IBM PROTECTIER: FROM BACKUP TO RECOVERY

NOVEMBER 2011

When it comes to backup and recovery, backup performance numbers rule the roost. It's understandable really: far more data gets backed up than ever gets restored, and backup length is one of most difficult problems facing administrators today. But a reliance on backup numbers alone is dangerous. Recovery may not happen as frequently as daily backup *but recovery is the entire reason for backup.* Backing up because everyone does it isn't good enough. Backing up for compliance won't cut it. Backing up because you'll get fired otherwise isn't the point. Because when a recovery operation looms -- especially if the recovery is large and the potential loss is huge -- fast recovery performance becomes more urgent than backup ever was.

Now, we know that backup is an important challenge in and of itself. Backup is the foundation for recovery and slow backup performance can threaten the integrity of the entire backup and recovery process. Today's greater data volumes already take a long time to backup which threatens service levels. IT must address these important issues by demanding acceptable backup performance. But for all the attention we pay to backup performance, recovery performance is equally crucial -- yet it remains an afterthought.

Why? Because it is frankly harder to plan for and deploy recovery than it is to install backup. Recovery is not just its speed but is also its replication options, its bandwidth impact, and its disaster recovery potential. To make it even more complicated, recovery planning must account for different service level agreements. It must appropriately provide for every scenario from "we can wait a week for this application to come up" to "if we lose the last half hour of transactions then we might as well just go home." Complicated, yes. But the consequences of not preparing for recovery are far too high to take the easy way out.

Fortunately there are backup and recovery products out there that can make recovery planning far easier and much more successful. One such product is IBM ProtecTIER, which has one of the fastest recovery rates in the business. In this paper we will cover the ProtecTIER deduplication system's very fast restore speeds and excellent DR architecture. These capabilities and more make ProtecTIER an exceptional leader in the crowded backup and recovery market.

Recovery Isn't Just a Good Idea

Disaster recovery (DR) is the process of returning to normal operational levels following a serious loss of applications and data. The reasons behind disasters come in lots of different flavors including physical disaster, external hacking, administrator errors, computer-caused data loss or corruption, and even internal malfeasance (fancy way of saying disgruntled employee revenge). In

87 Elm Street, Suite 900 · Hopkinton, MA 01748 · T: 508.435.2556 · F: 508.435.2557

all of these cases DR is accomplished by restoring applications and data to acceptable usage levels without serious negative consequences to the business.

It's a hard job. The modern data center struggles with hugely growing volumes of data that threaten data protection service levels. When disaster strikes, the sheer enormity of dealing with this volume of data (not to mention applications, systems, users and networks) can be overwhelming. Yet in spite of the urgency of recovery, IT often concentrates on backup performance when evaluating backup and recovery technology. Recovery planning is an afterthought and may be relegated to the DR document that is gathering virtual dust on a network share.

IT can handle quick and dirty file restores and the occasional volume recovery from backup, no problem. But then the Big One hits: the critical database goes down and stays down; the hurricane hits the central data center; the earthquake knocks the server rack to smithereens and takes attached storage with it. These things happen, and IT must be able to restore applications and data -- and they have got to restore them fast.

RECOVERY READINESS

There are a set of requirements that any company serious about recovery should meet. They include:

- **Recovery performance.** Recovery speed is the simplest of the recovery metrics and is crucial to timely restoration. Tape-based volume recovery can be very slow and combing backup catalogs for individual files is no picnic either. Virtual Tape Libraries (VTL) let users use disk-based recovery without making massive changes to their backup infrastructure. Adding deduplication deeply compresses the backup volume and is a basic method for speeding up disk-based backup and recovery.

- **Recovery Time and Point Objectives.** A recovery solution must address both Recovery Time Objective (RTO) and Recovery Point Objective (RPO). RTO is the maximum amount of time that a given application can be unavailable without damaging the business. Under 12 hours is a common RTO for priority applications and truly crucial applications may be set as low as a zero-hour RTO. RPO is the maximum allowable time that the replicated copy may lag behind the primary data. A period of 24 hours is common given nightly backup schedules, but in the case of critical data 12 hours down to zero data loss may be the only acceptable RPO.

- **Replication.** DR requires replicating data to a remote site for fast restoration to the primary site or for host failover. Replication is fast and bandwidth-efficient, protects data integrity, and enables quick virtual cartridge recovery at the secondary DR site.

- **Bandwidth**. Companies with remote DR sites need wide bandwidth but must control costs. Replication operations should optimize bandwidth to write and restore the replicated data within service level guidelines, at a price that won't break a reasonable budget.

- **Failover and failback.** Failover/failback is the process of stopping replication from the primary site and switching operations over to the remote site. Once the primary site is up failback operations transfers processing and data back to the recovered production site. The more automated the process the better, since minutes matter when critical applications become unavailable. Look for recovery platforms that help to automate this process using policies.

87 Elm Street, Suite 900 · Hopkinton, MA 01748 · T: 508.435.2556 · F: 508.435.2557

- **DR testing.** Disaster recovery can be difficult to test over time. There are multiple points of failure to consider: LAN and WAN network connections, communications, secondary site status, operational plans, automation triggers, and more. No one technology can test everything and some concerns such as telephone availability will be out of IT's hands. But data protection technologies should offer DR self-testing abilities, which removes some of the onus from overworked IT.

IBM ProtecTIER and Recovery Readiness

One of the best ways we know to accomplish these levels of recovery readiness is IBM's ProtecTIER deduplication system. ProtecTIER offers essential data protection with deduplicated backup and very fast recovery rates. Features include powerful and flexible replication, highly automated failover and failback, bandwidth optimization and DR testing capabilities. These features work together to make ProtecTIER an excellent choice for backup and also for recovery.

ProtecTIER builds the foundation for recovery by dramatically reducing backup time: it replaces tape with high performance backup disk, and deduplicates incoming data for big capacity and time savings. Processing is extremely fast using an efficient index that resides in memory.

Recovery speeds are exceptionally fast at an up to sustained 2800 MB/sec (10 TB/hr) or above, which lets it restore data very quickly when needed. The same rate of speed operates over the WAN with replication and bandwidth optimization. Since ProtecTIER replicates only unique and deduplicated data, it relieves the intensive load on the wide area network, and also provides bandwidth control options for administrators. Replication can occur during its own time-window or simultaneously with backup, which allows IT to fully protect the deduplicated data while it enters primary ProtecTIER storage. ProtecTIER replication operates in one-to-one, many-to-one and many-to-many modes to provide complete flexibility for data protection and disaster recovery.

Let's take a closer look at ProtecTIER and the recovery enablers we mentioned above: recovery speed, recovery time and point objectives, replication and bandwidth, failover/failback and DR testing. It is this combination of capabilities that provides tremendous recovery advantages to ProtecTIER users.

EXCEPTIONALLY FAST RECOVERY PERFORMANCE

ProtecTIER's restore performance is even faster than their already fast backup speeds of sustainable 2000-plus MB/sec. Recovery speeds hit 2800 MB/sec (10 TB/hr) and higher sustained recovery performance. ProtecTIER's architecture enables this exceptionally fast performance by only storing unique, deduplicated data. When it comes time to recover, data restores efficiently and quickly.

This is very good news, especially for Tier 1 applications like databases that require immediate or near-immediate recovery. For example, ProtecTIER is a popular choice with SAP administrators who back up large databases twice daily because they cannot afford data unavailability or corruption. ProtecTIER's fast backup and recovery speeds greatly benefit data protection and enable quick recovery of mission-critical applications.

MEETING RPO AND RTO

ProtecTIER fulfills RPO and RTO for even the most demanding recovery requirements. Replication schedules may be set to the proper service agreement level for any given

87 Elm Street, Suite 900 · Hopkinton, MA 01748 · T: 508.435.2556 · F: 508.435.2557

application, which may range from immediate RPO or RTO to hours or even days depending on the application priority.

RPO: A less critical application might be all right with restoring data from the point of the most recent backup, which might have run a maximum 24 hours ago with a single daily backup or a maximum of 12 hours ago with a twice-daily backup. But Tier 1 applications may require RPOs within a few minutes. Zero-loss RPO scenarios should be fully mirrored to redundant systems that take over immediately should there be an interruption in processing. For these circumstances, ProtecTIER supports running replication simultaneously with backup.

RTO: Using ProtecTIER to restore from disk instead of tape automatically speeds up the recovery process to the tune of a few hours instead of several days. As with RPO, one size does not fit all and IT needs to assign priority to application requirements. ProtecTIER enables fast and flexible recovery options for differing service level agreements. For example, IT can use ProtecTIER to create fully redundant backup and restore systems with immediate failover and up-to-the-second data processing.

POWERFUL REPLICATION

ProtecTIER offers integrated IP-based replication with the flexibility of one-to-one, many-to-one and many-to-many choices. (Many-to-one copies data from multiple source repositories to a single destination repository; many-to-many provides bi-directional replication between 2-4 systems in a replication grid.) All of these replication options allow IT to optimize backup data protection between data centers, DR sites and remote offices.

Users may schedule replication by using either preset times or concurrent to backup and deduplication. They may also choose to manually launch the process. Best practice is to set scheduled

ProtecTIER Customers

#1: FROM TAPE TO PROTECTIER DISK

This company had backed up files to tape for many years. One large recovery effort involved 1 million files stored on 10GB of tape. The restoration took over 17 hours to complete.

The company quickly made the switch to disk-based deduplication backup using IBM ProtecTIER. They expected a faster recovery time but they did *not* expect the extreme recovery speed of their next restoration project, which was even larger than the first. This time they had to restore 1.6 million files stored on 34GB on their ProtecTIER system. The grand total of recovery time? 1 hour and 38 minutes; a far cry from the previous tape-based project.

#2: TRIPLED DATA GROWTH

An IBM customer's data had tripled in just a few years. Over a petabyte of this data was contained in mission-critical databases. Over the years frequent tape-based backup had resulted in over 7 PB stored on physical tape. Backup to tape was taking 15 hours a day and recovery from this volume size was extraordinarily challenging.

The company adopted IBM ProtecTIER gateway clusters. In combination with TSM they experienced much faster backup time with better disk space reclamation and a smaller physical footprint. And recovery? Recovery time for their crucial Oracle database applications was slashed by more than 50%.

87 Elm Street, Suite 900 · Hopkinton, MA 01748 · T: 508.435.2556 · F: 508.435.2557

replication options in ProtecTIER's policy engine so it can automatically identify data status and priority in the replication queue. ProtecTIER's fast replication -- up to 128 cartridges simultaneously, each containing 32 concurrent data streams – provides for fast replication *and* fast recovery even over the WAN (256 cartridges and 64 concurrent data streams with a ProtecTIER dual-node cluster).

OPTIMIZED BANDWIDTH

ProtecTIER preserves replication bandwidth by only replicating deduplicated data that is new and unique. It also provides optional bandwidth management features that allows IT to support the maximum replication transfer rate allowed for a specific repository. This capability reduces bandwidth needs by 90% and more compared to uncompressed, undeduplicated data transfer.

This results in huge bandwidth cost savings, which allow users to protect all of their applications and not just a few chosen business-critical ones. Administrators can afford the level of bandwidth that they really need for restoring data between remote sites. Optimized bandwidth is crucial for this level of disaster recovery, where a remote host must restore critical data to a recently recovered primary site.

AUTOMATED FAILOVER/FAILBACK

Failover and failback events require extremely high recovery performance. This requires recovery speed at and between multiple sites in several situations: 1) restoring data from a replication site to an operational primary site, 2) during failover at a secondary site when restoring backup cartridges to the new host and 3) during failback when the secondary host restores data back to the restored primary host.

These scenarios all require high levels of preparation including secondary DR sites, creating ProtecTIER policies, and establishing replication procedures. But once the prep work is done, ProtecTIER will quickly accomplish failovers, failbacks and data restores as needed to get critical applications back up and running on time.

NATIVE DISASTER TESTING

ProtecTIER provides disaster recovery testing capabilities for the replicated repository at the DR site, which makes IT's job much easier. For example, IT might test DR settings by starting the remote ProtecTIER host and running a typical subset of deduplicated backup jobs. Production volumes will remain in read-only mode to prevent errors during testing. ProtecTIER also provides a command line interface (CLI) to allow many tools and capabilities to the user including checking for uncompleted replication jobs.

Taneja Group Opinion

Backup performance is relatively straightforward and it is simple to pay attention to the numbers surrounding it. Yet because recovery is more complex and more infrequent than backup, it goes begging far too often. But recovery is backup's end game and poor recovery practices result in sharp revenue losses, lost productivity, failed regulatory compliance, and unrecoverable critical data.

Before this happens to you, look to backup and recovery systems that offer deduplication, high backup performance AND high recovery performance. Make sure that those high recovery speeds

87 Elm Street, Suite 900 · Hopkinton, MA 01748 · T: 508.435.2556 · F: 508.435.2557

work locally for restoration data at the primary site *and* remotely for big disaster recovery. Then add the really hard questions: Does the system have native replication that I can customize to my needs? Does it optimize bandwidth so I can afford the level of DR protection I need? Can it automatically failover to a secondary site and then failback, and can it do these things fast?

When the answer to all of these questions is an unqualified Yes – as it is with IBM ProtecTIER – then we strongly suggest that you look very carefully at this platform for your backup, recovery and complete disaster protection needs.
