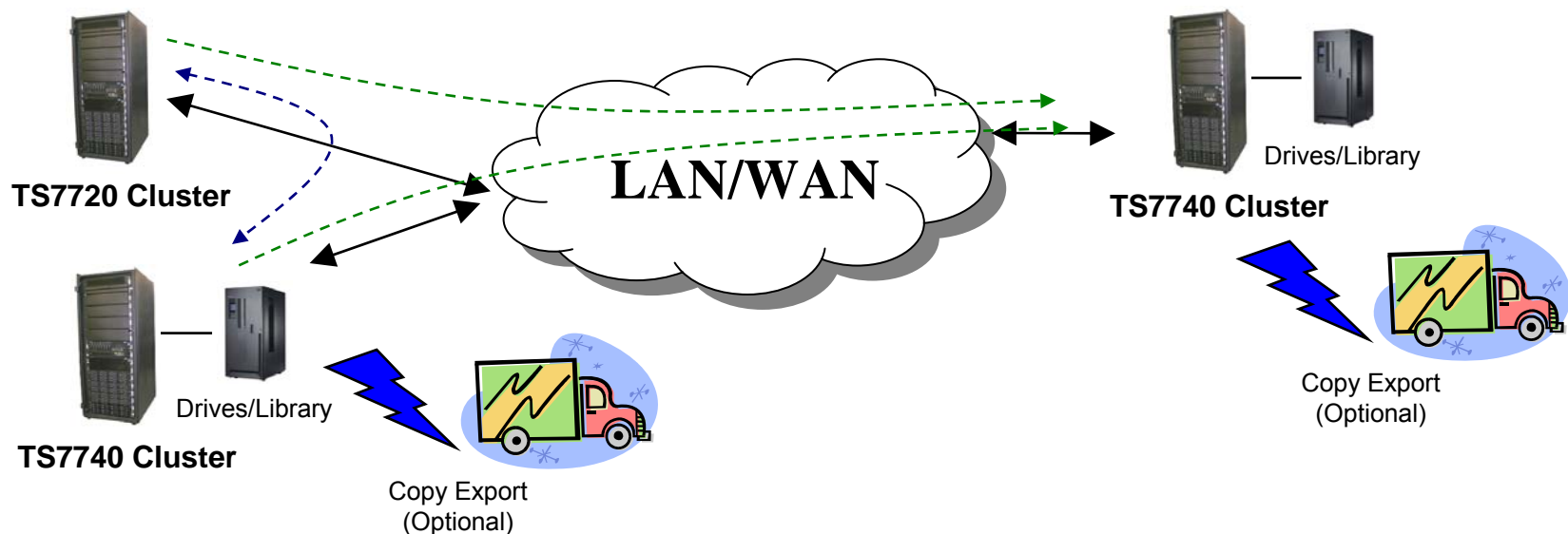- TS7720 clusters in a grid are designed to run at maximum capacity
- Through policy management, data can be:
  - **Pinned** – Reside in TS7720 disk cache indefinitely
  - **Retained** – Pinned for a duration of time since last access, then Keep or Remove
  - **Prefer Kept or Removed** – Two groups using LRU algorithm
- Data is removed from TS7720 disk cache after validating data already exists on TS7740s
- Automatic and nearly transparent integration of physical tape
  - Aged or archive data naturally ends up on tape within TS7740s
  - Other workloads can be pinned or retained in cache for 100% cache hits on reads

Pinned Volumes

Retained Volumes

Prefer Keep (LRU Order)

Prefer Remove (LRU Order)

TS7720 Cluster

LAN/WAN

Drives/Library

TS7740 Cluster
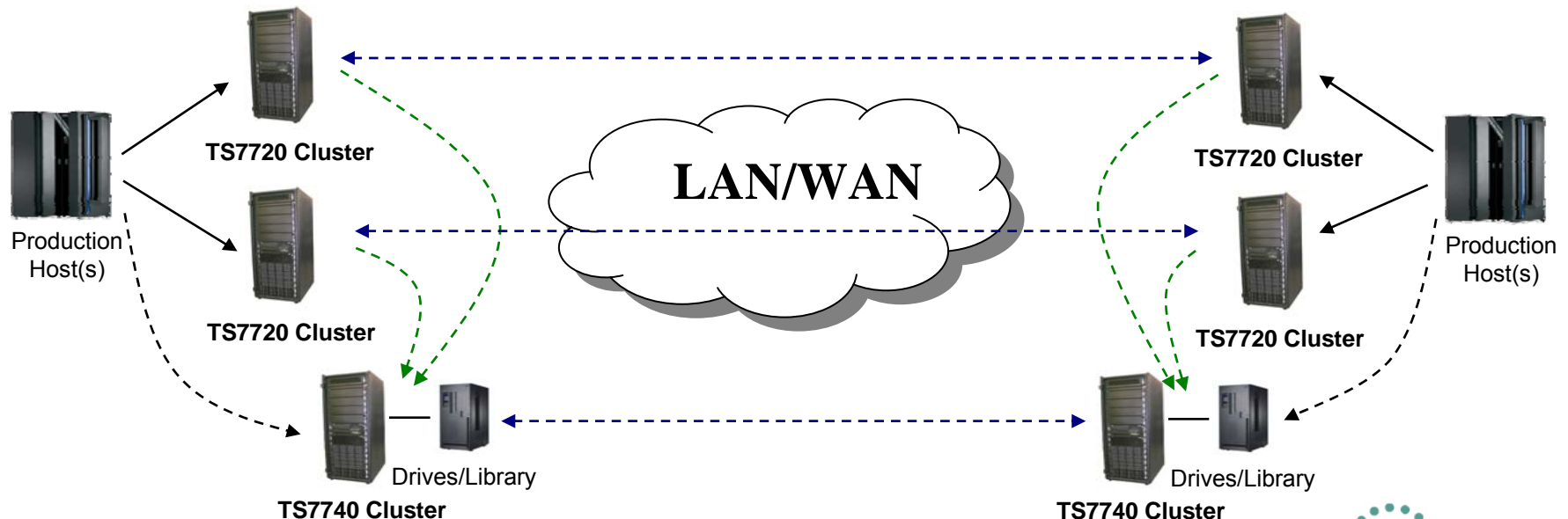
Drives/Library

TS7740 Cluster
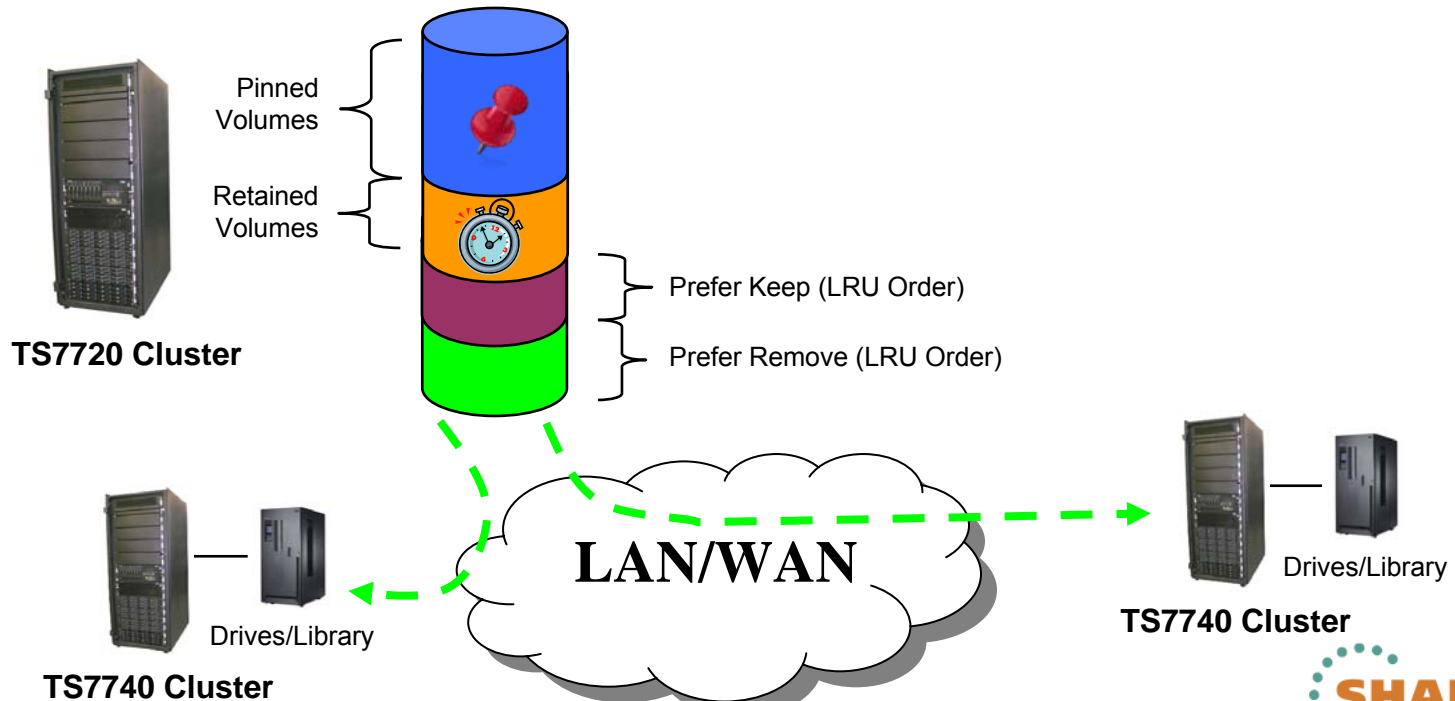
3-Way HA/DR Hybrid Grid Configuration

- **Production Hybrid Pair for HA both feeding into a common TS7740**
  - Deep disk cache at production for high cache hit ratios
  - Deep TS7740 capacity at both locations for cost effective long term archive & DR
  - Provides 580TB+ of HIGH PERFORMANCE production cache
  - Ability to have particular workloads favor the TS7720 while others favor the TS7740
- **Data migrates into the TS7740s**
  - If the TS7720 reaches capacity, the oldest data which has already been replicated to the TS7740 will be removed from the TS7720 cache.
  - Though not necessary, copy export can be used at both the production and DR locations

TS7720 Cluster

LAN/WAN

Drives/Library

TS7740 Cluster

Copy Export
(Optional)

Drives/Library

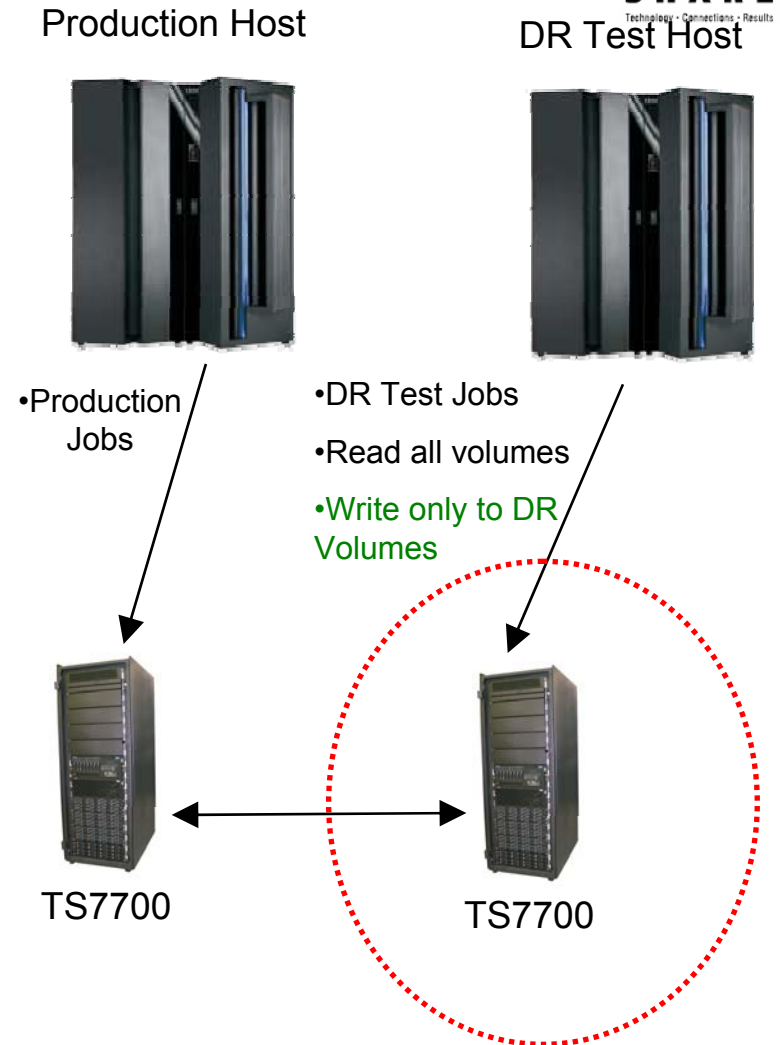**TS7740 Cluster**

Copy Export
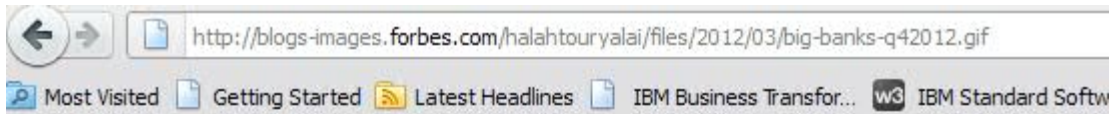(Optional)

Disk Centric Hybrid

- **Dual production data centers with deep disk cache**
  - Two TS7720 clusters with deep disk cache for high cache hit ratios
  - Deep TS7740 capacity at both locations for cost effective long term storage/archive
  - Ability to have particular workloads favor the TS7720 while others favor the TS7740
- **Data migrates into the TS7740s**
  - If the TS7720s reach full capacity, the oldest data which has already been replicated to the TS7740 will be removed from the TS7720 cache.
  - Policies can be designed so that specific workloads go directly to the TS7740s and not utilize TS7720 cache space
  - If mobile physical copies are required, copy export can be used within the TS7740 to eject duplexed copies

TS7720 Cluster · TS7720 Cluster · LAN/WAN · Production Host(s) · TS7720 Cluster · TS7720 Cluster · Drives/Library · TS7740 Cluster · Drives/Library · TS7740 Cluster

Disaster Recovery Testing

- Full support of concurrent testing
  - DR test host can run while production continues
  - Production data continues to replicate during the entire test
- Production data protection
  - Use TS7700 selective write protect to isolate DR test volumes by category for full read/write access while only allowing read access to production volumes
  - Access production volumes which have been returned to scratch as private volumes within DR location

Production Host

DR Test Host

•Production Jobs

•DR Test Jobs

•Read all volumes

•Write only to DR Volumes

TS7700

TS7700

SHARE
Technology · Connections · Results

http://blogs-images.forbes.com/halahtouryalai/files/2012/03/big-banks-q42012.gif

Most Visited | Getting Started | Latest Headlines | IBM Business Transfor... | IBM Standard Softw

Largest banks and thrifts in US by total assets
Pro forma for recent acquisitions

Q4'11 rank	Q3'11 rank	Change	Company (ticker)	City (state or territory)	Total assets ($B)	Total deposits ($B)
1	1	NC	JPMorgan Chase & Co. (JPM)[1]	New York (NY)	2,250.8	1,127.8
2	2	NC	Bank of America Corp. (BAC)	Charlotte (NC)	2,129.0	1,033.0
3	3	NC	Citigroup Inc. (C)	New York (NY)	1,873.9	865.9
4	4	NC	Wells Fargo & Co. (WFC)	San Francisco (CA)	1,313.9	920.1
5	5	NC	U.S. Bancorp (USB)	Minneapolis (MN)	340.1	230.9
6	7	↑	Capital One Financial Corp. (COF)[2]	McLean (VA)	328.6	211.2
7	6	↓	Bank of New York Mellon Corp. (BK)	New York (NY)	325.3	219.1
8	9	↑	PNC Financial Services Group Inc. (PNC)[3]	Pittsburgh (PA)	298.4	209.2
9	8	↓	HSBC North America Holdings Inc.[4]	New York (NY)	287.0	122.6
10	10	NC	State Street Corp. (STT)	Boston (MA)	216.8	157.3
11	11	NC	TD Bank US Holding Co.[5]	Portland (ME)	201.1	163.8
12	12	NC	SunTrust Banks Inc. (STI)	Atlanta (GA)	176.9	127.9
13	13	NC	BB&T Corp. (BBT)	Winston-Salem (NC)	174.6	124.9
14	14	NC	American Express Co. (AXP)	New York (NY)	152.3	43.0
15	15	NC	Citizens Financial Group Inc.[6]	Providence (RI)	129.8	93.0
16	16	NC	Regions Financial Corp. (RF)	Birmingham (AL)	127.1	95.6
17	17	NC	BMO Financial Corp.[7]	Chicago (IL)	117.4	75.2
18	18	NC	Fifth Third Bancorp (FITB)	Cincinnati (OH)	117.0	85.7
19	19	NC	Northern Trust Corp. (NTRS)	Chicago (IL)	100.2	82.7
20	22	↑	UnionBanCal Corp.[8]	San Francisco (CA)	95.5	69.0
21	21	NC	KeyCorp (KEY)	Cleveland (OH)	88.8	62.0
22	20	↓	Santander Holdings USA Inc.[9]	Boston (MA)	80.6	47.8
23	24	↑	BancWest Corp.[10]	San Francisco (CA)	78.1	55.0
24	23	↓	M&T Bank Corp. (MTB)	Buffalo (NY)	77.9	59.4
25	25	NC	Discover Financial Services (DFS)	Riverwoods (IL)	69.5	39.6
26	28	↑	Charles Schwab Bank[11]	Reno (NV)	66.1	60.9
27	26	↓	BBVA USA Bancshares Inc.[12]	Houston (TX)	63.1	46.1
28	27	↓	Comerica Inc. (CMA)	Dallas (TX)	61.0	47.8
29	29	NC	Huntington Bancshares Inc. (HBAN)	Columbus (OH)	54.5	43.3
30	30	NC	Zions BanCorp. (ZION)	Salt Lake City (UT)	53.1	42.9

TS7700 is used by more than 75% of the US Top Banks

Questions?

Thank You

Disclaimers and Trademarks 1 of 2

- Copyright© 2012 by International Business Machines Corporation.

- No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation.

- The performance data contained herein were obtained in a controlled, isolated environment. Results obtained in other operating environments may vary significantly. While IBM has reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. These values do not constitute a guarantee of performance. The use of this information or the implementation of any of the techniques discussed herein is a customer responsibility and depends on the customer's ability to evaluate and integrate them into their operating environment. Customers attempting to adapt these techniques to their own environments do so at their own risk.

- Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only

- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectually property rights, may be used instead. It is the user's responsibility to evaluate and verify the operation of any on-IBM product, program or service.

Disclaimers and Trademarks 2 of 2

- THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT.

- IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g. IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein.

- Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

- The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

  IBM Director of Licensing
  IBM Corporation
  North Castle Drive
  Armonk, NY 10504-1785
  U.S.A.

- The following terms are trademarks or registered trademarks of the IBM Corporation in either the United States, other countries or both.
  - IBM, TotalStorage, zSeries, pSeries, xSeries, S/390, ES/9000, AS/400, RS/6000
  - z/OS, z/VM, VM/ESA, OS/390, AIX, DFSMS/MVS, OS/2, OS/400, ESCON, Tivoli
  - iSeries, ES/3090, VSE/ESA, TPF, DFSMSdfp, DFSMSdss, DFSMShsm, DFSMSrmm, FICON,
- Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. Other company, product, and service names mentioned may be trademarks or registered trademarks of their respective companies.