IBM Storwize Family Replication
Agenda

- The Value of Storage Virtualization

- Replication Overview
  - Metro Mirror
  - Global Mirror
  - Global Mirror with Change Volumes

- Replication Topology Options/Limits
  - Linear
  - Hub
  - Cross Connect

- Replication Connectivity Options/Limits
  - Fibre Channel
  - IP with integrated SANSlide

- Replication Implementation Considerations
Setting the Data Agenda

Enabling Strategic Initiatives & Superior Data Economics through a Proven Data Architecture…

Strategic Initiatives

Data Architecture

Simplicity: Effort-Free Delivery of Data to its Use

Software - Value
(Workloads, Virtualization, Grid Scale, …)

Independence
Freeing Data from Devices

Orchestration
Data Mobility

Optimization
Efficiency and Protection

Devices – Commodity
(HDD, Flash, Tape…)

Data Resources
Primary Data…
Copy Data…
Archive Data…
Multi-vendor…
Data Virtualization… For All Data Centers
Setting the Data Agenda Provides Better Economics

Push button data mobility, 50% reduction in TCO

100% of data is virtualized, easing data center migrations, data reduction and protection

Continuous migrations with zero disruptions, 67% reduction in storage and facilities cost

Minutes not hours for provisioning, 70%+ reduction in data
Radically Optimize Your Data Architecture with Ease

Automated Data Migration
Technology Migration, Future Proofing...

Data Optimization
Manage data as a single pool, Common efficiency features

Continuous Operations
New Apps, Capacity Growth, Data Availability...

Workloads
(e.g., SAP, VMware, Database, Windows)

Real-Time Compression

Data Replication & Protection

Flash
Non-Disruptive Migration

Data Virtualization

Data Replication & Protection

Diverse Storage

Automated Data Migration
Think Data
Set The Data AGENDA
Replication Overview
Basics – Single Platform for Remote Copy Services

Single point for copy services

- Asynchronous remote copy/Global Mirror
- Synchronous remote copy/Metro Mirror
- Compatible between all members of the Storwize family
- Fibre Channel and/or IP Replication
- Mirroring types can be used in any combination

Use to meet business needs

- Disaster recovery up to 8000KM
- Business Continuance up to 300KM
- Works with native and externally virtualized storage
- Network Protocol Flexibility
- Environment Flexibility
The IBM Storwize Family

A comprehensive range of fully virtualized storage systems

- One code base on all platforms
- One set of functions (selectively licensed)
- One user interface
- One CLI
- Replication can occur between any systems in the family

Storwize V3700
Storwize V5000
Storwize V7000
Flex System V7000
SAN Volume Controller
Storwize V7000 Unified

© 2014 IBM Corporation
SVC Metro Mirror Function

- “Metropolitan” distance synchronous remote mirroring function
- Up to 300km between sites for business continuity
  - As with any synchronous remote replication, performance requirements may limit usable distance
- Host I/O completed only when data stored at both locations
  - Unless write rates are low, not a good candidate for IP based replication
  - Important to keep subsystem performance similar
- Designed to maintain fully synchronized copies at both sites
  - Once initial copy has completed
  - During resync data is not consistent; use a point-in-time copy
- All mirroring types/transport delivered as single feature
  - Offers great implementation flexibility
- Operates between members of the Storwize Family at each site
  - Local and remote volumes may be on any SVC supported disk systems
SVC Global Mirror Function

- Long distance asynchronous remote mirroring function
  - Up to 8000km distance between sites for business continuity
- Does not wait for secondary I/O before completing host I/O
  - Helps reduce performance impact to applications
- Designed to maintain consistent secondary copy at all times
  - Once initial copy has completed
  - Copies are not consistent during a re-sync
    - Use a FlashCopy for a good point-in-time
- Built on Metro Mirror code base
- All remote copy functions delivered as single feature
  - Offers great implementation flexibility
- Operates between Storwize members at each site
  - Local and remote volumes may be on any Storwize Family supported disk systems
Lower Bandwidth Remote Mirroring – Global Mirror w/Change Vols

- Potential Problem: SVC Global Mirror delivers very high data currency at remote site but requires network bandwidth to support this currency
  - No mechanism to control RPO
- New Global Mirror with Change Volumes (GMCV) option makes regular copies of data at primary location and transmits them to secondary location
  - RPO and network bandwidth consumption can be regulated
- User can select frequency of these copies, which determines network bandwidth required and data currency
- Provides much more flexibility to meet specific application currency and network cost requirements
- GMCV is limited to 256 relationships per system
  - With RPQ and 7.x code this can be lifted to 2048 relationships per system
Global Mirror with Change Volumes

- Initially copy all data from primary volume to secondary volume at point-in-time of when GM relationship started
- When copy completes a point-in-time copy of the secondary is taken to ensure data consistency
- Cycle is started again – Either immediately or when cycle timer trips
- 80ms roundtrip latency maximum (for Fibre Channel or 1 Gb Ethernet, 10ms for 10Gb)
- Bandwidth sized based on RPO desired (minutes to hours)
- No FlashCopy license is required
Cycling Period

- The user configures the cycling period
  - This is the minimum time between cycle starts
  - Default is 300 seconds (5 minutes)
  - If the link is slow or I/O throughput is high, it may take longer than this for the new data to be transferred to the secondary
    - If so then the next copy will start as soon as the last one has finished
  - RPO is up to 2X the cycling period
  - Actual RPO is determined by the Freeze Time listed on the relationship
Global Mirror Cycling Mode Options

- Options for `-cyclingmode` are `none` | `multi`
  - Specifies the behavior of Global Mirror for this relationship
- Specifying the default "`none`" gives identical behavior to regular Global Mirror even though you may have change volumes attached to the relationship
- Specifying "`multi`" uses the cycling protocol
- To start a relationship with cycling mode set to "`multi`" there must be Change volumes defined for the relationship
  - Admin can create Change volumes or let system create them via GUI when configuring GM with Change volumes option
  - Note that the cycling mode can only be changed when the relationship is stopped and in a stopped state
Replication Topology Options
Remote Copy Partnership Topologies

- Up to 4 Storwize family members in direct or indirect partnerships
- Topologies are related to remote copy “partnerships”
  - This differs from remote copy relationships (a volume and its mirrored partner)
- Only 1 IP partnership per Storwize family member is supported
  - The rest must be made up of Fibre Channel relationships.

- Topology Types
  - Quadrature or Triangle
  - Line
  - Star
Remote Copy Partnership Topologies

- Quadrature (or triangle) partnerships’ topology

- Since any system can only have 1 IP partnership only 2 of these partnerships could be IP
  - The rest have to be Fibre Channel
Remote Copy Partnership Topologies

- Line partnerships’ topology

- Since any system can only have 1 IP partnership only 2 of these partnerships could be IP
- The rest have to be Fibre Channel
Remote Copy Partnership Topologies

- Star partnerships’ topology

- Since any system can only have 1 IP partnership only 1 of these partnerships could be IP
- The rest have to be Fibre Channel
Replication Connectivity Options/Limits
Fibre Channel Communication Options

- Native Fibre Channel - Including FCoE
- Dark Fiber
- FCIP – Either Brocade or Cisco
- DWDM/CWDM

- Note: Storwize members have no knowledge of WAN

Max 80ms roundtrip latency
Native IP Remote Copy

- Enables use of **1Gbit or 10Gbit** Ethernet connections using TCPIP without the need for fibre channel or FCIP routers for replication
  - Supports all remote copy modes – MM and GM
    - GM with Change Volumes preferred mode
  - Covered by normal remote copy license
    - It is not a new replication offering, but rather a new transport versus using a fibre channel network

- **Configuration:**
  - Automatic path/session configuration via discovery of remote cluster using IP management addresses of both systems in the IP remote copy partnership
  - Configure IP partnership using either **1Gbit or 10Gbit** Ethernet ports but not both at same time
  - Currently cannot switch speeds anywhere in end-to-end link other then WAN
    - E.g. cannot connect 10Gbit port to 1Gbit switch and then connect 1Gbit out to WAN
    - E.g All 10Gbit->WAN<-10Gbit or 1Gbit->WAN<-1Gbit
  - CHAP-based authentication supported

- Includes Bridgeworks SANSlide IP network optimization technology
With TCP/IP, information transfer slows the further you go.

This is because of the latency caused by waiting for acknowledgment of each set of packets sent since the next packet set cannot be delivered until the previous one has been acknowledged by the remote system.
Bridgeworks SANSlide optimization (3/5)

Latency in standard IP network (1Gb link speed):
  - depends on the routing infrastructure
  - depends on distance

Latency = Round Trip Time (RTT) for a single packet set

Source: IBM Redpaper REDP5023
Enhanced parallelism by using multiple virtual connections (VC) that share the same IP links and addresses:
- When waiting for one VC’s ACK, it sends more packets across other VCs
- If packets are lost from any VC, data will be retransmitted
- Artificial Intelligence engine adjusts number of VCs, receive window size, and packet size as appropriate to maintain optimum utilization/performance of the links
Latency in IP network (1Gb link speed):
  depends on the routing infrastructure
  depends on distance

Latency = Round Trip Time (RTT) for a single packet set
Native IP Remote Copy – IP Partnership Requirements

- If you use IP replication, you must meet the following requirements:
  - TCP ports 3260 and 3265 are used by systems for IP partnership communications
    - These ports must be open on any firewalls
    - Port 3260 is the port used by the systems to initially discover each other
    - Port 3265 is for the actual IP Replication sessions used to transmit data
  - The maximum supported round-trip latency between sites is 80 ms for a 1 Gbps link
  - The maximum supported round-trip latency between sites is 10 ms for a 10 Gbps link
  - The recommended method of replication is Global Mirror with Change Volumes
  - The inter-cluster heartbeat traffic consumes 1 Mbps per link
  - The minimum bandwidth requirement for an inter-cluster link is 10 Mbps

- Maximum throughput is currently restricted based on use of 1gigE or 10gigE ports
  - One 1gigE port could do up to 111 MB/s unidirectional / 190 MB/s bidirectional
  - Two 1gigE ports could do 220 MB/s unidirectional / 325 MB/s bidirectional
  - One 10gigE ports could do up to 190 MB/s unidirectional (bidirectional not tested)
  - Two 10gigE ports could do up to 305 MB/s unidirectional / 325 MB/s bidirectional
    - V7.2.0.2 is supposed to give some moderate increases to speed
Native IP Remote Copy – Configuring Ethernet Ports
Bandwidth setting definition has changed in code version 7.2:
- Previously this bandwidth setting (called Partnership Bandwidth) default was 50 MBytes and meant that was the maximum transfer rate from primary to secondary site for initial sync/resyncs of volumes
- “Link bandwidth” setting is now configured using Mbits not MBytes and you set this to what the communication link actually can sustain or what is actually allocated for replication
- “Background copy rate” setting is now a percentage of the “link bandwidth” and determines the bandwidth available for for initial sync and resyncs or for GM w/chgvols
**Native IP Remote Copy – Partnership Properties**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>ATS_OXFORD6</td>
</tr>
<tr>
<td>Location</td>
<td>Remote</td>
</tr>
<tr>
<td>State</td>
<td>✔ Fully Configured</td>
</tr>
<tr>
<td>Type</td>
<td>IPv4</td>
</tr>
<tr>
<td>Partner system IP address</td>
<td>192.168.21.7</td>
</tr>
<tr>
<td>Link bandwidth</td>
<td>1500 MBps</td>
</tr>
<tr>
<td>Background copy rate</td>
<td>50 %</td>
</tr>
<tr>
<td>Mutual CHAP authentication</td>
<td>Disabled</td>
</tr>
<tr>
<td>CHAP secret</td>
<td></td>
</tr>
</tbody>
</table>

![Partnership Properties Window](image)
Remote copy port group = Set of local and remote Ethernet ports and associated TCP/IP addresses available to establish a session over the IP link(s)
- Path will be established between IP addresses (local + remote) when IP partnership first configured
- Minimum: 1 local + 1 remote port in the same remote copy port group
- Recommended to cover node failures: For single I/O Group
  - 2 local + 2 remote ports in the same remote copy port group
- Recommended to cover node failures: For two or more I/O Groups
  - 2 local + 2 remote ports from each of 2 I/O Groups maximum in the same remote copy port group
Either Ethernet port 1 or 2 on a node/node canister can be configured for IP replication but not both.

The remote copy port group on each system includes two IP address, one from each node, and when initially configured the system will establish which one from each site is used.

- This pairing can not be chosen or changed by an administrator.
- In this example H1 and M2 have established a session.

If H1 node fails then the session between H1 and M2 fails and the system automatically establishes another session between H2 and either M1 or M2.

- The IP address used for replication on H1 does not failover to H2.
The remote copy port group on each system includes four IP address, one from each node, and when initially configured the system will establish which one from each site is used.

If H1 node fails then the session between H1 and M2 fails and the system automatically establishes another session between H2, H3 or H4 and M1, M2, M3 or M4.
One link – More than two I/O Groups

NOTE: Nodes H5/6 and M5/6 do not have IP addresses configured as the limit as only the Ethernet ports of 2 I/O Groups can be configured with IP addresses and included in remote copy port groups.
Remote copy port group setup for dual redundant links:
- Port group 1: two IP addresses, each on a different node in different system (green)
- Port group 2: two IP addresses, each on a different node in different system (orange)

- Maximum of 2 simultaneous IP replication sessions at any time
  - Only one active per node, so in this configuration if node A1 would fail then Link 1 would no longer be available and only link 2 would be active
    - Result is loss of half of the bandwidth available for replication
    - If bandwidth sizing dictates dual links required and if one link fails resulting in MM or regular GM not able to operate properly, recommendation is to have 2 I/O Groups at each site to allow the failed link to restart on other nodes available in systems
Two Links – One I/O Group and one port per node

- Again maximum of 2 simultaneous IP replication sessions at any time
  - Only one active per node so in this example if node M2 would fail then Link 1 would no longer be available and only link 2 would be active
    - Result is loss of half of the bandwidth available for replication
Each remote copy port group on each system includes two IP address, one from I/O Group in this example, and when initially configured the system will establish the pairings used.

If H1 node fails then the session between H1 and M2 fails and the system automatically establishes another session between H3 and either M2 or M4 since they are all in the same remote copy port group with H1.
Native IP Remote Copy – Miscellaneous Information

- Cannot today non-disruptively switch from a FC partnership to an IP partnership
  - Must delete relationships and delete partnership and start over basically
  - If currently replicating over FC and can take outage on hosts replicating volumes, we have a method using –sync flag on “mkrcrelationship”

- SANSlide – What it does and doesn’t do:
  - It optimizes link utilization, but not latency
  - It uses standard TCP/IP – no UDP
  - It doesn’t do any encryption, compression or de-duplication

- Only two systems can be in one IP Replication partnership
  - Can do A->B partnership
  - Cannot do A->B->C->D or A/B/C->D

- A system can be in an IP partnership with one other system and in FC partnership with others
  - Still a maximum of 4 systems in partnership with each other

- FC replication and IP replication can be used with any combination of SVC and Storwize systems all running V7.2
Native IP Remote Copy – Best Practices

- Configure two physical links between sites for redundancy and to provide greater bandwidth, if needed, between the two systems
  - A minimum of 4 nodes at each site is required for dual redundant IP links to work across a node failure at either site today
  - With a 2 node system at each site, in the event of a node failure, one link will be lost
- Dedicate Ethernet ports for replication if possible
  - Recommendation is to not use a port for both replication and iSCSI host access
  - Ok to use 1gigE port for both replication and system management
- Multi-protocol partnerships are not supported
  - Cannot configure an additional FC/FCoE partnership between two systems if an IP partnership already exists between them
- Configure CHAP secret based authentication for added security
- A whitepaper is in the works on how to step-by-step configure IP replication
Replication Implementation Considerations
Port Usage

- Although not required, it is highly recommended to dedicate ports to the remote copy process
  - Dedicate SAN ports for FC mirroring
  - Dedicate IP ports for IP replication
    - ISCSI and IP replication not recommended on the same ports

- Port dedication can help with the following problems:
  - SAN buffer to buffer credit issue for FC replication
    - If ports are shared, this can cause production traffic slowdown
  - 1920 errors due to insufficient WAN bandwidth
  - SVC Node to Node communication issues

- On SVC with code 7.x, dedicating ports will require that you use port masking to take node to node traffic off the dedicated ports
  - The local_fc_port_mask parameter can be set on a cluster to restrict the node to node traffic to specific ports
  - Other members of the Storwize family (V3700, V5000, V7000) do not need this set, because node to node traffic prefers the back-plane to the FC SAN.
SVC CG8 2nd 4 Port 8Gb HBA

- The SVC 2nd 4 port HBA was designed for environments where Fibre Channel ports need to be dedicated to certain aspects of the system.

- Available for all CG8 models (both 4 core or 6 core CPU):
  - Cannot be used with 10Gb Ethernet
  - Cannot be used with in-node SSD’s
  - Can be used with iRPQ 8S1296
    - Extra processor and memory for compression

- Currently, the extra HBA ports can be used for the following:
  - Metro/Global Mirroring
    - Dedicating ports helps with known buffer credit issues
  - Node to Node communication
    - Particularly relevant in a stretch cluster environment
  - Host Traffic
  - Storage traffic (as of 7.2 code)
Port Designations

1. Fibre-channel port 1
2. Fibre-channel port 2
3. Fibre-channel port 3
4. Fibre-channel port 4
5. Power supply 1
6. Power supply 2
7. Serial port (RS232 connection to UPS)
8. Ethernet port 2
9. Ethernet port 1
10. Fibre-channel port 5
11. Fibre-channel port 6
12. Fibre-channel port 7
13. Fibre-channel port 8

Port 1 Adapter 1 – 50050768014XXXXX
Port 2 Adapter 1 – 50050768013XXXXX
Port 3 Adapter 1 – 50050768011XXXXX
Port 4 Adapter 1 – 50050768012XXXXX
Port 5 Adapter 2 – 50050768015XXXXX
Port 6 Adapter 2 – 50050768016XXXXX
Port 7 Adapter 2 – 50050768017XXXXX
Port 8 Adapter 2 – 50050768018XXXXX
Which Type of Remote Copy Do I Use?

- **Use Metro Mirror when:**
  - You absolutely have to have a base RPO of 0
    - Base RPO - Meaning rolling logs forward does not count
  - Only when you have the bandwidth to sustain it
    - Production traffic will be slowed if you don't

- **Use Global Mirror when:**
  - You don't need base RPO of 0, but want a small base RPO (2 min or less)
  - You don't quite have the bandwidth for Metro Mirror
  - You don't want your production applications to be affected by somewhat slower replication
  - Use of 7.2 code can improve situations when bandwidth is not an issue

- **Use Global Mirror with Change Volumes when:**
  - You can tolerate a base RPO of 5 minutes or greater
  - Your production applications are being affected by slow downs from Global Mirror
  - Your WAN link is not sufficient to keep up with the peaks from Global Mirror
What Type of Partnership Do I Use?

- Use an IP partnership when:
  - FCIP equipment is not an option
  - Only when IP network is stable
    - Does not work well on IP networks that are not healthy
  - When bandwidth requirements do not exceed 200-300 MB/s
  - Only use 10Gb when latency does not exceed 10 ms

- Use a Fibre Channel partnership when
  - You have reached the maximum (1) IP partnerships per system and need another
  - High bandwidth is a requirement (greater than 200-300 MB/s)
  - You have a method of taking FC over the WAN (FCIP routers)
Thank You
Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.
Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

* AS/400®, e-business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, zVM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.