Data Deduplication Makes It Practical to Replicate Your Tape Data for Disaster Recovery
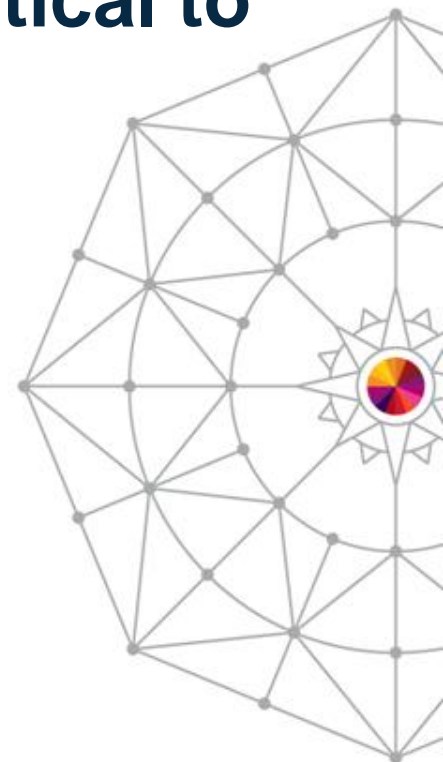
Scott James
VP Global Alliances
Luminex Software, Inc.

Randy Fleenor
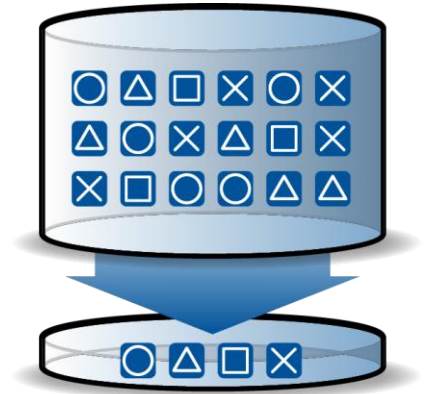*Worldwide Data Protection Management*
IBM Corporation

Discussion Topics

- Data Deduplication: "It's not just for reducing storage requirements"

- Compression/Deduplication Technologies

- Typical Customer Goals & Requirements

- Luminex Tape Analysis & Sizing

- ProtecTIER Mainframe Edition (ME)

- What Can ProtecTIER ME Do For You?

Data Deduplication: It's not just for reducing storage requirements

When virtual tape solutions include **data deduplication**, the **network bandwidth requirements** for replication are **dramatically reduced**.

Host Activity   **Deduplication**   **Replication**

Host Data

Replicated Data

WAN
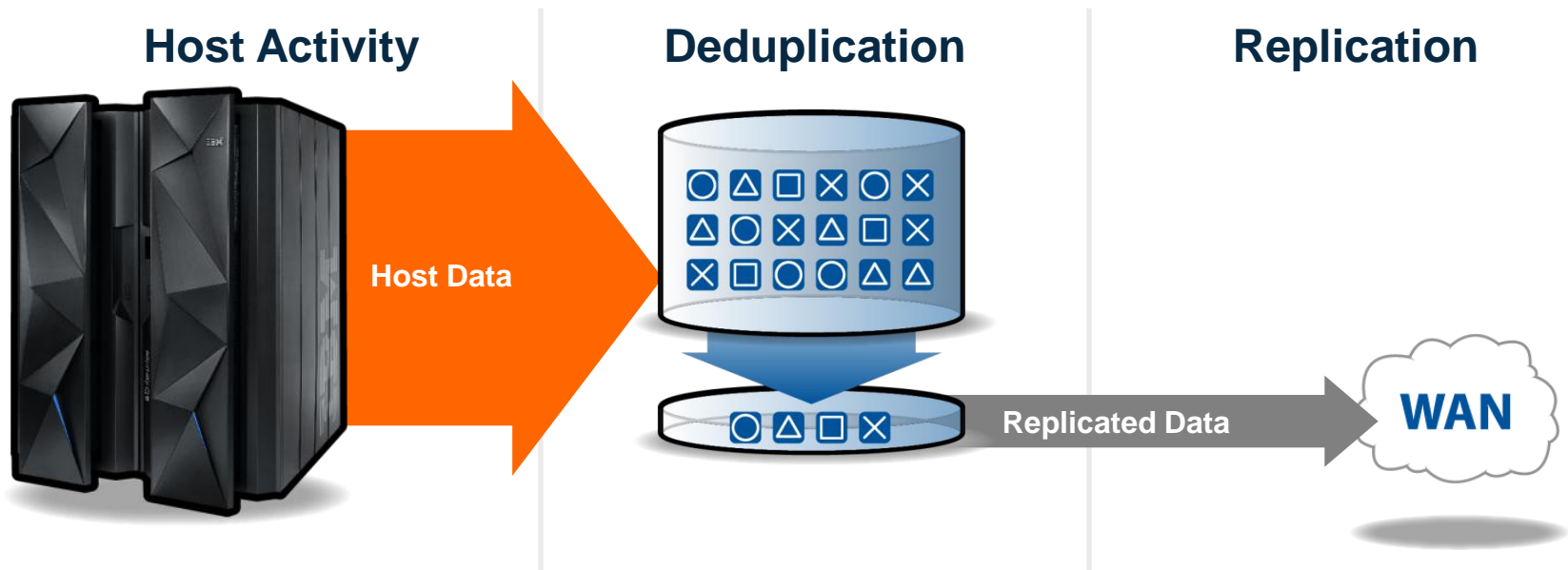
Customer Goals and Requirements

Virtual Tape Solutions with Deduplication
It's Not Just for Replication for DR

Replication of Tape Data for DR	**Replace Aging Infrastructure**	**Improve Data Security**
Reduce or Eliminate Physical Tape	**Increase Performance/ Efficiency**	**Higher Availability**
Reduce Costs	**Improve Disaster Recovery Plan**	**Common Storage Infrastructure**

Our Experience with Mainframe Compression & Data Deduplication Technologies

- **Compression (Typically 2–4x)**
  - Host-based
  - Physical Tape
  - Disk Storage
  - Control Unit

- **Data Deduplication**
  - WAN Optimization (up to 2x)
  - Disk Storage (8x–20x+ depending on data type)
    - IBM ProtecTIER, HP D2D/StoreOnce, Quantum DXi & EMC Data Domain

- **Other Considerations:**
  - Compressing or encrypting data before dedupe will reduce your dedupe rates
  - Read performance – Varies substantially by dedupe storage vendor
  - Support for both mainframe, iSeries and open systems required?

Luminex At-A-Glance

- Enterprise-class solution vendor focused on mainframe virtual tape

- 25+ years delivering mainframe virtual tape solutions worldwide

- Installed in all major industries on 6 continents

- Solutions scale seamlessly from entry level to multi-PB data centers

- ***The clear majority*** of mainframe deduplication customers rely on Luminex Channel Gateways/CGX

Sizing – Storage Capacity Requirements
TMC Analysis

Program	Description	Category	Capacity (GB)	Comp. Rate	Comp. Capacity (GB)	Dedupe Rate	Dedupe Capacity (GB)
ARCCTL	IBM DFSMShsm Migration	Archive and Compliance	27,018.30	3.00	9,006.10	8.00	3,377.29
IDCAMS	IBM VSAM Utility (Define, Repro, Export, etc.)	Backup and Recovery	22,177.91	3.00	7,392.64	15.00	1,478.53
ADRDSSU	IBM DFSMSdss Full Volume / File Disk Backup Utility	Backup and Recovery	3,559.18	3.00	1,186.39	15.00	237.28
IFASMFDP	IBM SMF Data Offload Utility	Archive and Compliance	3,413.77	3.00	1,137.92	8.00	426.72
SAS	SAS Institute, Inc. Statistical Analysis Program	Primary Operational	1,829.80	3.00	609.93	8.00	228.73
DSNUTILB	DB2 Utilities (Load/Unload, etc.)	Backup and Recovery	1,805.20	3.00	601.73	15.00	120.35
IEBGENER	IBM Sequential File Copy Utility	Primary Operational	919.44	3.00	306.48	8.00	114.93
BAKPROD8		Other	431.96	3.00	143.99	8.00	54.00
DRCPYDAT		Other	302.47	3.00	100.82	8.00	37.81
SORT	IBM SORT Utility	Primary Operational	291.78	3.00	97.26	8.00	36.47
Other		Other	1,943.64	3.00	647.88	8.00	242.96
		Summary	**63,693.45**	**3.00**	**21,231.15**	**10.02**	**6,355.05**

Legend:
- ☐ Archive and Compliance
- ☐ Backup and Recovery
- ☐ Primary Operational
- ☐ Other
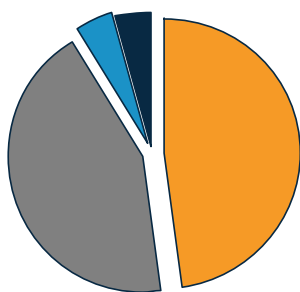
Uncompressed Capacity   Compressed Capacity   Deduplicated Capacity

Sizing – Network Bandwidth Requirements
TMC & SMF Analysis

Gigabytes Created by Day
Host versus Compressed versus Dedup

2,099 GB

2,099.26

700 GB

1,171.77

Host_GB
Compressed_GB
Dedup_GB

699.73

390.58

140 GB

139.95

88.27

<-- Least Recent --   Date Created   -- Most Recent -->
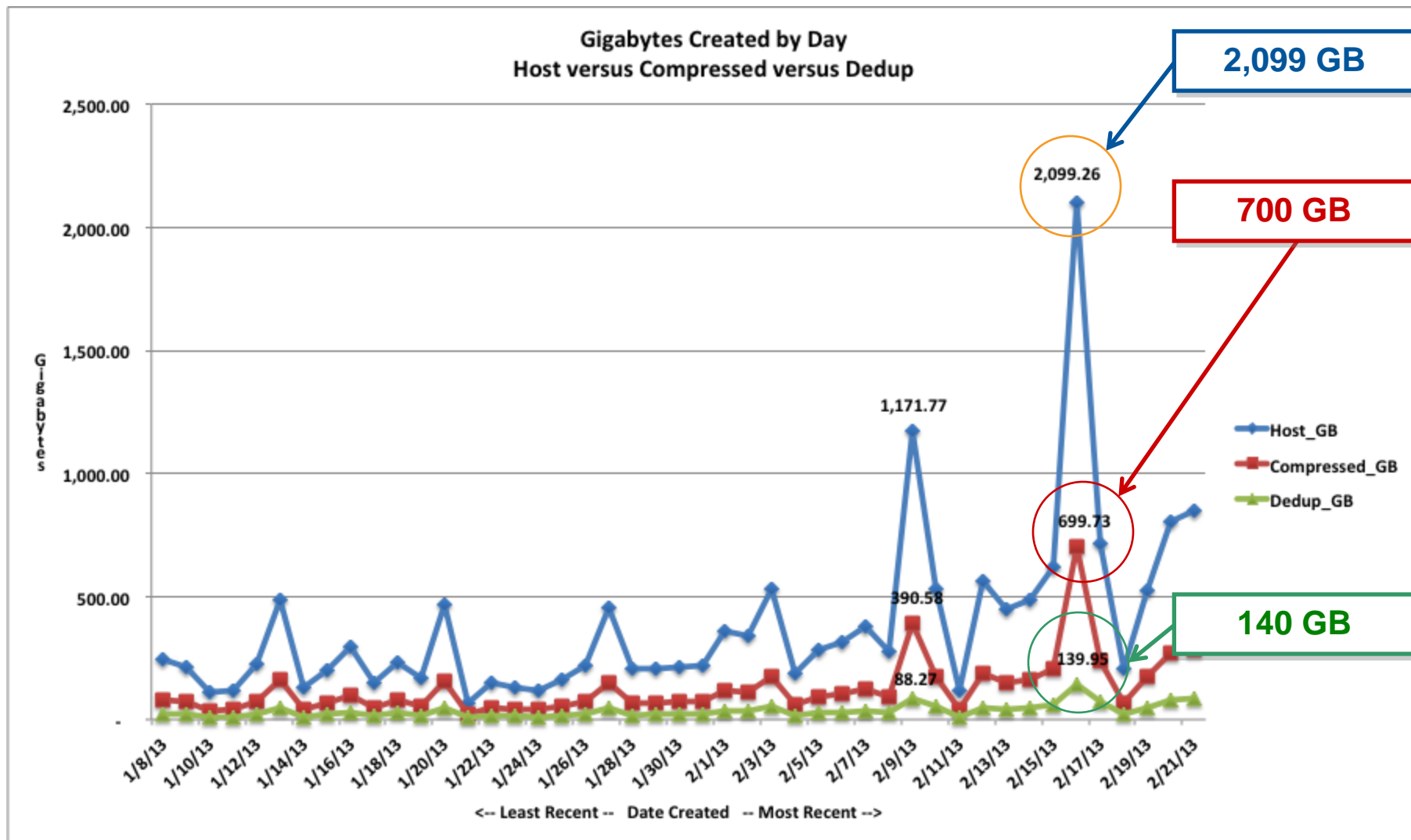
What Can Dedupe Do For You?
100 TB Tape Catalog Example

100 TB of Virtual Tape Data　　**1 TB of Daily Backup Data**　　**100 TB of Virtual Tape Data**

WAN

Primary Storage Capacity Requirement　　Replication Bandwidth Requirement　　DR Storage Capacity Requirement

Primary Storage		Replication Bandwidth		DR Storage	
100 TB	1:1	1000 GB　1:1		100 TB	1:1
33.3 TB	3:1	333 GB　3:1		33.3 TB	3:1
10 TB	**10:1**	**67 GB　15:1**		**10 TB**	**10:1**

- ■ No Compression
- ■ Compression
- ■ Deduplication

Note: Best Practice – Seed DR storage at the Primary Site before shipping

SHARE in Pittsburgh 2014

How Much Bandwidth is Required?

Example: Replicate 1 TB daily

WAN Class	MB/s	GB/hr	Replication Time w/ 3:1 compression	Replication Time w/ 15:1 deduplication
DS3	5.59	19.66	16.95 hours	3.39 hours
OC-1	6.48	22.78	14.63 hours	2.93 hours
OC-3	19.44	68.34	4.88 hours	0.98 hours

If replication begins at midnight, a DS3 line will complete replication long before a courier would show up to begin transporting a physical tape.

ProtecTIER Mainframe Edition

Luminex CGX with ProtecTIER

ProtecTIER ME

- **The solution makes it practical to replicate tape data over a WAN**

- **Deduplication Tuned for the Mainframe**
  Offering performance for both reading and writing mainframe tape data

- **High Deduplication Ratios**
  Coupling Luminex's DataStream Intelligence with HyperFactor allows customers to achieve much higher data deduplication rates for all mainframe data (not just full volume dumps…)

- **Open and Mainframe Synergy**
  Offers a common virtual tape storage target for both mainframes and open systems

- **Floor Space Savings**
  Utilizes existing customer racks for minimal floor space utilization.

IBM ProtecTIER Mainframe Edition

Enterprise class deduplication powered by Luminex

Specifications per cluster	ProtecTIER ME
Number of Virtual Devices	Up to 8192
Usable Physical Cache	Up to 1 PB
Effective Cache Capacity (15:1)	up to 15 PB
Number of Virtual Volumes	unlimited
FICON channels	1 or 4
Logical paths per FICON channel	256
8 GB FICON support	Yes

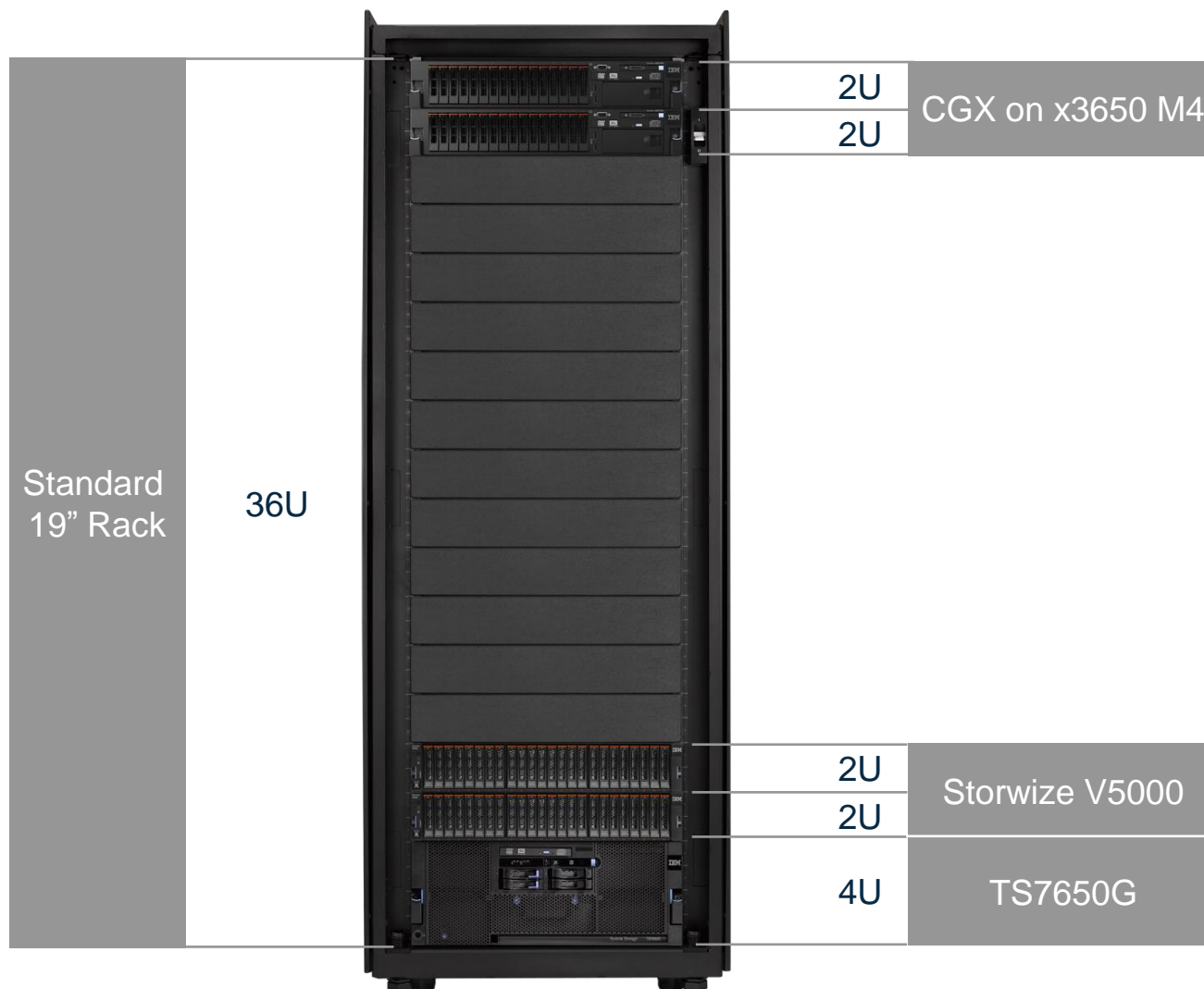# CGX with IBM System Storage ProtecTIER TS7620 Appliance

- In-Line Data Deduplication

- Automated replication capabilities

- TS7620 Appliance can provide 35TB of physical storage capacity per system

  - Up to 525TB of backup capacity with 15:1 deduplication

CGX with IBM System Storage ProtecTIER TS7650G

- In-Line Data Deduplication
- Replication Options:
  - One-to-One, One-to-Many and Many-to-Many
- TS7650G can provide up to 1PB of physical storage capacity per system
  - Up to 15PB of backup capacity with 15:1 deduplication
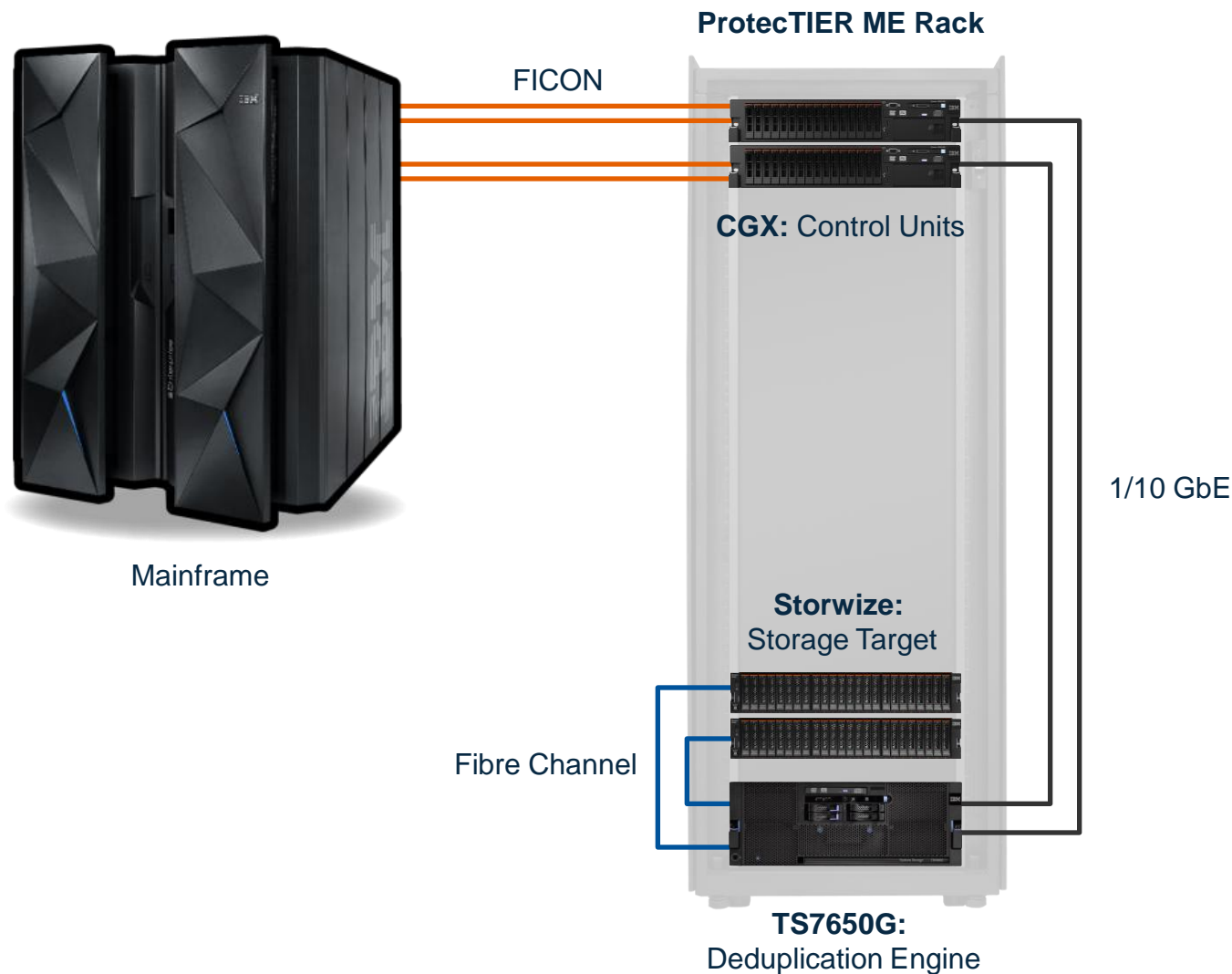
IBM ProtecTIER ME Rack Space
Smaller Footprint for Floor Space Savings

Standard 19" Rack

36U

2U	CGX on x3650 M4
2U	
2U	Storwize V5000
2U	
4U	TS7650G

IBM ProtecTIER ME Configurations

Multiple CGXs for Redundancy

ProtecTIER ME Rack

FICON

CGX: Control Units

1/10 GbE

Mainframe

Storwize:
Storage Target

Fibre Channel

TS7650G:
Deduplication Engine

IBM ProtecTIER ME Configurations

For Increased Capacity & Throughput

ProtecTIER ME Rack

ProtecTIER ME Rack

GROWTH OPTIONS

FICON

FICON

Two Dimensional Growth Options

- Expandable capacity
- Performance Enhancements

1/10 GbE

1/10 GbE

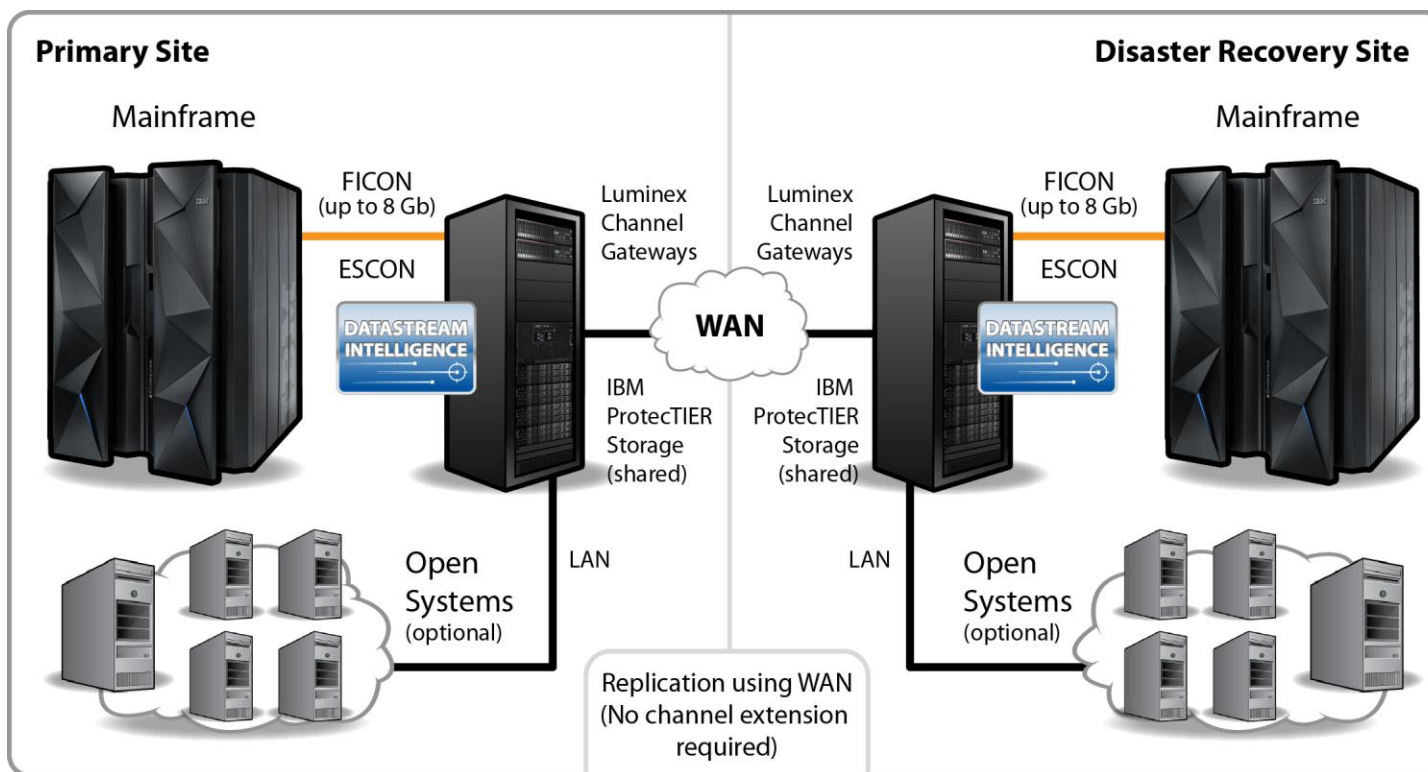FC

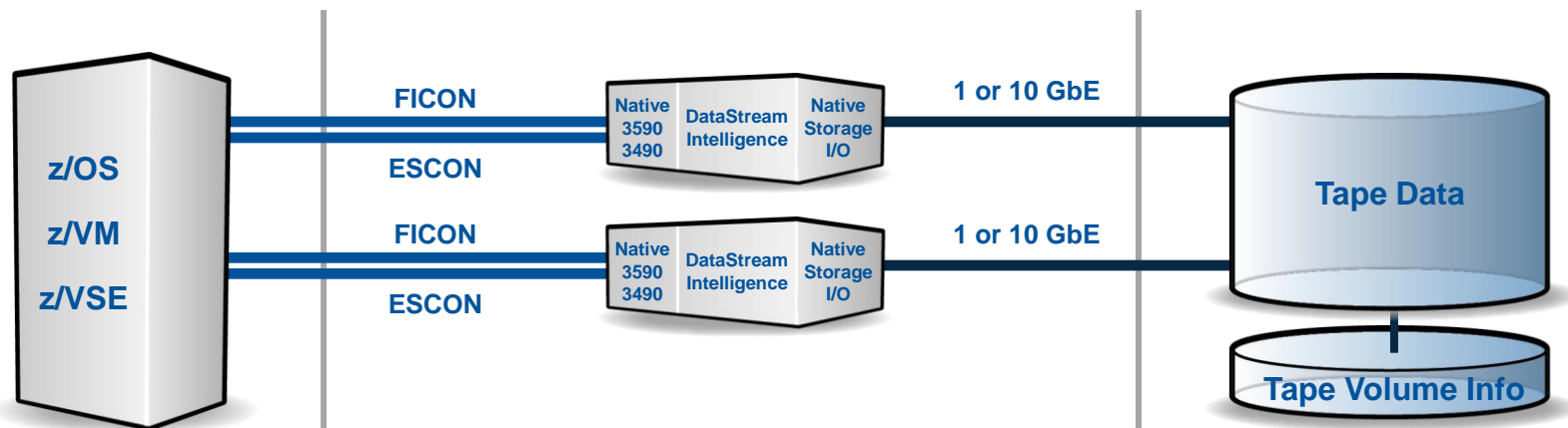FC
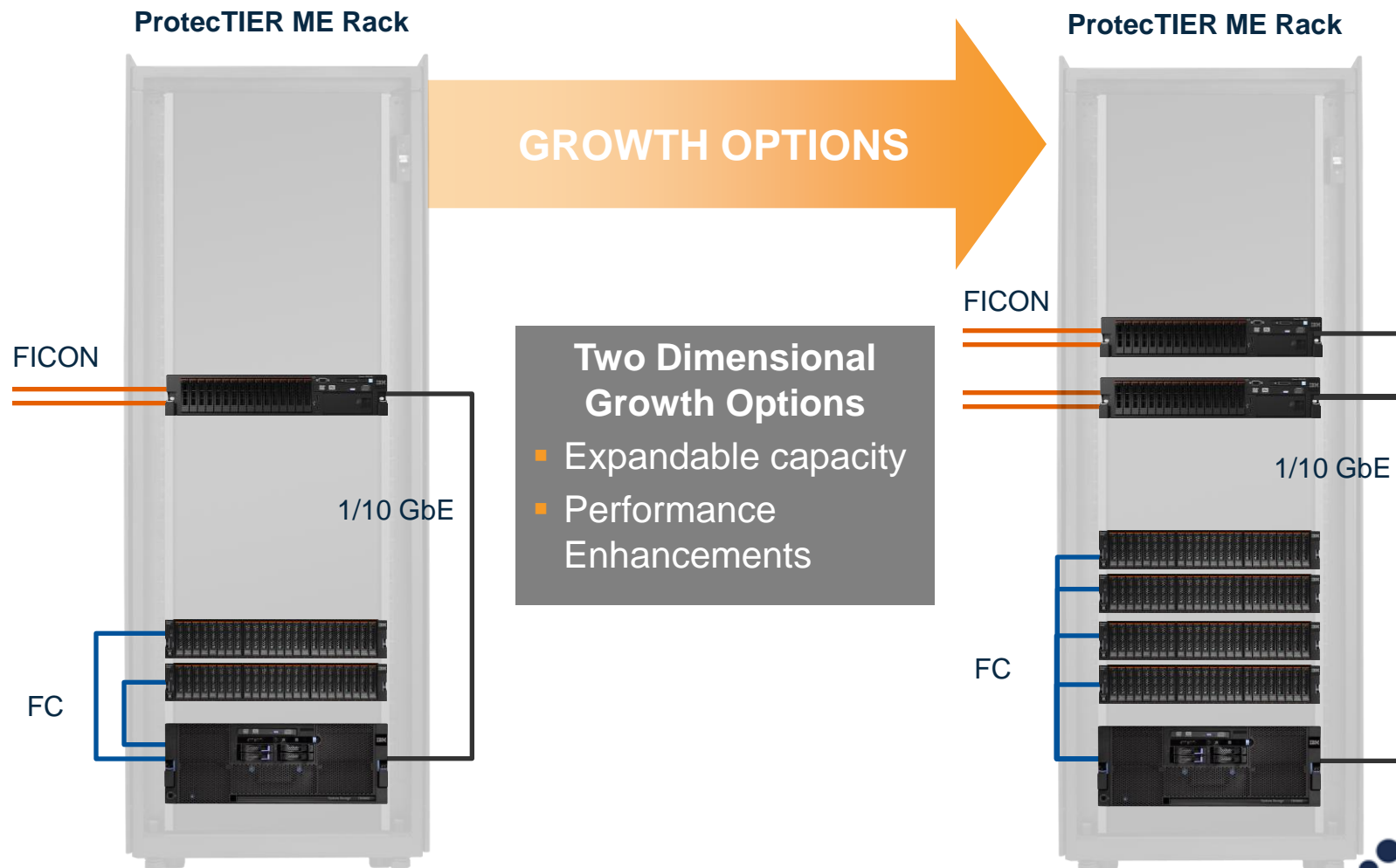
ProtecTIER ME – Production and DR Site Example

Shared Infrastructure
Common open and mainframe backup and tape processing solution

How the Solution Works
IBM ProtecTIER ME

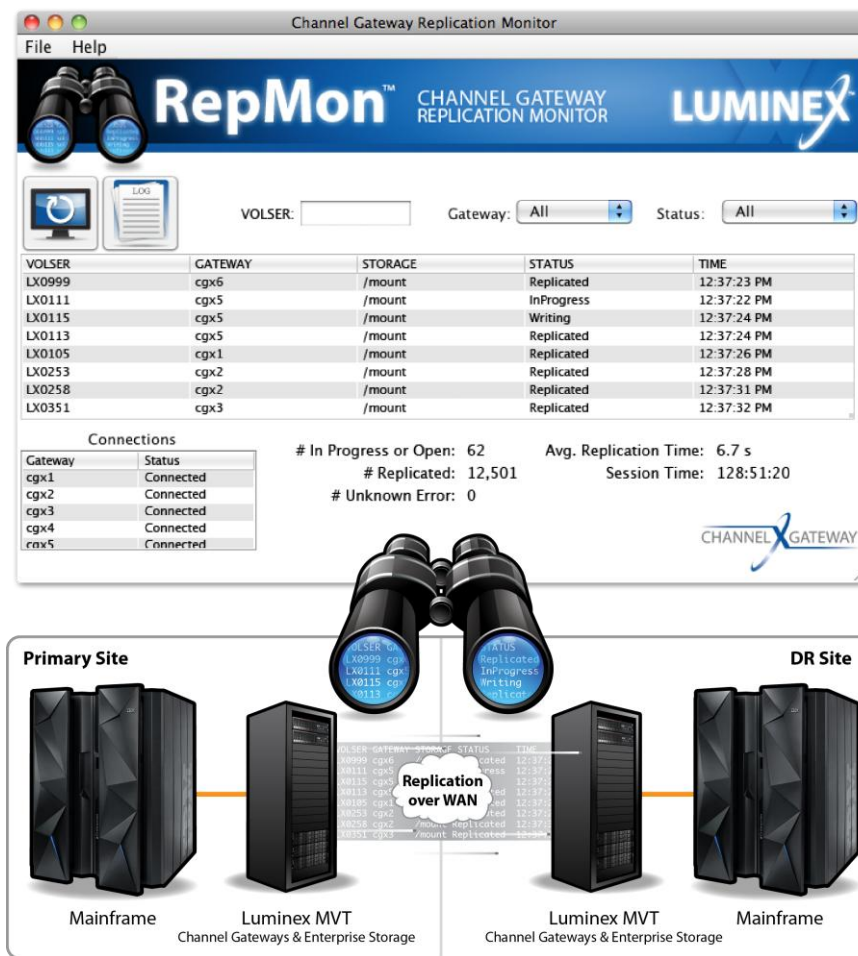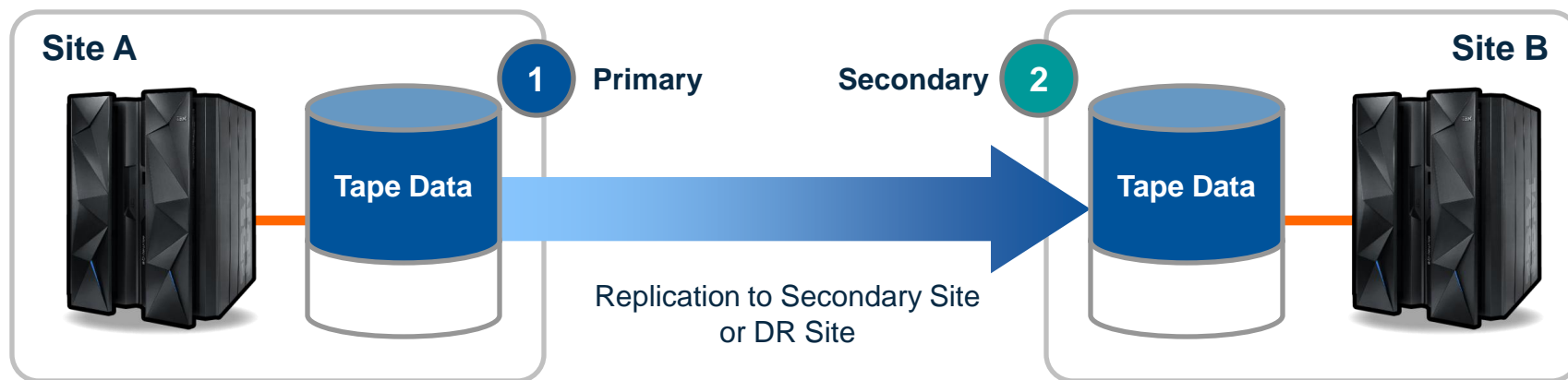| Mainframe	Luminex Channel Gateways	IBM ProtecTIER Storage
- Application transparent – non-intrusive - No MIPS required - z/OS, z/VM and z/VSE supported - Works well with all major tape management systems - SMS via MTL or Esoterics can be used - Optional host-based software for management	- Emulates 3490 or 3590 mainframe tape drives - "Wire Speed" up to 8 Gb FICON - DataStream Intelligence™ optimizes deduplication - No dependencies on separate code to modify a backup VOLSER - Active – Active with NSPOF - Modular design makes scaling throughput and/or capacity easy and cost effective	- Network Attached Storage (using NFS over 1 or 10 GbE) - Mainframe tape volumes stored as standard files - Replication for backup/DR - RAID Data Protection

Replication Monitoring

Provides **real-time status monitoring and logging** of virtual tape data writes and replication to a remote disaster recovery site at the VOLSER level

- Identifies Write and Replication Status of Mainframe Tape VOLSERs

- Identifies if virtual tape data at DR is still consistent with the primary datacenter

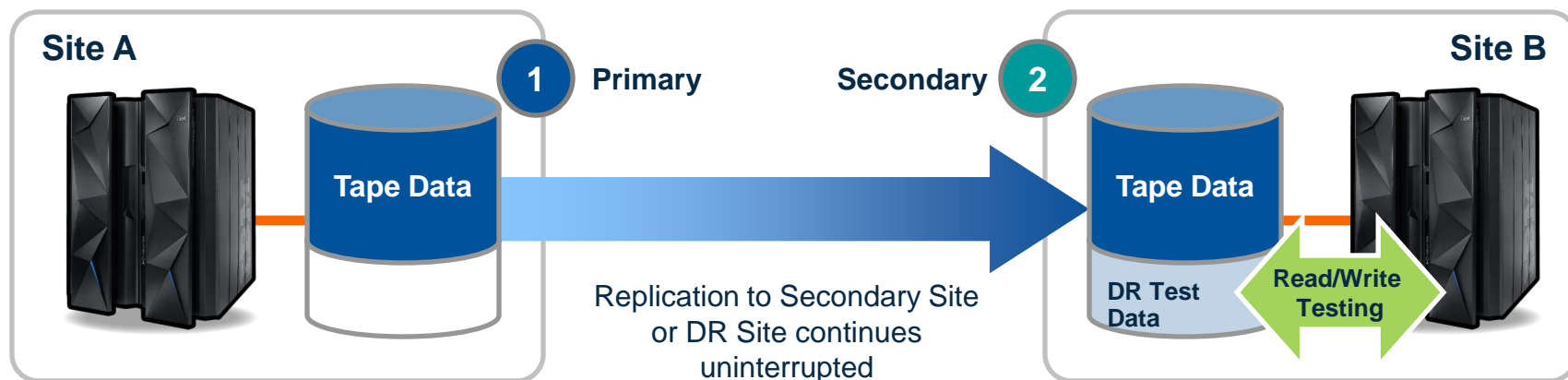- Provides visual and audit capabilities to confirm when backups reach DR

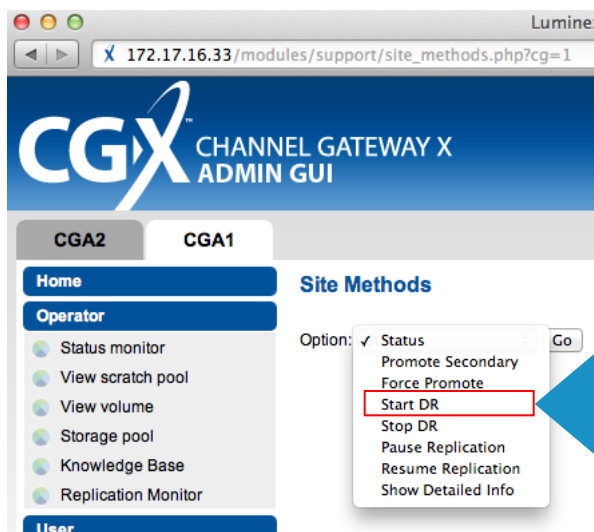# Simplified DR Testing
Replication During Normal Operations

Site A

1 Primary

Secondary **2**

Tape Data

Tape Data

Site B

Replication to Secondary Site
or DR Site

Simplified DR Testing
Replication During DR Testing

Site A

(1) Primary   **Secondary (2)**

Site B

Tape Data

Tape Data

DR Test Data

Read/Write Testing

Replication to Secondary Site or DR Site continues uninterrupted

Prepares DR environment for read/write testing; original Tape Data remains untouched

CGX CHANNEL GATEWAY X ADMIN GUI

X 172.17.16.33/modules/support/site_methods.php?cg=1

CGA2   CGA1

Home

Operator

Status monitor

View scratch pool

View volume

Storage pool

Knowledge Base

Replication Monitor

User

Site Methods

Option: ✓ Status   Go
Promote Secondary
Force Promote
Start DR
Stop DR
Pause Replication
Resume Replication
Show Detailed Info

Start DR

Simplified DR Testing
After DR Testing is Completed

Site A

① Primary   Secondary **②**

Site B

Tape Data

Tape Data

Replication to Secondary Site
or DR Site continues
uninterrupted
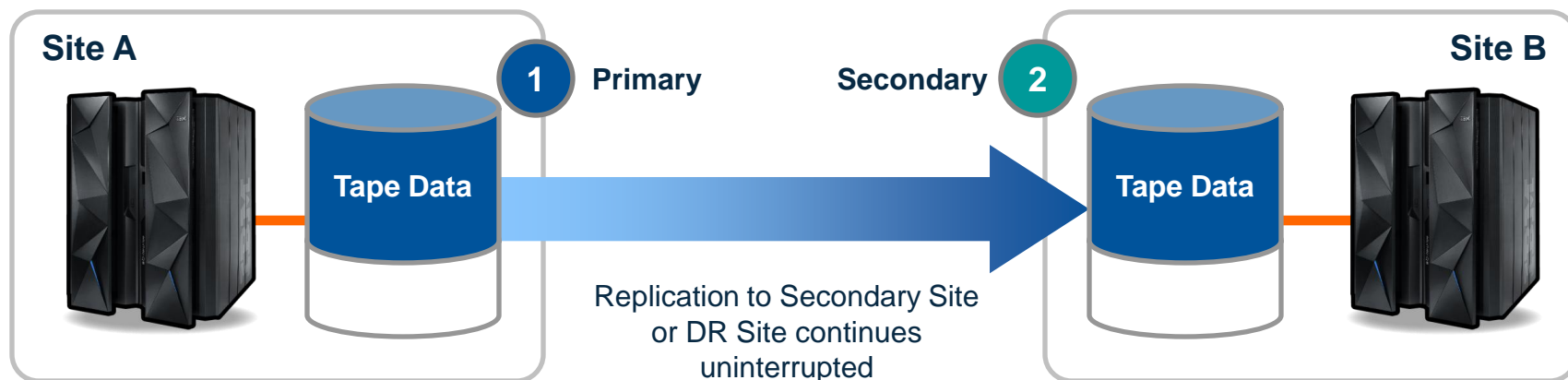
DR Test Data is purged

Optionally, DR Test Data can be
automatically replicated back to Site
A for auditing purposes

Luminex

X 172.17.16.33/modules/support/site_methods.php?cg=1

CGX CHANNEL GATEWAY X
ADMIN GUI

CGA2   **CGA1**

Home

Operator

- Status monitor
- View scratch pool
- View volume
- Storage pool
- Knowledge Base
- Replication Monitor

User

Site Methods

Option: ✓ Status   Go
        Promote Secondary
        Force Promote
        Start DR
        Stop DR
        Pause Replication
        Resume Replication
        Show Detailed Info

Stop DR

Tape Migration Services and Software

- **Luminex offers Tape Migration Services to migrate to ProtecTIER ME**

  – Elegantly designed to work with TMACS to move tape data without touching the tape catalogs

  – Current VOLSER #s and all historical information are retained in the new environment as well

  – Supports all existing tape library and virtual tape environments for z/OS

- **TMACS (Tape Monitoring and Allocation Control Software)** is optional host-based software to automate device allocation steering for complex environments

- **For other Luminex virtual tape environments**

  – Luminex offers Media Migration to migrate non-disruptively to ProtecTIER

Media Migration Services & Software

For other Luminex virtual tape environments

- Luminex offers Media Migration to non-disruptively migrate to the new storage target

- Entirely off-host, no mainframe MIPS required

- Current VOLSER #s and all historical information are retained in the new environment (no changes to tape catalogs)

- Volumes will acquire the characteristics of the new configuration
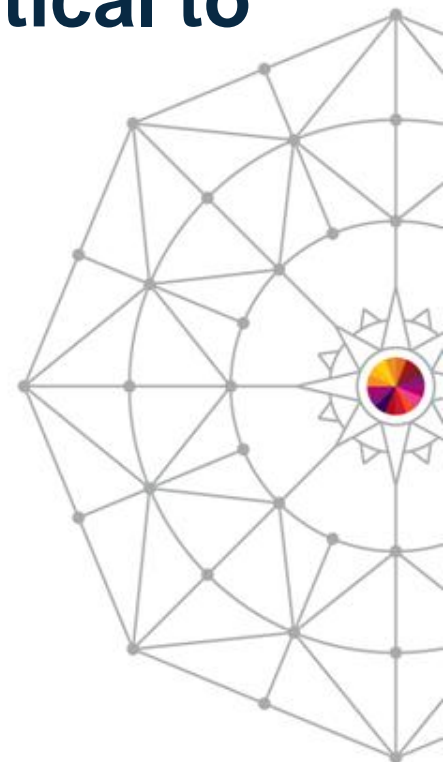
What Can ProtecTIER ME Do For You?

1. Increase FICON throughput and reduce tape storage requirements
2. Offer better read performance for HSM, batch and other tape applications
3. Reduce data center floor space requirements
4. Further reduce or eliminate the cost of maintenance, media, shipping and off site storage for physical tape
5. Improve tape data security
6. Simplify Disaster Recovery via Push Button DR
7. Offer Replication Monitoring at the volser level
8. The solution can make it practical to replicate tape data over a WAN
9. Improve disaster recovery preparedness
   - Improve RPO and RTO
10. Improve all aspects of tape operations
    - Batch processing, HSM, Archiving, Backup & Recovery

Data Deduplication Makes It Practical to Replicate Your Tape Data for Disaster Recovery

Scott James
VP Global Alliances
Luminex Software, Inc.

Randy Fleenor
Worldwide Data Protection Management
IBM Corporation

SHARE in Pittsburgh 2014