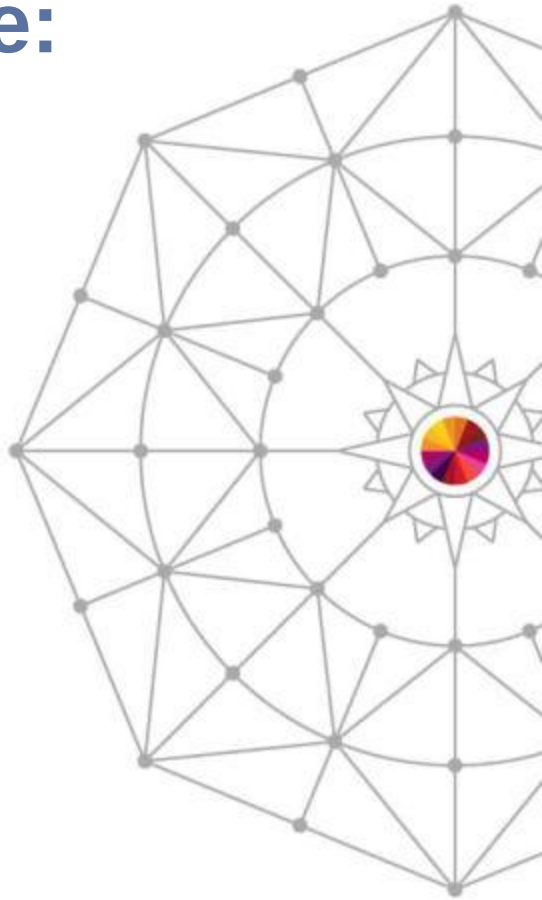
#SHAREorg

SHARE
Technology · Connections · Results

EMC Customer Experience: Universal Storage Consistency of DASD and Virtual Tape

Jim Erdahl
U.S.Bank

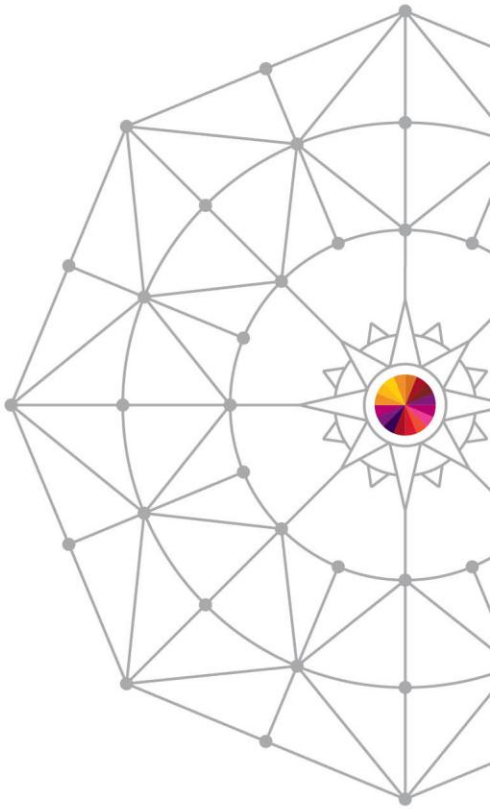
March 12, 2014
Session 14390



SHARE in Anaheim

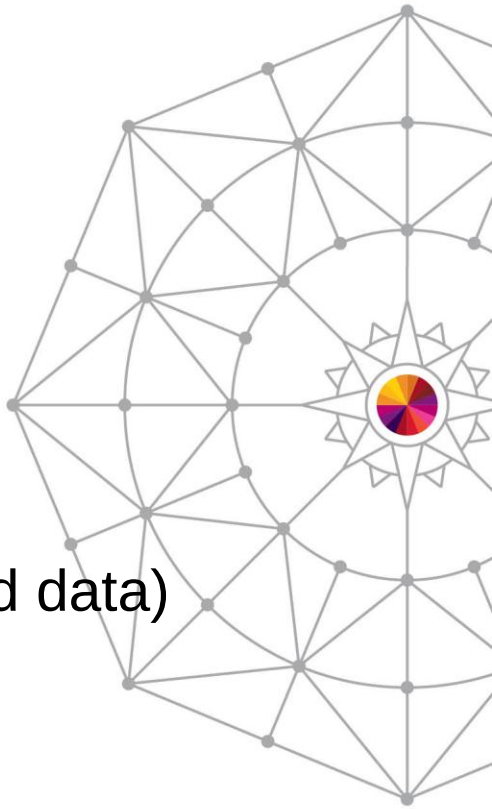
Agenda

- Tape background
- Storage consistency and resiliency
- Motivation for DLm8000
- DLm8000 implementation
- GDDR automation for DLm8000



Connotation of “tape”

- Nearline storage
- Lower cost than DASD
- Serial access
- Sequential stream of data
- Special catalog and retention controls (CA-1)
- No space allocation constructs
  - (5GB x 255 volumes = 1.275TB of compressed data)



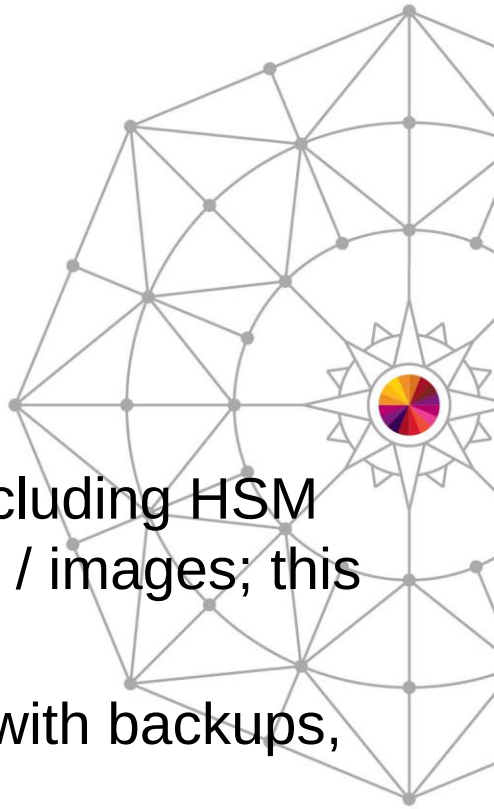
Disclaimer: tape no longer refers to physical media and drives

Characterization of tape data

- Archive / Retrieval
- Operational files
- Backup / Recovery

- Profile at US Bank
  - At least 60% of overall tape data is archive, including HSM ML2, CA View reports, and CMOD statements / images; this data is retrieved extensively
  - No more than 25% of tape data is associated with backups, including image copies and archive logs

***Tape data is critical!

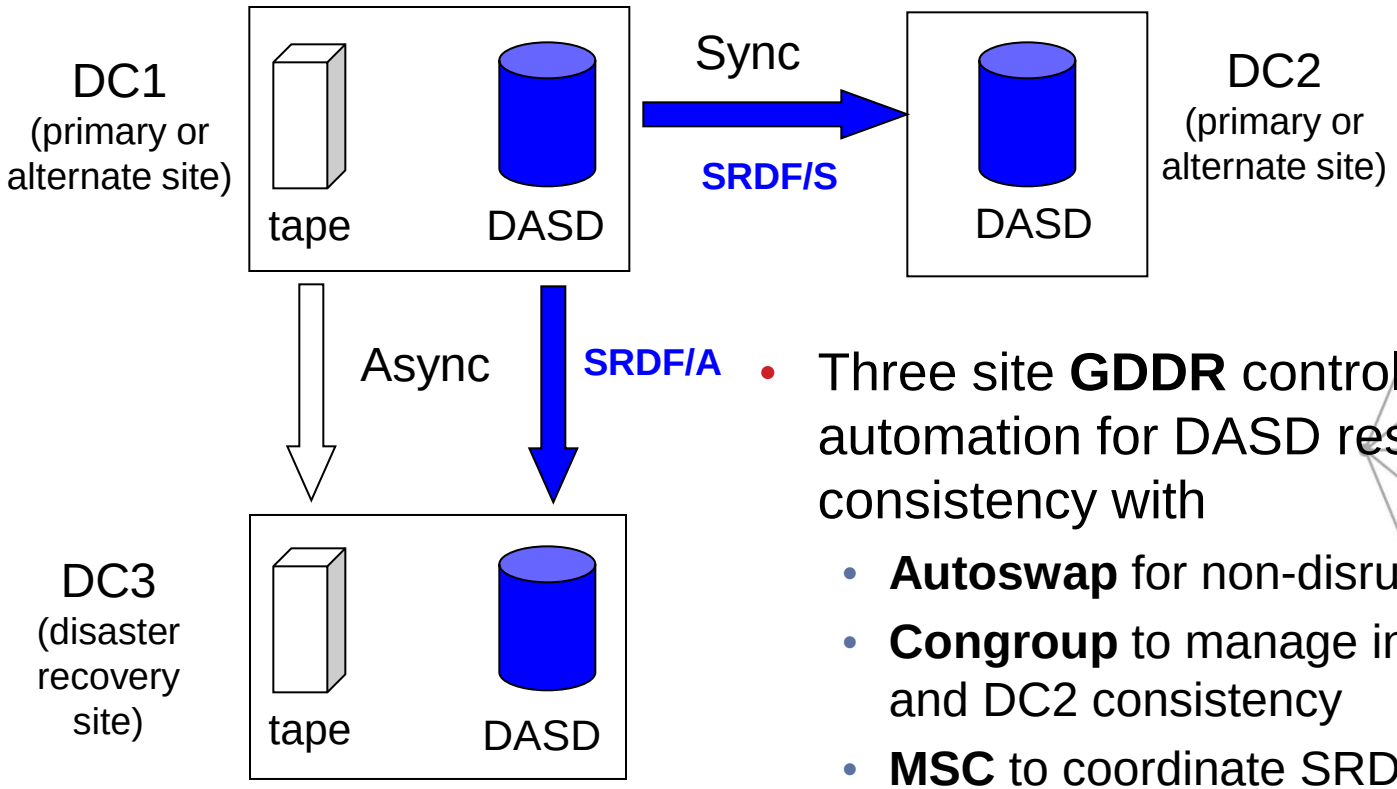


DLm Origin / Primer

Origin	Key Component	Role
Bus-Tech MDL	VTEs	Tape drive emulation, FICON interface, compression, tape library portal
EMC NAS (Celerra / CLARiiON)	Data Movers	File System sharing over an IP network / replication
	SATA disks	Data storage

- Tape files reside in File Systems; the name of the file is the Volser
- Multiple file systems are defined within a tape library to provide for sufficient volsers and capacity
- A single tape library is typically created per Sysplex and is mounted on specified VTE drives

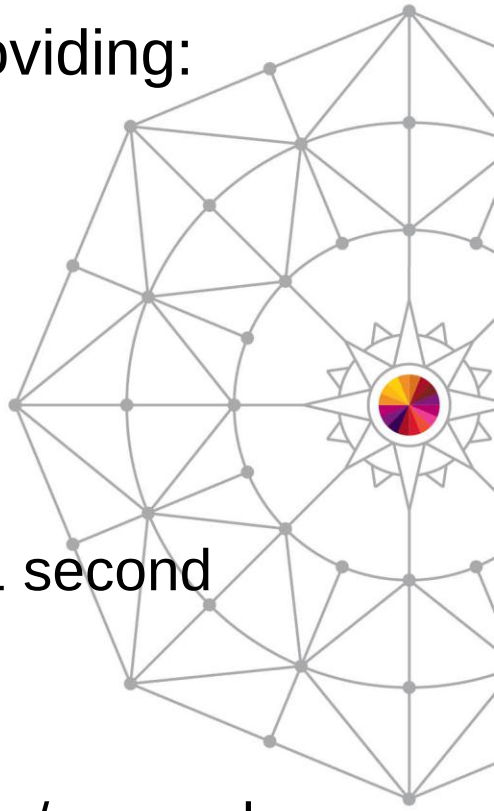
Original US Bank Storage Resiliency Configuration



- Three site **GDDR** control LPARs / automation for DASD resiliency and consistency with
  - **Autoswap** for non-disruptive failover
  - **Congroup** to manage integrity of SRDF/S and DC2 consistency
  - **MSC** to coordinate SRDF/A cycle switching across DASD arrays and DC3 consistency
  - **STAR** for differential resync to DC3 after a DASD swap

DLm6000 Experience at US Bank

- DLm implemented in 2008 for a technology refresh in conjunction with a Data Center migration, providing:
  - Consistently good performance
  - Low maintenance
  - High scalability and redundancy
- Notable metrics
  - Over 50,000 mounts per day
  - 99.9% of all tape mounts fulfilled in less than 1 second
  - 4.5 to 1 compression
  - Over 720 terabytes of usable capacity
  - Peak host read / write rate of 1,200 megabytes / second



If Life is so good, what is the motivation for DLm8000?

Problem #1 – Disaster Recovery Scenario
The “missing tape” problem at DC3 (DR site)

- Tape replication lags behind DASD replication (RPO measured in minutes versus seconds)
- In an out-of-region disaster declaration, tens and maybe hundreds of tape files closed immediately before the “Disaster” have not completely replicated
- But these files are defined in catalogs replicated on DASD (TMS, ICF catalogs, HSM CDSs, IMS RECON, DB2 BSDS, etc.)
- Hence, there are critical data inconsistencies between tape and DASD at DC3 (DR Site)

Problem #1 (continued)

During a Disaster Recovery at DC3….

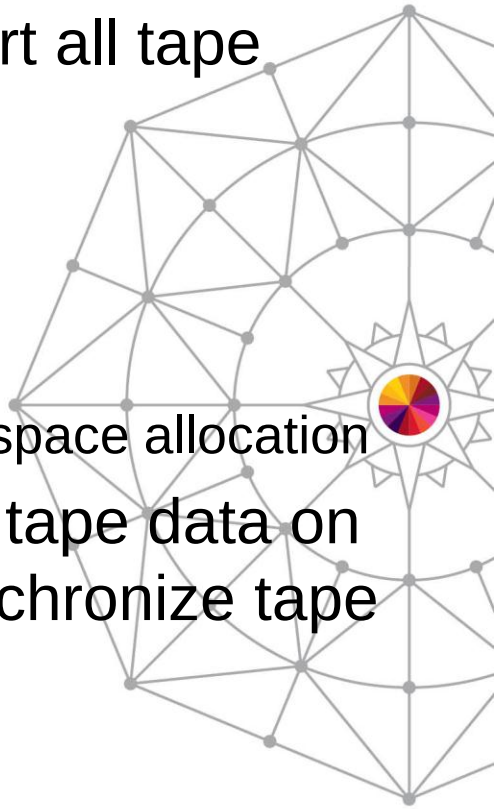
1. What if HSM recalls fail because of missing ML2 tape data?
2. What if the DBAs cannot perform database recoveries because archive logs are missing?
3. What is business and customer data archived to tape is missing?
4. How does this impact overall recovery time (RTO)?
5. Are Disaster Recovery capabilities adequate if tapes are missing?

Problem # 2 – Local Resiliency Scenario Local DASD resiliency is not sufficient

- Three site DASD configuration with synchronous replication between DC1 and DC2 and asynchronous replication to DC3. Two site tape configuration and asynchronous replication from DC1 to DC3.
- In the event of a DASD catastrophe at the primary site, a local DASD failover is performed non-disruptively with zero data loss, and asynchronous replication is re-instated to DC3.
- But, what if a catastrophe occurs to tape storage at DC1?
  - Mainframe processing will come to a grinding halt.
  - A disruptive recovery can be performed at DC3, but this is a last resort.

What are the options?

- To solve problems #1 and #2, why not convert all tape allocations to DASD?
  - The cost is prohibitive, but
  - Space allocation is the impediment
    - SMS and Data Classes are not adequate
    - Massive JCL conversion is required to stipulate space allocation
- To solve problems #1 and #2, why not store tape data on the same storage platform as DASD and synchronize tape and DASD replication and failover?



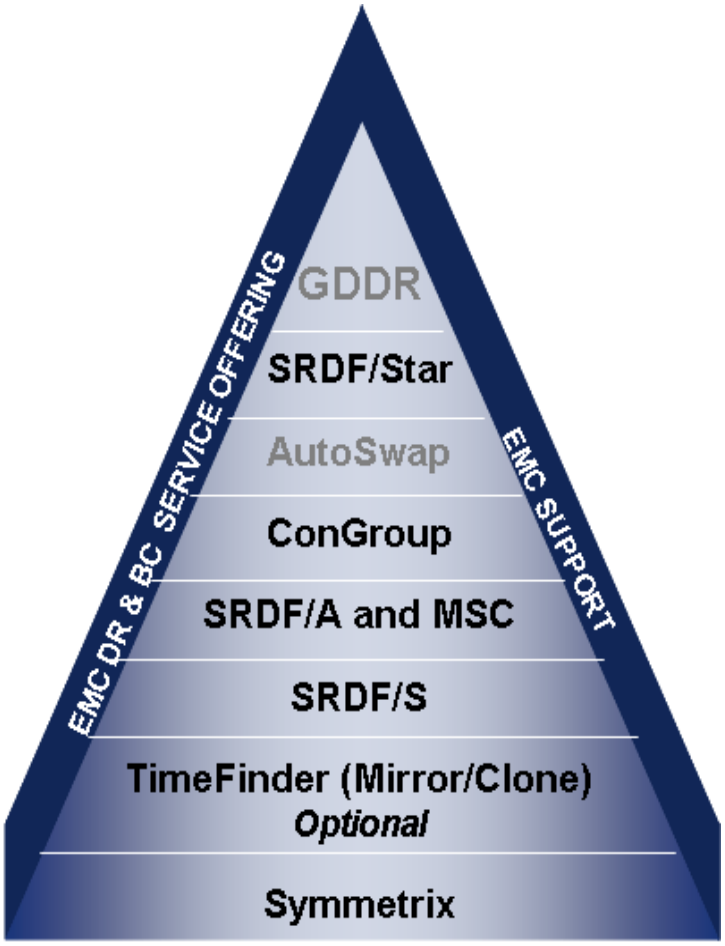
Today's Storage Management Mission - Protection

- Storage consistency: control mechanisms for storage replication which preserve dependent write consistency of data, ensuring that replicated data is in a consistent state
- Storage resiliency: mechanisms which protect or recover from storage failures or disruptions.

Tape backups
Hardware redundancy
RAID

Synchronous replication
Asynchronous replication
Automated failover
Continuous Data Protection

EMC has the technologies, but…



EMC Foundation Technologies

1. Can DLm support a Symmetrix backend?

Yes

2. Can SRDF/S and SRDF/A handle enormous tape workloads without impacts?

Yes

3. Can we achieve “universal data consistency” between DASD and tape at DC2 and DC3?

Yes

4. Can GDDR manage a SRDF/STAR configuration with Autoswap which includes the tape infrastructure?

Yes

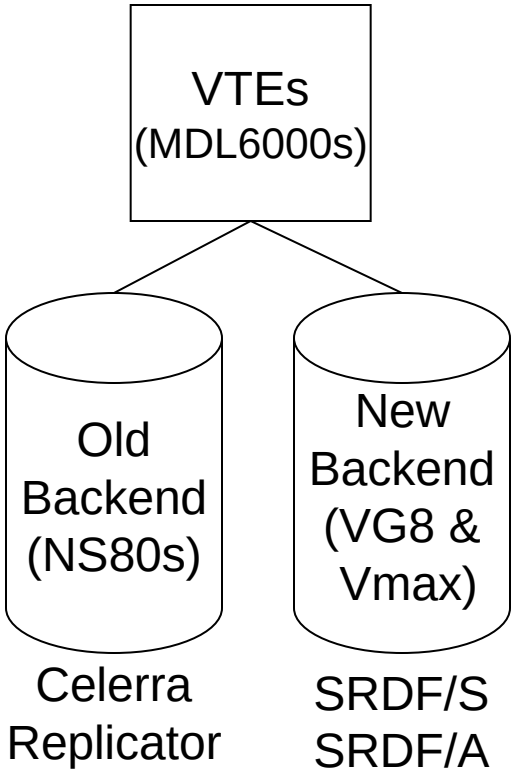
Creation of DLm8000

- US Bank articulated tape consistency and resiliency requirements in executive briefings and engineering round tables
- EMC solicited requirements from other customers and created a business case
- EMC performed a proof of concept validating overall functionality including failover processes
- EMC built a full scale configuration, based on US Bank's capacity requirements; performed final validation of replication, performance, and failover capabilities
- GDDR automation designed and developed with significant collaboration across product divisions
- US Bank began implementation in June, 2013

DLm8000 – What is under the hood?

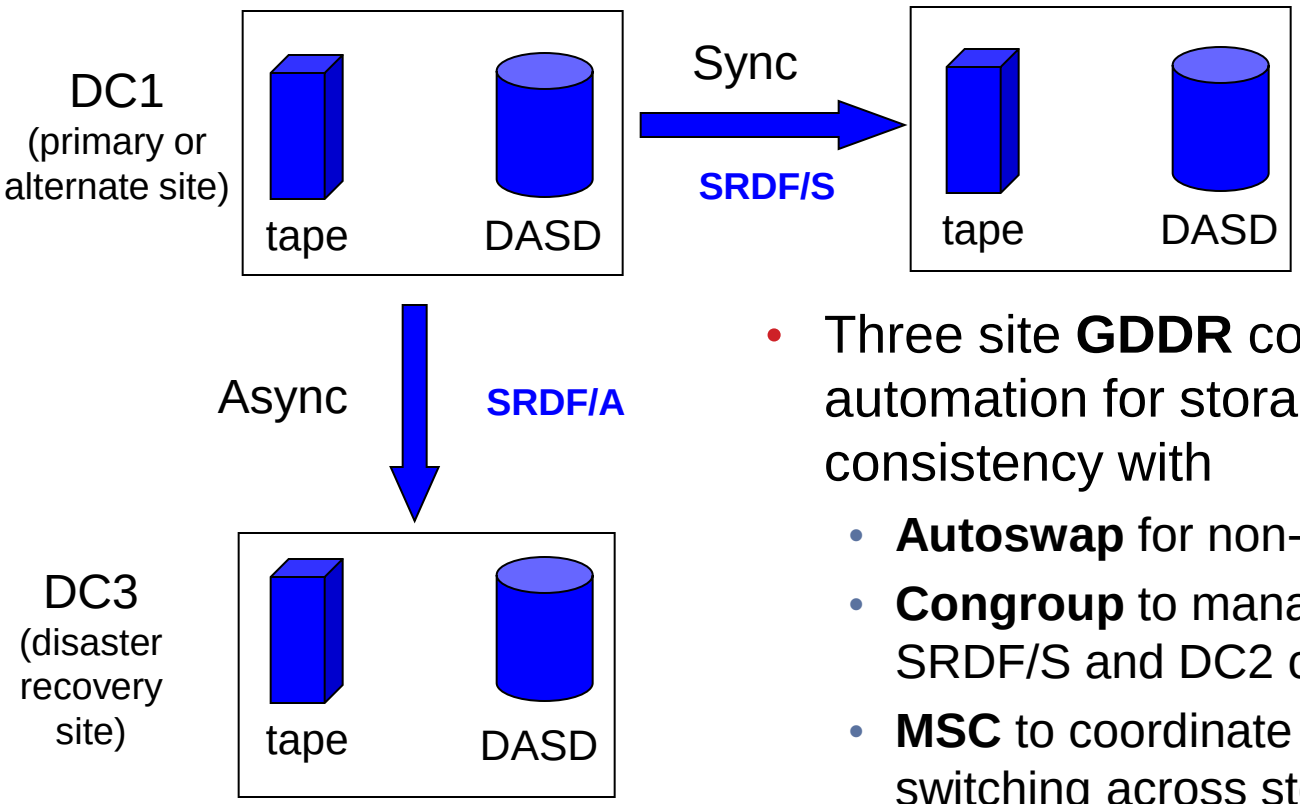
Product Line	Key Component	Role
MDL 6000	VTEs	Tape drive emulation, FICON interface, compression, tape library portal
VNX VG8	Data Movers	File System sharing over an IP network
Vmax	SATA FBA drives / SRDF	Data storage / replication

Migration to the DLm8000



1. Configured new backend at all three sites
2. Setup SRDF/A and SRDF/S
3. Defined File Systems on new backend
4. Partitioned “old” and “new” file systems with DLm storage classes
5. Updated Scratch synonyms to control scratch allocations by storage class
6. Deployed outboard DLm migration utility to copy tape files
7. Maintained dual replication from old backend and new backend during the migration
8. Incorporated tape into GDDR along with DASD (ConGroup, MSC, STAR, etc.) – March, 2013

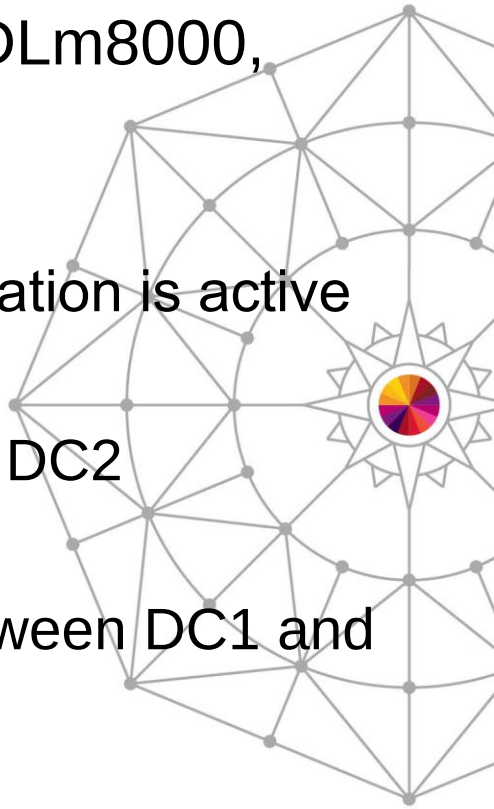
Current US Bank Storage Resiliency Configuration



- Three site **GDDR** control LPARs / automation for storage resiliency and consistency with
  - **Autoswap** for non-disruptive failover
  - **Congroup** to manage integrity of SRDF/S and DC2 consistency
  - **MSC** to coordinate SRDF/A cycle switching across storage arrays and DC3 consistency
  - **STAR** for differential resync to DC3 after a storage swap

How do we monitor / control all of this?

- GDDR provides comprehensive support for DLm8000, including scripts to:
  - Perform a disaster restart at DC3
  - Test from BCV's at DC3 and DC2, while replication is active
  - Restart SRDF/A replication to DC3
  - Restart SRDF/S replication between DC1 and DC2
  - Planned storage swap between DC1 and DC2
  - Recover from an unplanned storage swap between DC1 and DC2



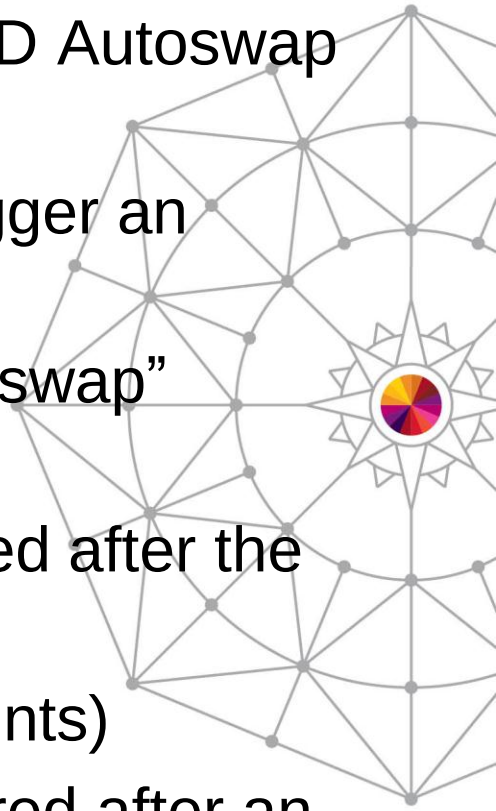
GDDR automation for a planned storage swap between DC1 and DC2 (high level steps)

Once tape workload is quiesced, GDDR script is initiated…

1. Tape drives varied offline
2. Tape configuration disabled at “swap from” site
3. DASD Autoswap and SRDF swap (R1s not ready, R2s ready and R/W)
4. Failover of Data Movers to “swap to” site
5. VTEs started at “swap to” site
6. Tape drives varied online and tape processing resumes
7. Recovery of SRDF/S, SRDF/A, and STAR

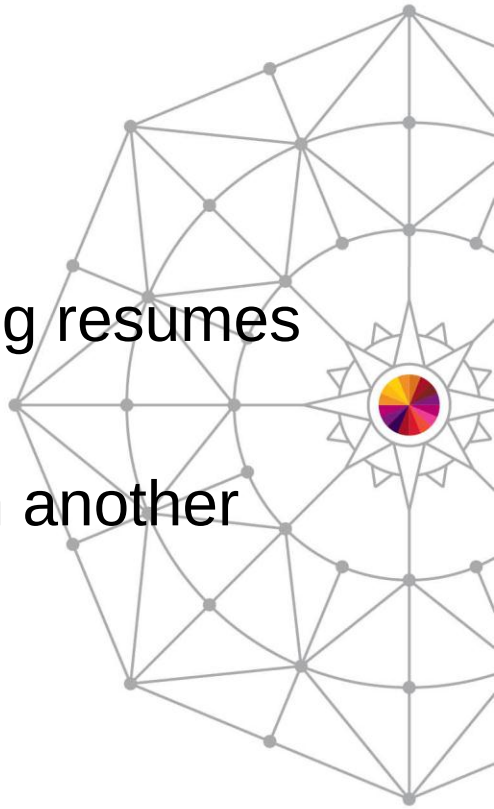
Unplanned storage swap between DC1 and DC2

- Loss of access to CKD devices triggers DASD Autoswap and SRDF swap
- Note: a tape infrastructure issue does not trigger an unplanned swap
- In-flight tape processing fails – no tape “Autoswap” functionality
- Failed in-flight tape jobs need to be res-started after the tape swap
- No data loss for closed tape files (or sync points)
- GDDR recovery script is automatically triggered after an unplanned swap



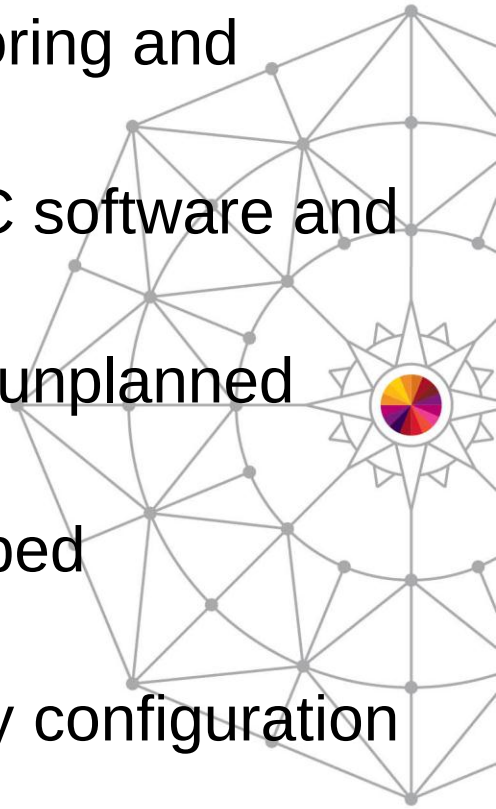
GDDR script to recover from unplanned swap (high level steps)

- Tape drives are varied offline
- Failover of Data Movers to “swap to” site
- VTEs started at “swap to” site
- Tape drives varied online and tape processing resumes
- Recovery of SRDF/A
- Recovery of SRDF/S and STAR initiated with another script



GDDR Value at US Bank

- Provides sophisticated automation for monitoring and management
- Handles coordination of comprehensive EMC software and hardware stack
- Provides recovery for critical events such as unplanned swaps and SRDF/A outages
- Minimizes requirements for internally developed automation and procedures
- Indispensable for a multi-site, high availability configuration



DLm8000 Value Proposition – Universal Storage Consistency and Resiliency

- Identical storage platform for DASD and tape
- Synchronous and asynchronous replication without impacts
- Universal consistency between tape and DASD at DC2 and DC3, enabled with ConGroup and MSC
- GDDR automation to manage overall storage replication, failover, and recovery

Questions ?