TANEJA GROUP
TECHNOLOGY ANALYSTS

SOLUTION PROFILE

IBM PROTECTIER REPLICATION FOR OUTSTANDING DATA PROTECTION

OCTOBER 2011

Traditional tape backup is pulling apart at the seams. Tape requires manual intervention to manage which leaves backup data wide open to human error. Transporting tape also exposes data to the risk of loss and off-site storage facilities are costly. Testing disaster recovery (DR) plans in this manual environment is an exercise in futility, and even when the process works restoring data from tape is painfully slow.

The basic answer is disk-based backup with replication. However, this very basic scenario is no miracle cure. Storing straight backup data on disk can quickly overwhelm disk storage resources, at which point IT sets up large-scale copy-to-tape – and ends up with the same tape problem they started with. Replication too can be very expensive in terms of bandwidth, and even a single set-and-forget unidirectional copy can be quite costly. It also lacks flexible replication features that today's backup environments require for their Tier 1 and even Tier 2 applications.

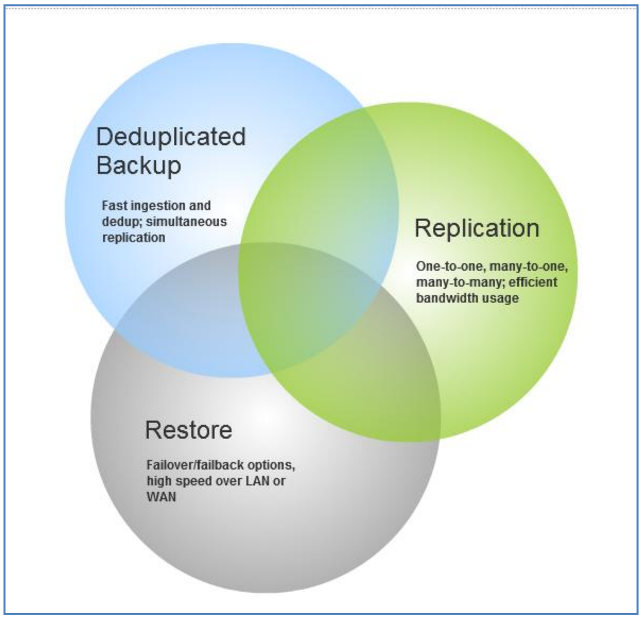
Deduplicated Backup
Fast ingestion and dedup; simultaneous replication
Replication
One-to-one, many-to-one, many-to-many; efficient bandwidth usage
Restore
Failover/failback options, high speed over LAN or WAN

Fig. 1: The Data Protection Trifecta

Modern backup and recovery needs much robust business continuity (BC) and DR than garden variety approaches can provide. Organizations require a trifecta of data protection operations: high-performance deduplicated backup; simultaneous replication that is bandwidth-efficient, flexible and offers 100% data integrity; and high-speed restore with failover and failback options.

This is why IBM ProtecTIER with advanced replication is a vital piece of backup and recovery. ProtecTIER deduplicates incoming backup streams for extremely efficient disk-based storage and simultaneously replicates data. ProtecTIER has offered native unidirectional replication for years but IBM has now developed far beyond that to include many-to-one and many-to-many replication options. Replicating only deduplicated, new and unique data makes it extremely bandwidth-efficient.

This Solution Profile makes the case for replication as a data protection cornerstone, positions replication in the context of customer DR environments, and explores how IBM ProtecTIER's powerful replication dramatically enhances BC and DR across mid-tier business and the enterprise.

Replication: Beyond the Basics

Replication is hardly new to data protection. Once backup administrators have set up a replication pair then they are inclined to do a set-and-forget. And indeed, one-to-one replication will serve simple environments and companies that are new to replication. In these cases one-to-one will work just fine *if* you only need basic replication to protect a single backup site, or *if* you don't mind setting up and managing multiple replication pairs for multiple backup sites or *if* you do not need to replicate mission-critical information to multiple secondary sites.

Because if these conditions are not true – if you do need to protect multiple backup sites; to consolidate replication management; to replicate to multiple secondary sites – then you need to go beyond basic one-to-one. For you, *the set-and-forget plan is playing with fire.* Here are some considerations that will help you decide what level of replication you actually need:

Replication Direction	Considerations
One-to-one	• You just have one backup site that needs to be replicated and do not plan on adding more • You have additional sites that need to be replicated but you don't mind managing separate replication pairs • DR strategy is not mission-critical and does not require multiple replication sites
Many-to-one	• You want to easily extend replication to more backup sites • You want to send replication from multiple backup sites to a single secondary site for ease in recovery • You want to centrally manage replication instead of managing separate pairs
Many-to-many	• You have mission-critical data that has to be protected on more than one secondary site • You need bidirectional replication among multiple sites with all the advantages of many-to-one • You want the option to centrally manage replication between multiple sites

WHAT REPLICATION SHOULD OFFER NOW

- **Many-to-one and many-to-many options.** Replication should offer dramatic improvements in data protection, data integrity and recovery times. Achieving this goal takes flexible replication architecture to optimize the backup infrastructure. Simple environments can use unidirectional replication but critical data and complex backup environments need more. Many-to-one replication copies data from multiple source repositories to a single destination repository. Many-to-many replication enables bi-directional replication between multiple systems. Both approaches allow IT to optimize backup data protection between data centers, DR sites and remote offices.

- **Serious cost savings.** Effective replication generates immediate cost savings and lowers risk by replacing physical tape transport and vaulting. The solution should also save on bandwidth costs which are the largest price tag associated with replication. The process should only copy deduplicated, new and unique data to achieve cost savings. Adopting this approach instead of copying uncompressed incremental backup over the WAN results in very significant bandwidth cost savings.

- **100% data integrity.** Replicated data must maintain 100% data integrity for correctly restoring lost or corrupted data. The replication operation should assure 100% data consistency at write time on the replication storage target. Asynchronous copy is best for this level of data verification when working with virtual cartridges.
- **Simplified management.** Managing complex replication operations requires policy-driven automation. Management tools that simplify policies and replication settings are enormously useful to backup administrators and a smooth replication process, and a central console should be a given.
- **Maximize recovery point objectives (RPO).** RPO governs the frequency of replication to a secondary site depending on an application's RPO service level agreement. The replication solution should serve RPOs that vary according to data criticality and needs. Administrators should be able to set varying replication schedules to suite different RPOs. (Shorter RPOs represent greater loads on the host and bandwidth, so settings must match the value of the data being replicated.)
- **Optimize recovery time objectives (RTO).** RTO is the length of acceptable time until a given application or data set is available to users. Electronically restoring replicated data sets is much faster than restoring from physical tape but IT still needs to assign replication and restore priority to different applications. Acceptable RTOs vary according to the business value of applications and data.
- **Failover and fast data recovery.** If a primary site is compromised then the solution should allow for a failover to the replication target systems and give the user the opportunity to immediately restore the replicated data sets for fast recovery. Failback to a restored primary system should be equally efficient with the secondary system restoring data from replicated virtual cartridges.

IBM ProtecTIER® and Replication

IBM ProtecTIER is a virtual tape system that replaces physical tape with deduplicated backup. ProtecTIER combines high performance deduplication with integrated IP-based replication for powerful data protection. ProtecTIER appliances and gateways simultaneously replicate deduplicated backup in a variety of options including one-to-one, many-to-one and many-to-one. ProtecTIER only copies deduplicated, new and unique metadata and customer data for dramatic bandwidth savings and fast restores.

More and more organizations are adopting disk-based deduplication and replication to protect their critical application data. ProtecTIER replicates deduplicated data between multiple sites including data centers, remote and branch offices, and offsite DR systems. This is a massive improvement over physical tape backup and long-term vaulting. By replacing physical cartridges with virtual cartridges, backup and recovery become faster, more efficient, less expensive, and more secure.

HOW IT WORKS

ProtecTIER's replication is native to the appliances and gateways. The initial set up of the replication grid is done through the ProtecTIER Replication Manager (RM) module. Replication may be scheduled using policies or may be manually launched. Normally ProtecTIER works on policy-driven schedules where it identifies data status and priority and places the data in priority order in the queue. ProtecTIER can simultaneously replicate up to 128 cartridges and each cartridge will have up to 32 concurrent data streams, which makes for very fast and efficient replication even on large amounts of data. ProtecTIER users may choose either scheduled replication that occurs at preset times or that runs concurrent to backup and deduplication. Retries are automatic up to seven days and users may clone virtual tape cartridges to physical tape libraries if desired. Asynchronous ProtecTIER replication verifies each virtual cartridge for 100% data integrity on the fly, only marking it complete after the data finishes replicating and is fully validated.

Bandwidth costs for remote replication can be very expensive. ProtecTIER offers sound bandwidth management so it can keep costs under control and data transfer rates high. ProtecTIER preserves bandwidth by only replicating deduplicated data that is new and unique. This approach returns a bandwidth reduction of 90% or more over uncompressed data transfer, which optimizes throughput and dramatically reduces the cost of backup replication. ProtecTIER also provides optional bandwidth management features that allows IT to support the maximum replication transfer rate allowed for a specific repository. Replication Rate Control for example enables users to set limits on replication performance to protect host throughput and bandwidth transfer rates.

The same bandwidth reduction also results in fast recovery rates. Following a disaster or corrupted data event, the source destination ProtecTIER can immediately begin restore operations to the primary ProtecTIER system. If the primary is unavailable then this secondary system can failback and internally restore the virtual cartridges as the new primary system. When the primary system comes back online, the secondary system fails back and fast data restoration begins. The systems then resume regular replication operations.

IT uses the ProtecTIER Manager and Replication Manager interface to maintain full access to the affected ProtecTIER systems during the disaster. Best practices strongly suggest that users install ProtecTIER Replication Manager at the secondary site for just this purpose.

MANY-TO-ONE REPLICATION

ProtecTIER many-to-one replication enables IT to replicate ProtecTIER data from remote sites to a centralized ProtecTIER destination. Many-to-one replication creates topology groups containing up to 12 source repositories (spokes) and a single destination repository (hub). Spokes backup data directly at their locations and may replicate some or their entire repository to the hub. The Hub can perform local backups and receive replication data from all or any of the spokes (ProtecTIER provides capacity monitoring tools and the user can enforce capacity limits so replication activity cannot use space allocated for the hub local backup data.) The hub acts as the disaster recovery site for any of the spokes. Spokes cannot be replication targets and the hub cannot replicate outside of failback. The spokes need not be physically connected with each other but spokes and hubs must be connected over the network to the Replication Manager node.

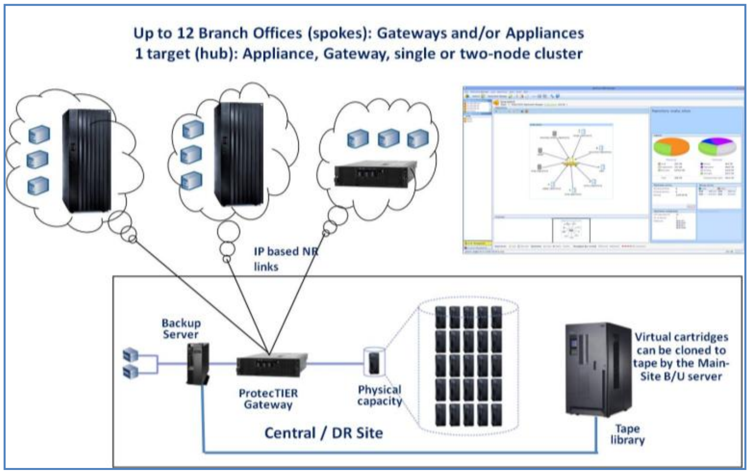
Up to 12 Branch Offices (spokes): Gateways and/or Appliances
1 target (hub): Appliance, Gateway, single or two-node cluster
IP based NR links
Backup Server
ProtecTIER Gateway
Physical capacity
Central / DR Site
Virtual cartridges can be cloned to tape by the Main-Site B/U server
Tape library

Fig. 2: ProtecTIER many-to-one replication

When a ProtecTIER customer upgrades to many-to-one from ProtecTIER's one-to-one replication, the RM will automatically upgrade the grid's database to set all replication pairs as hubs and spokes. ProtecTIER's many-to-one replication supports both single node and dual-node clustered configurations.

MANY-TO-MANY REPLICATION

ProtecTIER has introduced many-to-many bidirectional replication. Up to four ProtecTIER systems may participate in a many-to-many topology group with each system acting as both a source and hub (destination repository). Each hub can receive local backups directly and may also replicate to and receive replication from other hubs within the replication group. Users may choose to replicate different backups within the hub group. The groups are not restricted to pairs; each ProtecTIER system with the group may replicate to one another. For example, IT may replicate SAP DB2 data from backup target Hub #1 to Hub #2, replicate Exchange backup from Hub #2 to Hub #1, and replicate both hubs to the DR site's Hub #3, which is set to failover in the event of primary hub failure.

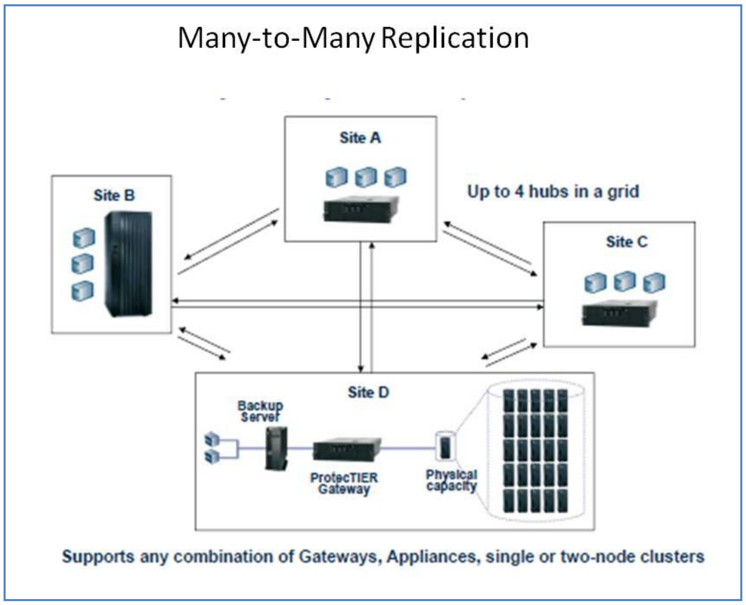

Fig. 3: ProtecTIER many-to-many replication

PROTECTIER RM AND REPLICATION GRIDS

The ProtecTIER Replication Manager coordinates ProtecTIER replication throughout the organization. ProtecTIER Manager allows users to manage and monitor replication grids, policies, cartridge replication and priorities, centralized replication schedules, cartridge visibility and virtual locations, and additional operations.

IBM ProtecTIER RM enables IT to manage multiple many-to-one and many-to-many configurations within centrally managed replication grids. RM software may be deployed on a dedicated host or on a ProtecTIER node. The RM manages all assigned ProtecTIER members and replication activity over individual replication subnets. The RM uses agents installed on each ProtecTIER node to interact with systems belonging to specific grids.

POLICY MANAGEMENT

ProtecTIER provides a Policy Manager that administers replication policies across replication groups and grids. Very flexible policies range from granular to global, and may be set at a number of different levels such as individual virtual cartridge level, grouped cartridges, topology groups, or full replication grids.

Examples include setting policies around failover operations, timeframes in support of service level agreements, replication rate limits, storage capacity, and bandwidth control for optimizing throughput.

A replication policy defines a set of objects within its range, such as virtual cartridges storing database data from the SAP ERP, for example. The policy directs action on these objects including replication priority, schedules and destination repositories. Once set the policy runs automatically according to its trigger event.

Sample Scenarios

CUSTOMER SCENARIO #1: MANY-TO-ONE REPLICATION

A mid-tier manufacturing company has a ProtecTIER gateway in their compact data center and deploys ProtecTIER appliances in three remote offices. The largest of the offices has a ProtecTIER cluster and the others have single node configurations. Three of the offices backup locally and replicate deduplicated backup to the data center. Another small office backs up remotely to the data center ProtecTIER, which maintains protected storage capacity for direct backup.

Before they purchased the ProtecTIER systems the company looked at some third party replication products that would have required them to upgrade their WAN connections. ProtecTIER's native replication and bandwidth reduction let them easily leverage existing connections.

IT set up many-to-one replication between three remote office spokes and a data center hub. If the smallest office needs to add replication then it can easily be added to the group. The customer also plans on adding ProtecTIER systems to several of their factories. The existing data center ProtecTIER hub can take up to 12 spokes, which will be adequate for some time. If the company grows its ProtecTIER deployments past that point they may create another replication grid while retaining centralized management capabilities.

CUSTOMER SCENARIO #2: MANY-TO-MANY REPLICATION

A major office products retailer has two regional data centers that act as secondary DR sites for one another. IT also wants to add backup replication from some of larger regional stores. The company owns a replication product that works with their backup application but bandwidth costs are scaling sharply up.

The customer keeps their backup application but replaces the replication product with ProtecTIER deduplication and replication systems at both data centers. The company starts with a pilot project that replicates from the stores to one of the data centers and then replicates Tier 1 data to the second data center. Once they realize that ProtecTIER has reduced bandwidth requirements by 85% they create a many-to-many replication group between the two data centers. Each ProtecTIER system replicates back and forth to each other.

They are also building a compact DR site where a third ProtecTIER system will serve as another repository within the many-to-many group.

Benefits of ProtecTIER Replication

ProtecTIER replication offers exceptionally strong benefits to customers who need to significantly improve data protection at a reasonable cost.

1. **High performance replication and recovery.** Fast and flexible replication protects data and enforces service level agreements. Replication happens concurrently with backup and inline deduplication, which decreases backup windows and meet service level agreements. Data recovers from high speed disk across fast bandwidth connections, which enables ProtecTIER's enviable recovery speeds.

2. **Save on bandwidth costs.** Replication in general vastly improves on tape-based backup and restore but it can also raise bandwidth prices through the roof. ProtecTIER replicates only new

and unique deduplicated data, which take just a fraction of WAN bandwidth compared to incremental backup replication. For example, a replication customer who switches to ProtecTIER can replace its OC48 WAN connections with OC12 for an annual cost savings of over half a million dollars. Companies can choose to take the hard cost savings or keep their existing connections and replicate more application data for even better data protection.

3. **Strong DR and business continuity.** ProtecTIER replication is optimized for DR and business continuity. Failover and failback options between primary and remote sites ensure continuity during disaster recovery. ProtecTIER's compact bandwidth usage also enables more frequent testing of DR plans, while many-to-many and many-to-one configurations optimize ProtecTIER replication for individual customers.

4. **New level of application support.** Traditionally high replication costs and overhead limit replication to Tier 1 applications. ProtecTIER continually grows its replication support of production applications and associated data, which increases data protection for many types of additional data and applications. This improves DR and business continuity down the line.

5. **Meeting SLAs.** ProtecTIER cost-efficiently replicates deduplicated data to remote sites, which enables customers to improve data protection SLAs for many applications. Enabling rapid restores is even more critical to service levels. Fast disk-based recovery lets IT restore application data well within Recovery Time Objectives. ProtecTIER's policy engine also aids with meeting SLAs by automating replication actions around timeframes, capacity, and throughput.

Taneja Group Opinion

Replication is a basic data protection technology but IBM ProtecTIER replication has become anything but basic. Bandwidth reduction, powerful policies, many-to-one and many-to-many options, as well as centrally-managed replication grids have all made a huge difference in ProtecTIER's ability to optimize data protection.

Ever since ProtecTIER's launch, even before the advent of advanced replication, we liked and endorsed ProtecTIER for its superior deduplication algorithm, high capacity, scalability and high performance backup and recovery operations. Then IBM introduced a native replication option that was perfectly adequate for most ProtecTIER customer environments. But IBM was already a replication leader in its other product lines and they set out to vastly improve replication options for ProtecTIER. They have succeeded, and the latest powerful many-to-many, bidirectional, replication capabilities have helped to cement ProtecTIER's place as a data protection market leader.

This document was developed with IBM funding. Although the document may utilize publicly available material from various vendors, including IBM, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.

The information contained herein has been obtained from sources that we believe to be accurate and reliable, and includes personal opinions that are subject to change without notice. Taneja Group disclaims all warranties as to the accuracy of such information and assumes no responsibility or liability for errors or for your use of, or reliance upon, such information. Brands and product names referenced may be trademarks of their respective owners.