



# LAB VALIDATION REPORT

## **IBM TS7650G ProtecTIER™** Enterprise-class Data De-duplication

**By Brian Garrett**  
With Claude Bouffard

**November, 2008**

# Table of Contents

<b>Table of Contents</b> .....	<b>i</b>
<b>Introduction</b> .....	<b>1</b>
<i>Background</i> .....	1
<i>Introducing the IBM TS7650G</i> .....	2
<b>ESG Lab Validation</b> .....	<b>4</b>
<i>Getting Started</i> .....	4
<i>Performance</i> .....	6
<i>Capacity Savings</i> .....	10
<i>Fault Tolerance</i> .....	12
<b>ESG Lab Validation Highlights</b> .....	<b>14</b>
<b>Issues to Consider</b> .....	<b>14</b>
<b>ESG Lab's View</b> .....	<b>15</b>
<b>Appendix</b> .....	<b>16</b>

## ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about emerging technologies and products in the storage, data management and information security industries. ESG Lab reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objective is to go over some of the more valuable feature/functions of products, show how they can be used to solve real customer problems and identify any areas needing improvement. ESG Lab's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments. This ESG Lab report was sponsored by IBM.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. Copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482.0188.

# Introduction

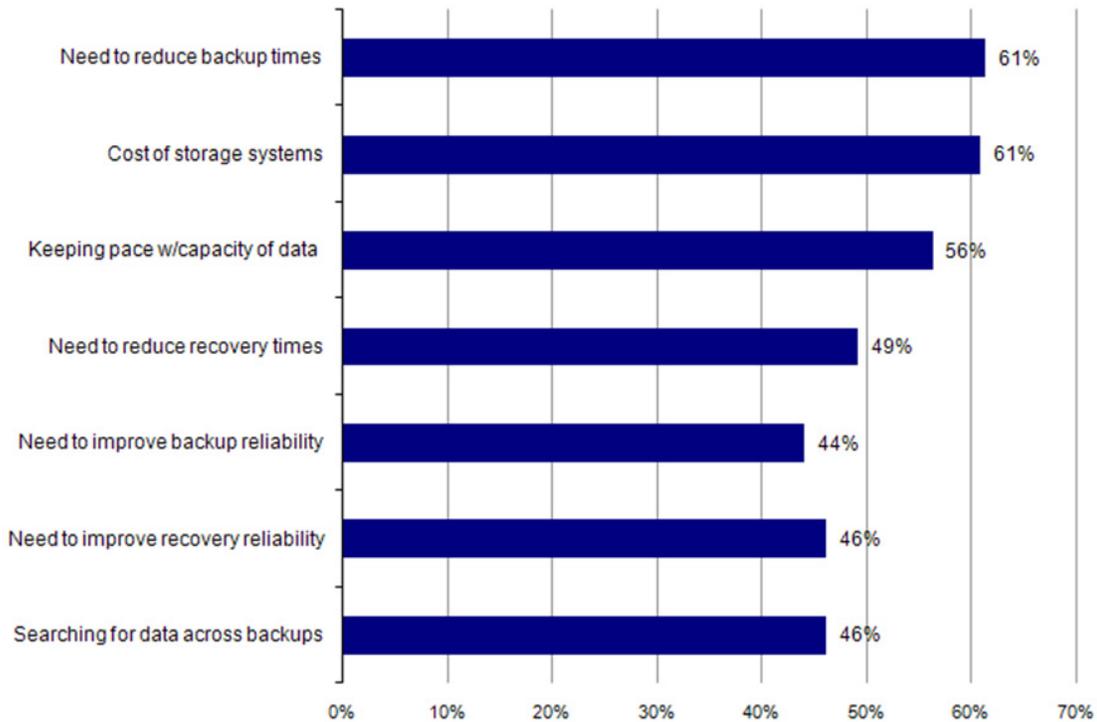
The IBM TS7650G, based on innovative Diligent ProtecTIER technology acquired by IBM in 2008, was designed to deliver the scalability, performance, and ease of use required to meet the data protection needs of midrange and enterprise-class data centers. This ESG Lab Report examines the enhanced ease of use, performance, and fault tolerance of the latest version of ProtecTIER inline de-duplication software running within a clustered pair of IBM TS7650G appliances.

## Background

A recent ESG survey of IT decision makers within enterprise-class organizations with more than 1,000 employees indicates that the need to reduce backup times and the cost of storage systems top the list of data protection challenges.<sup>1</sup> At the root of these challenges is an inability to keep pace with the growing capacity of information that needs to be protected. As a matter of fact, enterprise-class organizations are more likely to report capacity growth as a challenge (61% as shown in Figure 1 vs. 52% for all respondents). These are also more likely to be faced with stringent service level agreements, compliance initiatives, and legal discovery requests—all of which are driving the need to reduce recovery times and improve the reliability of backup and recovery processes.

**FIGURE 1. ENTERPRISE-CLASS DATA PROTECTION CHALLENGES**

Which of the following are challenges to your organization's current data protection processes and technologies  
(1,000 or more employees, N=197, multiple responses accepted)



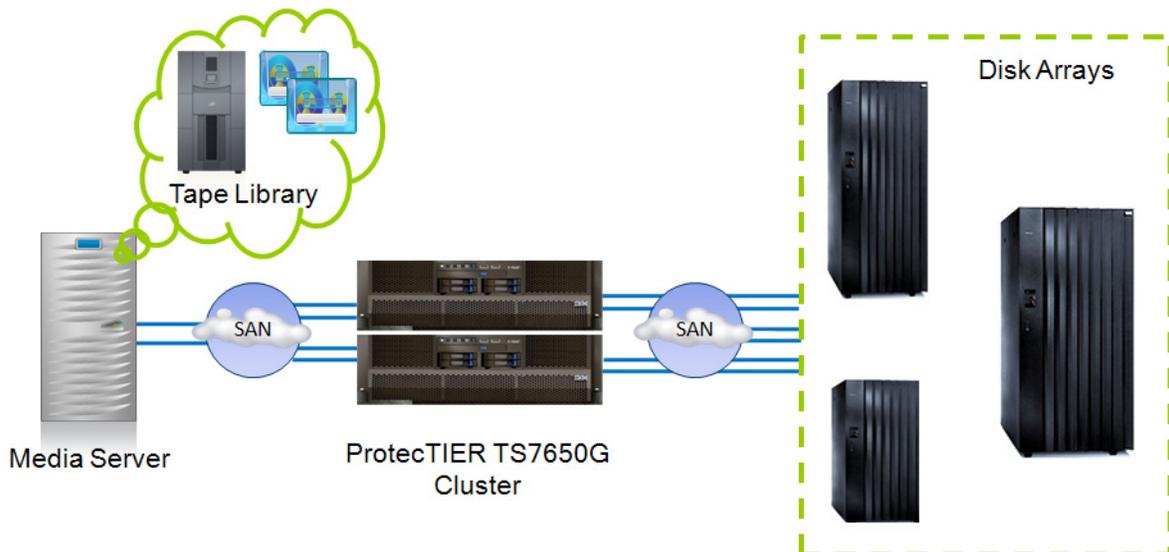
<sup>1</sup> Source: ESG Research Report, *Data Protection Trends*, January 2008

Enterprise-class data protection challenges are driving IT managers to adopt a number of technologies, including backup to disk, virtual tape library (VTL), and data de-duplication. As a matter of fact, 73% of enterprises surveyed by ESG indicate that they are using a disk-to-disk-to-tape backup process, 28% have deployed a VTL solution, and 22% have deployed data de-duplication with another 38% planning to do so. Clearly, these technologies are changing the way that organizations deal with growing backup and recovery challenges.

## Introducing the IBM TS7650G

The IBM TS7650G is a gateway-based VTL solution used for disk-to-disk backup as shown in Figure 2. Backup software running on a media server connects with one or more TS7650G servers over a Fiber channel (FC) storage area network (SAN). ProtecTIER software running on the TS7650G servers emulates one or more tape libraries. From the media server's perspective, disk capacity managed by the TS7650G behaves like a tape library, so there is no need to change existing backup software, policies, or procedures. Acting as a gateway, the IBM TS7650G servers manage access to virtualized pool of disk capacity within one or more FC attached disk arrays.

**FIGURE 2. AN ENTERPRISE-CLASS DATA PROTECTION ARCHITECTURE**



ProtecTIER software running on TS7650G gateway servers provides:

- Virtual tape library emulation for capacity within one or more SAN attached disk arrays
- Easy deployment and integration with existing backup software, infrastructure, and processes
- Enterprise-class capacity and performance scalability
- Enterprise-class data integrity
- Inline data de-duplication, which can be used to reduce retained capacity requirements by 90% or more

Inline data de-duplication is powerful technology capable of drastically reducing the capacity required to store backup data on disk. The concept of data de-duplication is simple—when multiple copies of the same data are sent to a system, the system finds the redundancy and stores only one copy of the data as it maintains an index to keep track of all data within the system. The motivation for data de-duplication is also simple—storing fewer copies of the same data can significantly reduce the capacity required to keep backup images on disk for quick and reliable restores.

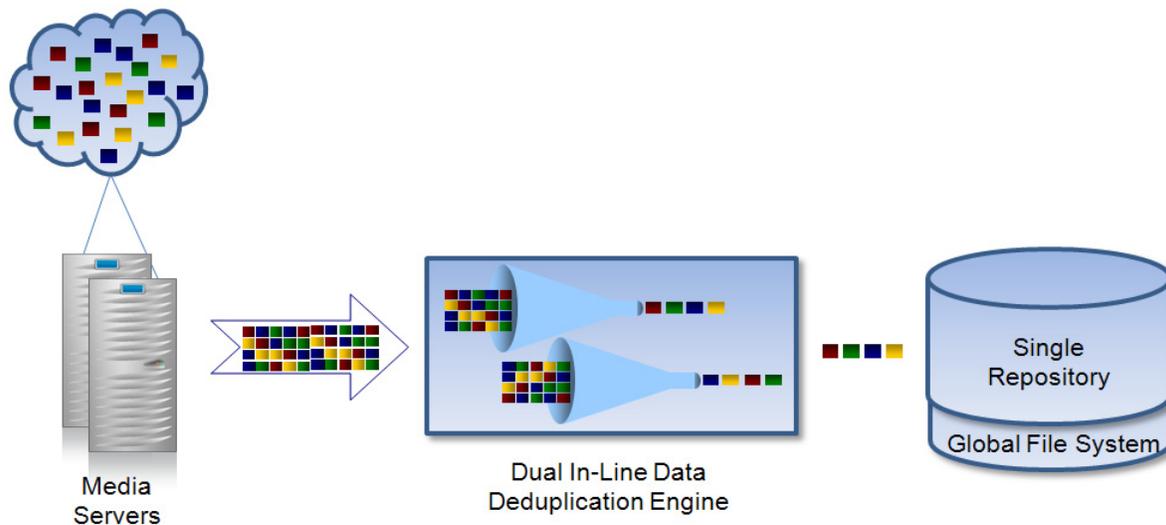
The concept of inline de-duplication is also simple. It eliminates redundant data as it is being backed up. Compared to post-process de-duplication, which removes redundant data after backup has landed on disk, inline de-duplication eliminates the capacity required to store backup data in its native format. Inline de-duplication also eliminates the performance impact that post-process de-duplication can have on processes that typically run after backup job have completed (e.g. a clone to physical tape for offsite vaulting or the attempted recovery of corrupt data).

To better understand data de-duplication technology, consider the example of a PowerPoint presentation attached to an e-mail. If the e-mail is sent to multiple recipients and then forwarded to yet another set of recipients, data de-duplication technology can be used to store the presentation only once. This is an example of data de-duplication technology working at the file level. Next, consider what happens when one of the e-mail recipients modifies a slide in the presentation and forwards it to a group of colleagues. Block level data de-duplication algorithms, like those utilized by ProtecTIER, can be used to store only the new, unique data associated with the changed slide.

The inline de-duplication algorithm at the core of the ProtecTIER platform is called HyperFactor™. Using an efficient in-memory index to keep track of where duplicate data is stored, HyperFactor was designed to provide high-speed de-duplication services for organizations with enterprise-class performance, scalability, and availability requirements.<sup>2</sup>

The latest version of ProtecTIER software includes support for a two node cluster of inline de-duplication gateways. As shown in Figure 3, a pair of TS7650G gateways running ProtecTIER software is used to create a single repository of de-duplicated backup data. The two gateways share a global file system stored on disk to multiply the performance and capacity savings that can be achieved with a single node. With state information stored safely on disk, this approach can also be used to ensure that backup and recovery operations are available after a hardware failure.

**FIGURE 3. PROTECTIER HYPERFACTOR DATA DE-DUPLICATION**



The balance of this report presents the results of ESG Lab testing, which was designed to confirm IBM's claims of up to 900 MB/sec or more of aggregate backup performance, retained capacity savings of 90% or more, and enterprise-class fault tolerance of ProtecTIER version 2.1 software running on a clustered pair of IBM TS7650G gateways.

<sup>2</sup> The TS650G is ideally suited for organizations that need to backup more than 10 TB of data nightly.

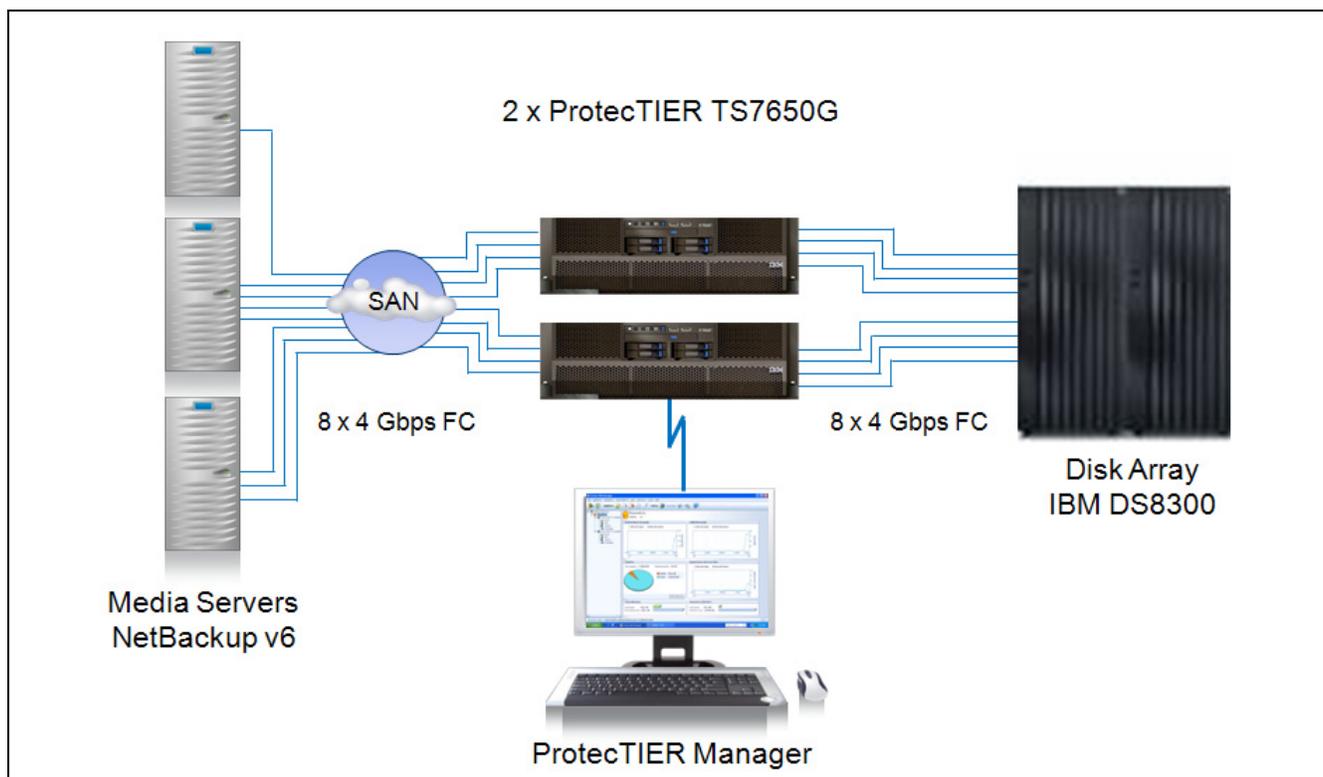
## ESG Lab Validation

ESG Lab performed hands-on testing of an enterprise-class TS7650G inline de-duplication solution at an IBM facility located in Tel Aviv, Israel. A pair of TS7650G gateways were deployed as a fault tolerant active-active cluster to test the speed, reliability, and recoverability of backup and recovery operations while avoiding the inconveniences of changing existing backup software, infrastructure, policies, and processes.

### Getting Started

Three quad core servers running Red Hat Linux and Symantec NetBackup software were used as media servers for backup and recovery operations.<sup>3</sup> The media servers were connected to the TS7650G appliances through a FC SAN. A total of eight 4 Gbps FC connections provided up to 32 Gbps of theoretical bandwidth for backup and restore operations. The TS7650G gateways were connected to an IBM DS8300 disk array using eight 4 Gbps FC connections. A FC SAN was implemented between the media servers and the TS7650G appliances using a QLogic SANbox 5200 FC switch. The storage array was direct connected to the TS7650G gateways.<sup>4</sup>

**FIGURE 4. ESG LAB TEST BED**



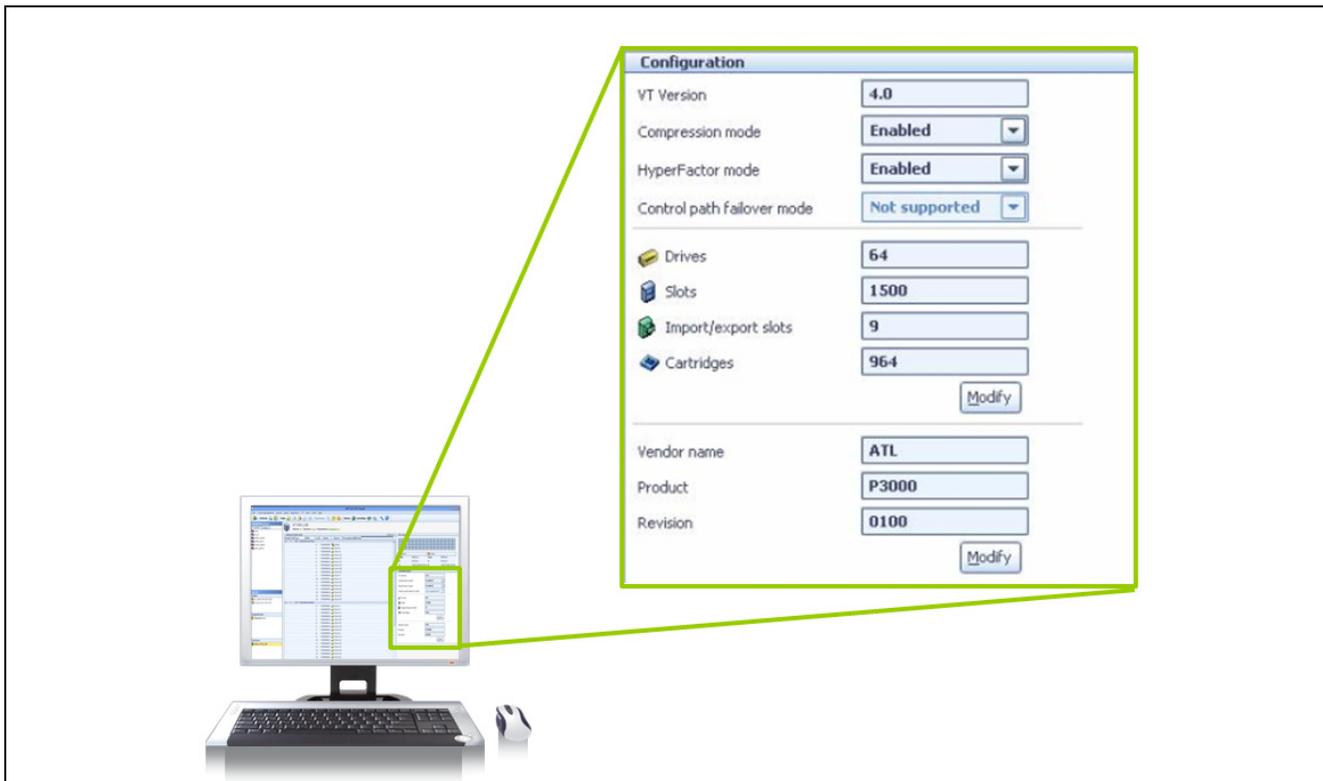
A two-bay IBM DS8300 disk array housed 256 300 GB 15K RPM FC drives. The DS8300 was configured with 50 TB of usable backup capacity accessed through 36 RAID-5 LUNs.<sup>3</sup> The ProtecTIER management console was used to configure and monitor the solution through an Ethernet connected web browser.

ProtecTIER Manager was used to browse the test bed, beginning with the virtual tape library configuration. As shown in Figure 5, the system was configured as an ATL P3000 tape library with 64 tape drives and 964 tape cartridges.

<sup>3</sup> See the Appendix for more configuration details.

<sup>4</sup> While IBM supports SAN connections as well, it should be noted that the direct connect methodology used during ESG Lab testing reduces the number of FC switch connections which lowers the cost of the total solution.

**FIGURE 5.** VIEWING VIRTUAL TAPE LIBRARY CONFIGURATION WITH PROTECTIER MANAGER



ESG Lab noticed that the GUI has been enhanced since it was tested by ESG for the first time in 2006. Most noticeable was the addition of tabs to manage from either the system- or node-level due to the introduction of clustered support in ProtectIER version 2.1. The user interface has also improved in terms of usability. The clear and valuable depiction of used versus nominal disk space, which shows the capacity utilized and savings provided by ProtectIER, are more prominently displayed. The excellent performance and trending graphs, absent in a number of early data de-duplication solutions tested by ESG Lab, have been enhanced as well. It's obvious that IBM's excellent GUI design principles are working their way into the ProtectIER management console. ESG was also pleased to see that ProtectIER documentation has been expanded and enhanced in accordance with IBM standards.

Pre-configured Veritas NetBackup resource definitions and backup policies were reviewed. From a backup administrator's perspective, managing the 64 tape drives presented by the VTL software running on the TS7650G gateways felt exactly the same as a physical ATL P3000 tape library.

### Why This Matters

Unrelenting capacity growth and shrinking backup windows are driving a growing number of enterprises to deploy backup to disk processes. As a matter of fact, a recent ESG survey indicates that 70% of enterprise-class organizations have already deployed a backup to disk solution due to the fast backup and recovery performance of disk compared to tape. FC-attached VTL technology simplifies the deployment and management of a disk-based backup solution—especially for large organizations that have standardized on FC SAN technology. ESG Lab recently spoke with a customer, who summed it up well when he said, "Deploying a pair of IBM TS7650G gateways was simple. My backup administrators didn't have to learn anything new."

## Performance

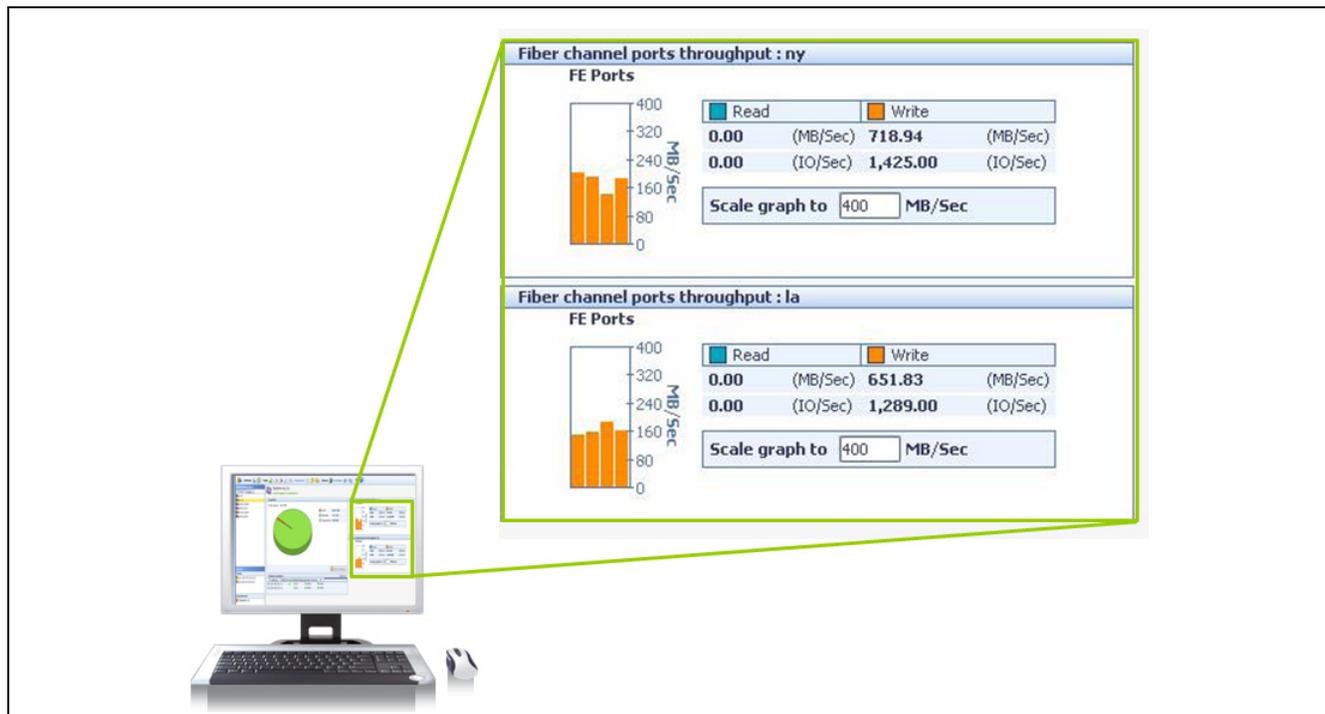
Midrange and enterprise-class data centers backing up ten terabytes of data or more nightly are faced with a number of conflicting challenges. The system needs to be fast to avoid a missed backup window. The system needs to store and retain tens, or hundreds, of terabytes of backup data for quick and reliable restores. And the cost of the system, including the cost of the relatively expensive disk drives, has to fit the budget. While inline de-duplication can drastically reduce the amount of disk capacity (and hence the cost), it is a complex operation that must be done in real-time, it has to be fast, and it needs to support a large pool of capacity so that it can find and eliminate duplicate data throughout the enterprise.

The HyperFactor technology at the heart of the IBM TS7650G solution was designed to meet the performance and capacity needs of enterprise-class data center environments. ProtecTIER version 2.1 software extends the performance and capacity scalability of HyperFactor technology using a clustered pair of TS7650G gateways, which implement de-duplication over a shared pool of capacity. Supporting up to a petabyte (1PB) of physical capacity and twenty petabytes (20 PB) or more of retained backup capacity, IBM claims that a pair of TS7650G gateways can deliver up to 900 MB/sec of sustained aggregate backup performance.

### ESG Lab Testing

IBM tested the performance of a single and dual-node TS7650G solution using the configuration depicted in Figure 4 and documented in the Appendix. Backup and restore performance was monitored using the ProtecTIER manager GUI as shown in Figure 6. The aggregate performance of 64 backup jobs running in parallel was used to confirm that ProtecTIER was reporting the same performance as NetBackup.<sup>5</sup>

**FIGURE 6. MONITORING AGGREGATE BACKUP PERFORMANCE WITH PROTECTIER**



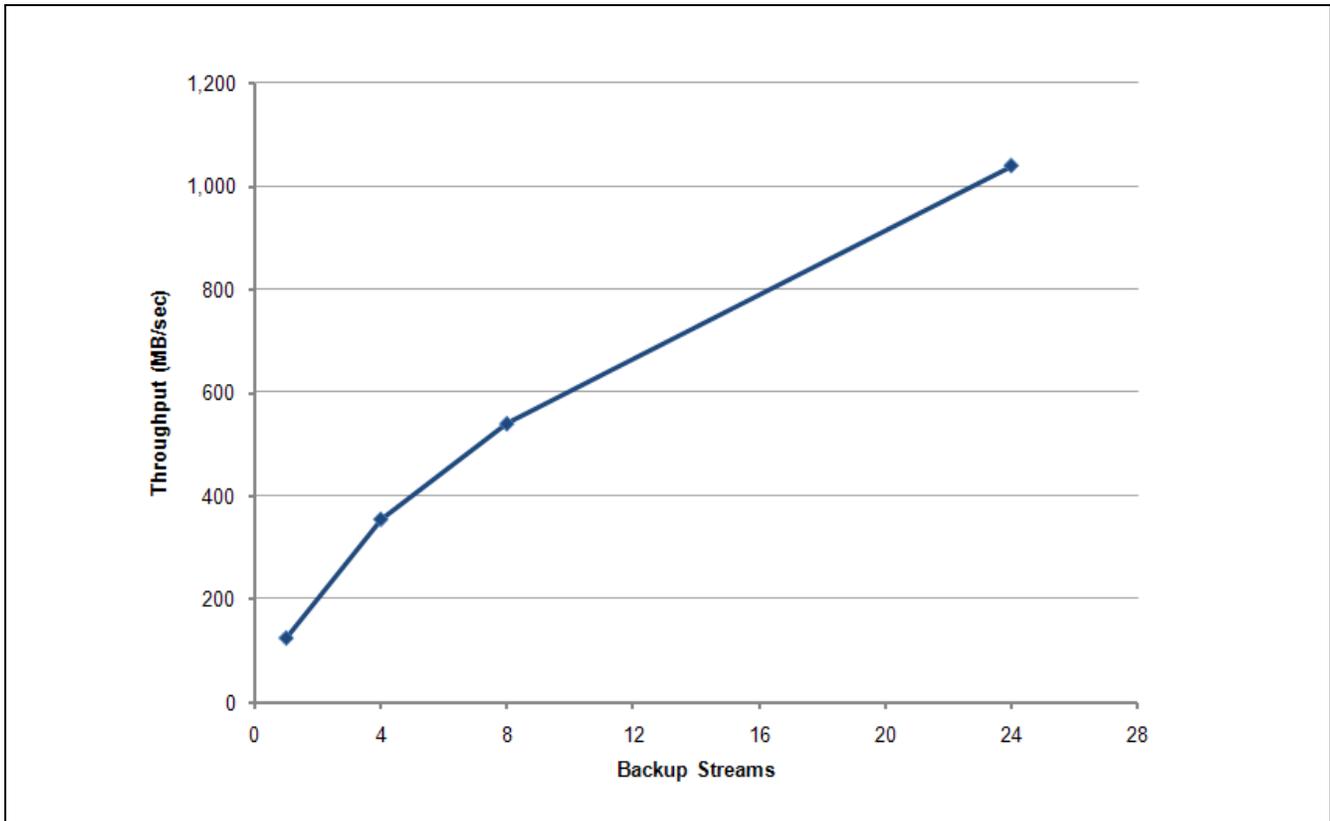
The performance screenshot shown in Figure 6 was taken as 60 backups ran a 3<sup>rd</sup> consecutive full backup with a data change rate of 3% between each backup. Results for the full series of backup and restore jobs will be presented in detail later in this section. This particular screen shot is presented to document performance during

<sup>5</sup> The sum of the throughput reported by each Veritas NetBackup job was compared to the performance shown in the ProtecTIER GUI.

the fastest aggregate performance recorded by ESG Lab (1,371 MB/sec). It also illustrates how performance is spread relatively evenly across each of the TS7650G gateways (NY and LA) and each of the front-end FC interfaces within each gateway (FE ports).

ESG Lab performance testing began with an analysis of how aggregate backup scales as the number of backup jobs running in parallel are increased. The performance scalability of a clustered pair of TS7650G gateways for the first set of full backups is depicted in Figure 7.

**FIGURE 7. ADDING STREAMS TO INCREASE AGGREGATE BACKUP PERFORMANCE**

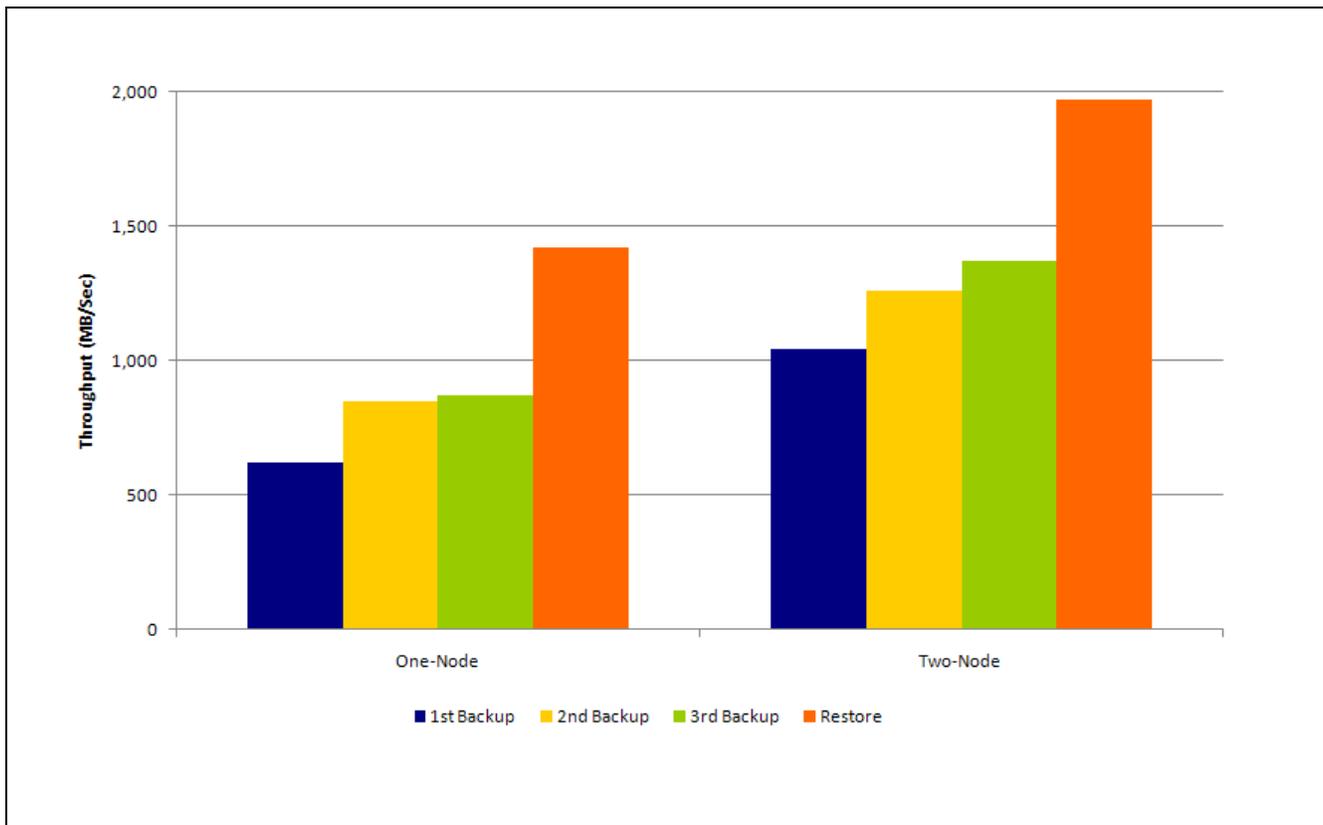


*What the Numbers Mean*

- The first full backup job, running alone (single stream), sustained 125 MB/sec of performance as inline de-duplication processed backup data in real-time and reduced the amount of data stored on disk.
- ESG Lab's experience testing inline de-duplication solutions indicates that 125 MB/sec for a single backup stream is an excellent level of performance.
- A sustained performance level of 125 MB/sec can be used to protect up to 3.6 TB of data in an eight hour shift.
- Performance scales in a near linear fashion as the number of backup-job streams is increased.
- A maximum performance of 1,040 MB/sec (1 GB/sec) for 24 full backup streams running in parallel exceeds IBM's conservative claims of up to 900 MB/sec of aggregate backup performance for a pair of clustered TS7650G gateways.

A series of daily full backup jobs with a daily change rate of 3% between each backup job was run. The aggregate performance for 64 backup jobs (streams) running in parallel was recorded. A Veritas NetBackup verify job was run after the third full backup to measure maximum theoretical restore performance.<sup>6</sup> Single and clustered dual-node TS7650G solutions were used to measure performance scalability when moving from a single node to a cluster of two TS7650G gateways. The results are shown in Figure 8.

**FIGURE 8. ONE VERSUS TWO NODE AGGREGATE PERFORMANCE**



**TABLE 1. PROTECTIER AGGREGATE PERFORMANCE ANALYSIS**

Operation	One Node (MB/sec)	Two Nodes (MB/sec)
1 <sup>st</sup> full backup	622	1,043
2nd full backup	845	1,260
3rd <sup>t</sup> full backup	870	1,371
full restore	1,420	1,965

<sup>6</sup> Normally, a verify job is used to verify the integrity of a previously backed up data. In this case, the verify operation was used to measure the maximum restore performance of TS7650G appliances without incurring the potential performance overhead of the file system and disk array where the data is being restored.

### *What the Numbers Mean*

- The aggregate backup performance of a two node cluster is significantly greater than a single node—even as the two nodes work together to de-duplicate a common pool of backup data.
- Compared to a single node, two nodes working together increases backup performance for the first full backup by 67% percent
- Aggregate performance increases for the second and third backup jobs. This is due in part to the fact that less data is being written to disk (only the 3% of data that has changed since the last backup). The speed and efficiency of HyperFactor inline de-duplication is also a factor.
- A maximum aggregate performance of 1,371 MB/sec (1.37 GB/sec) was recorded for the 3<sup>rd</sup> set of full backups. This is well in excess of IBM's conservative claims of 900 MB/sec.
- An aggregate backup rate of 1,371 MB/sec (4.9 TB/hour) can be used to protect 39.5 TB of data in an eight hour shift.
- A peak aggregate restore performance of nearly 2 GB/sec is impressive, considering the fact that HyperFactor is running the de-duplication algorithm in reverse. This clearly demonstrates the power and efficiency of the in-memory indexing algorithm at the heart of HyperFactor technology.

ESG spoke recently with an IBM customer that has deployed a single TS7650G gateway connected to an IBM XIV Storage System full of SATA disk drives. He currently has 280 TB of IBM TSM backup data stored on IBM XIV disk. An upgrade to a two node TS7650G solution is scheduled for the next maintenance window. The single node system has improved performance by 30%, compared to the enterprise-class tape library used previously. While he's impressed with the backup performance and expects that it will improve when he upgrades to a clustered pair of TS7650G gateways, he's most impressed with his new-found ability to more quickly respond to ad hoc restore requests.

### **Why This Matters**

ESG research indicates that the number one data protection challenge reported by IT professional within enterprise organizations is the need to reduce backup times.<sup>7</sup> Quicker recoveries are also needed to meet service level agreements and increase end-user productivity.

ESG Lab has confirmed via hands-on testing that IBM can easily exceed its aggregate backup throughput claims of 900 MB/sec for a two node TS7650G solution. A peak aggregate backup performance of 1.37 GB/sec and 1.97 GB/sec for restores was observed. Based on experience testing a number of backup to disk solutions, ESG Lab is extremely impressed with the performance of a clustered IBM TS7650G solution—especially given the fact that it performs inline de-duplication to reduce the amount of disk capacity required to retain backup data on disk.

---

<sup>7</sup> Source: ESG Research Report, *Data Protection Trends*, 2008

## Capacity Savings

Data de-duplication is a resource-intensive process of examining data to identify and eliminate redundancy. It can have a significant impact on the capacity of data stored, which, in turn, can deliver significant economic benefits. IBM TS7650G solutions use an inline de-duplication method, which eliminates duplicate data as it is being backed up. The solutions use an efficient algorithm that uses an in-memory index to keep track of duplicate data. This efficient approach, referred to as HyperFactor, is designed to meet the extreme scalability and performance requirements of enterprise-class data center environments.

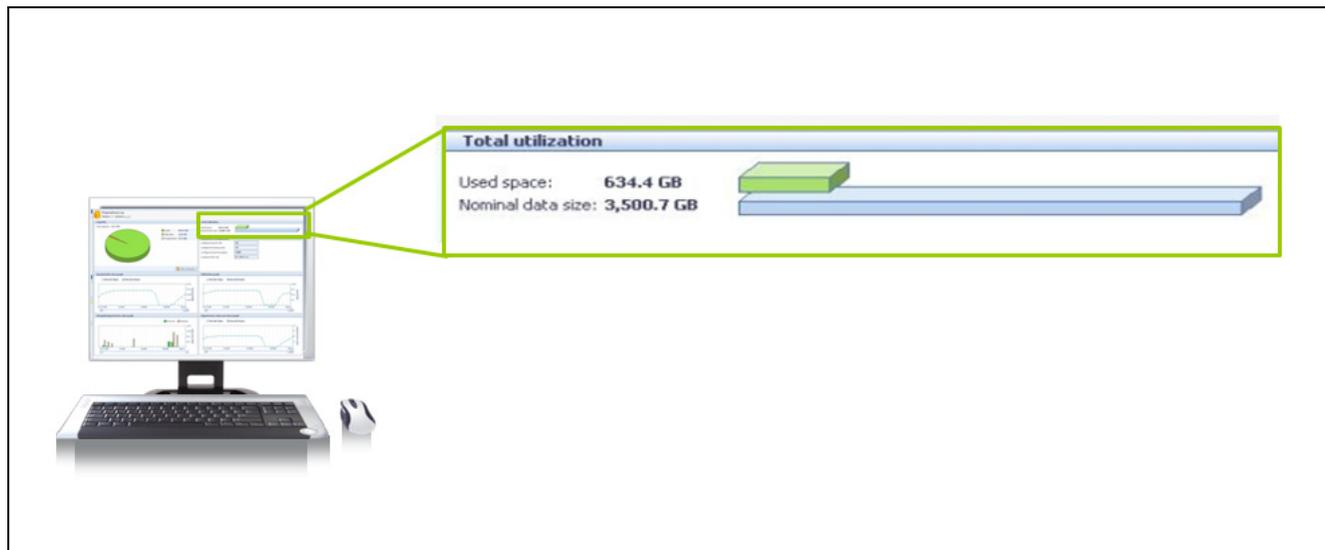
The level of disk capacity savings that can be achieved with data de-duplication varies according to the backup policy in use, the number of backup images retained on disk, (a.k.a. retention policy), the rate at which data is changing the amount of duplicate data found within an organization.. The ratio of capacity backed up versus capacity stored on disk after de-duplication is generally known in the industry as the data de-duplication factor. An ESG survey of early adopters of de-duplication solutions indicates that factors between 10:1 and 20:1 are commonly achieved.<sup>8</sup> A de-duplication ratio of 10:1 reduces disk capacity requirements by 90%.

### ESG Lab Testing

ESG Lab recorded the capacity backed up and stored on disk during the performance tests presented earlier in this report. An 875 GB collection of file data designed to mimic the contents of typical office productivity files (e.g., documents, spreadsheets, presentations) was used during this round of ESG Lab testing. The data de-duplication results presented below are similar to those recorded during a round of ESG Lab testing performed in 2006 using Oracle, Exchange, office productivity, and audio files harvested from production backup data sets. Daily changes were emulated using an IBM utility that randomly changed 3% of the data after each set of backup jobs had completed.

The amount of disk capacity backed up by Veritas NetBackup and consumed on disk after de-duplication was recorded using the ProtecTIER GUI after each backup job. A sample screen shot is shown in Figure 9. In this example, 3,500.7 GB of backed up data consumes only 634.4 GB of disk capacity.

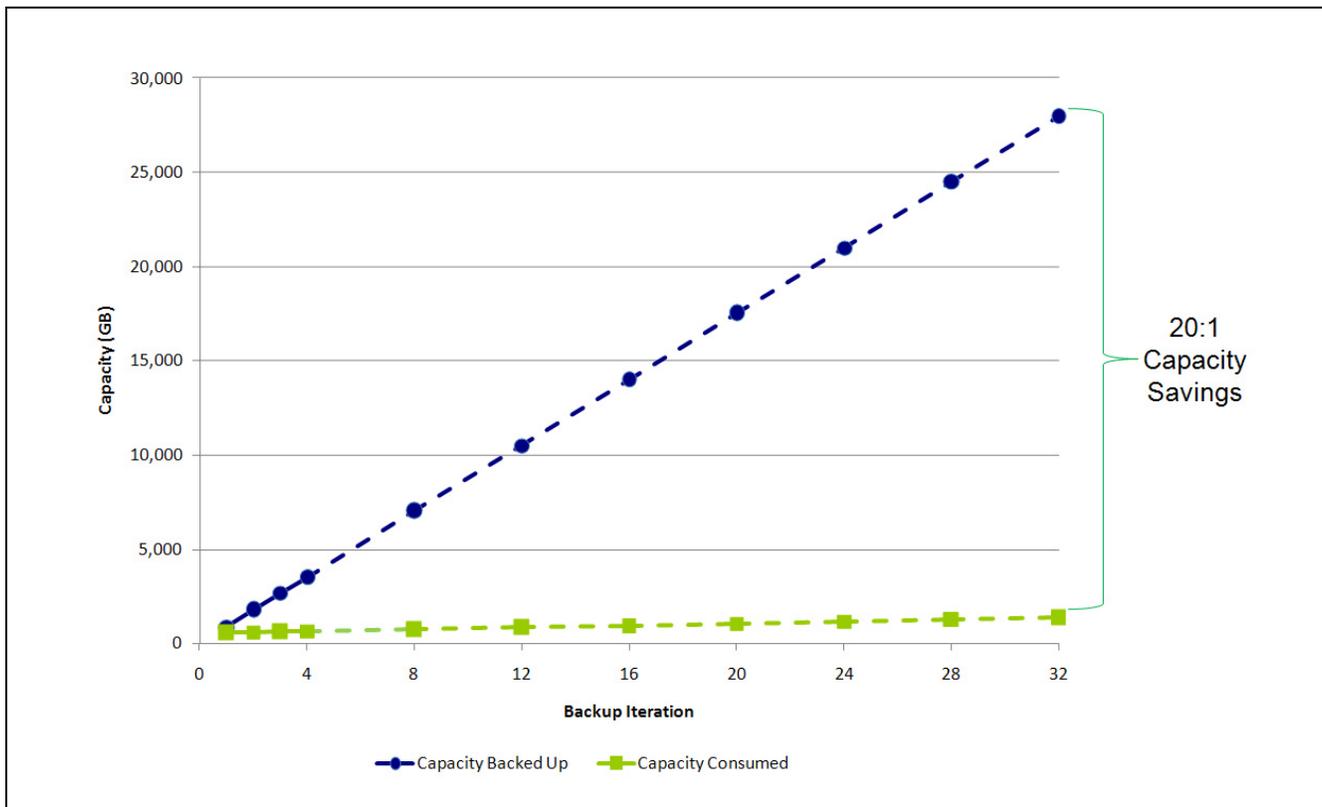
**FIGURE 9. PROTECTIER CAPACITY UTILIZATION CONSOLE**



Values recorded for the first four backups were used to project the savings for the 8<sup>th</sup> thru 32<sup>nd</sup> backup jobs. The results are shown in Figure 10.

<sup>8</sup> 28% report up to 10x reduction, 40% between 10x and 20x, 16% greater than 21x and 17% don't know. Source: ESG Research Report, *Data Protection Trends*, 2008

**FIGURE 10. DE-DUPLICATION CAPACITY SAVINGS**



**TABLE 2. PROTECTIER DATA DE-DUPLICATION CAPACITY SAVINGS**

Backup Iteration	Backed Up (MB)	Consumed (MB)	De-duplication Ratio
1 <sup>st</sup>	875	555	1.6
2 <sup>nd</sup>	1,750	582	3.0
3 <sup>rd</sup>	2,626	608	4.3
4 <sup>th</sup>	3,501	634	5.5
8 <sup>th</sup>	7,001	825	8.5
28 <sup>th</sup>	24,500	1,256	19.5
32 <sup>nd</sup>	28,000	1,361	20.6

### Why This Matters

ESG research indicates that cost is the leading reason why non-adopters have not yet embraced a disk-based backup strategy.<sup>9</sup> Data de-duplication addresses the cost issue by reducing the disk capacity required to maintain multiple generations of backup data on disk for quick and reliable restores. ESG Lab has confirmed that HyperFactor inline de-duplication can be used to reduce retained backup to disk capacity requirements by a factor of up to 20 to 1.

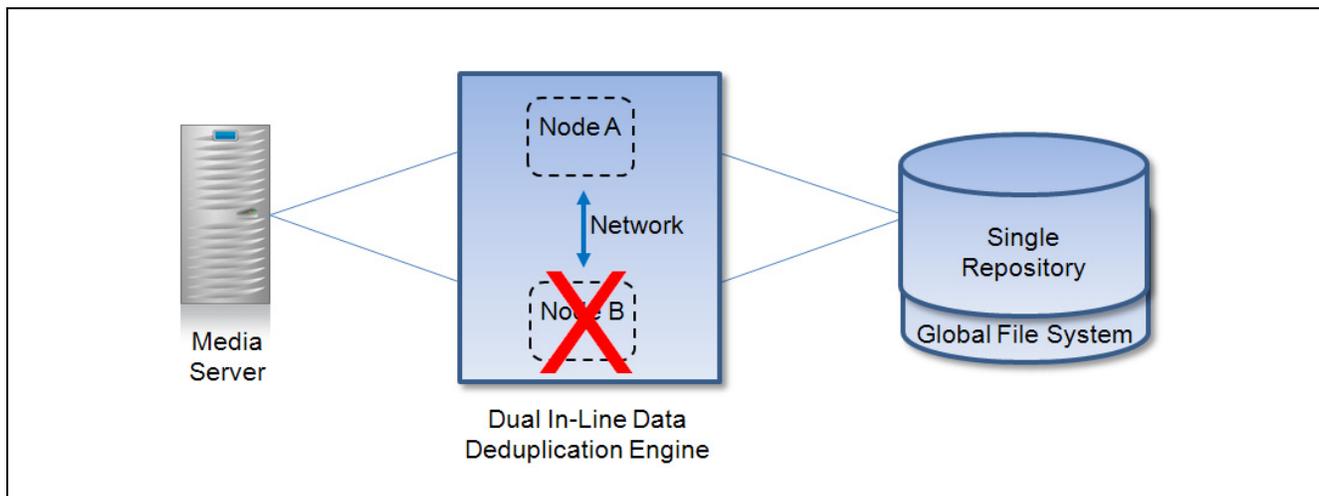
<sup>9</sup> Source: ESG Research Report, *Data Protection Trends*, 2008

## Fault Tolerance

Two ProtecTIER gateways can be configured as a cluster. The ProtecTIER clustered approach not only maximizes performance and de-duplication savings as shown earlier in this report, it also minimizes downtime in the unlikely event of a ProtecTIER TS7650G hardware failure.

ProtecTIER uses an active-active cluster approach that ESG refers to as a true cluster. As shown in Figure 11, a global file system, which is maintained on disk and shared by each node in the cluster, is used to create a single pool of highly available backup to disk capacity. With system configuration, state, and status information stored safely on disk, multiple nodes can work together while sharing a single repository. Any node can service backup or restore requests for any virtual cartridge. Clustering and the global file system ensure that backup and restore services are highly available.

**FIGURE 11. HIGH AVAILABILITY/DR TEST CASE**



If one of the nodes in the cluster fails, the system continues to be available for new backup and restore requests. Backup jobs running on virtual tape drives being serviced by a failing node eventually time out and fail in ProtecTIER version 2.1. Jobs running on the surviving node continue without error. Failed jobs are restarted and serviced by the surviving node. Virtual cartridges that appear to be stuck in a failed tape drive from a backup software perspective can be reassigned and found with a routine backup software inventory job if needed. A higher level of non-stop availability using host-based failover drivers is planned for a future release of ProtecTIER software.

### ESG Lab Testing

The configuration and data sets used during the performance and de-duplication tests presented earlier were used during this phase of testing. To simulate an outage, errors were injected during the fourth generation of backup jobs running on three media servers. Ten Veritas NetBackup jobs running in parallel were used to assess the performance and availability of a ProtecTIER cluster after a node restart, a node shutdown, and a node power failure.

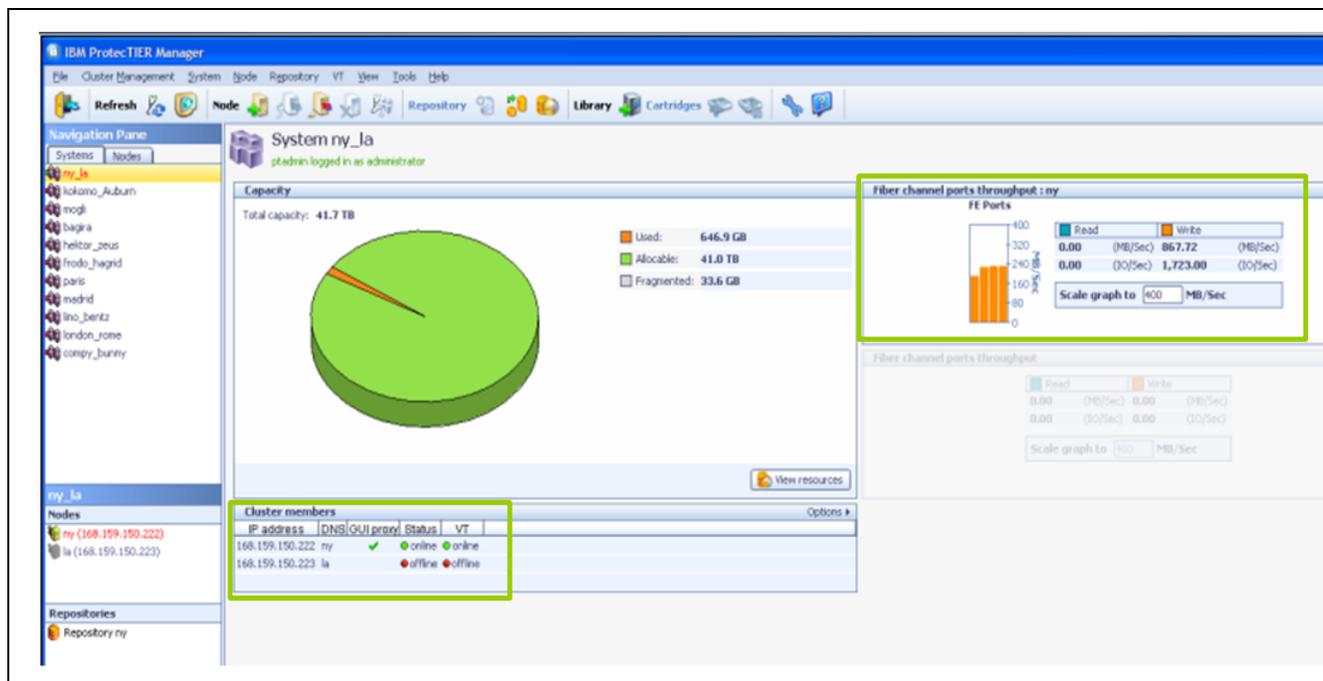
### Auto Restart

A manual restart of a ProtecTIER node was performed as backup jobs were balanced across both nodes in the cluster. A sustained aggregate backup rate of 1.3 GB/sec was noted ten minutes after the backup jobs had started. As one of the nodes was restarting, aggregate performance dropped sharply for a minute and then increased to 600 MB/sec. As expected, the aggregate throughput during the failure was roughly half that of a healthy two node system. Most of the jobs continued to completion after the restart. One of the five NetBackup jobs that had timed out was restarted and completed without error. ProtecTIER Manager was used to confirm that the job had been restarted and was being serviced by the node that had been restarted.

### Manual Shutdown

A manual shutdown of a ProtecTIER node was performed to assess the ability of the system to ride through a planned service event (e.g., a code or hardware upgrade). This test was also used to confirm that jobs running on a node that had been shut down are automatically migrated to a surviving node when the backup job is restarted. The screenshot shown in Figure 12 shows the state of the system after the shutdown. Note that the offline node is depicted in red. The surviving node, depicted in green, is online and acting as a proxy for the management GUI. It should also be noted that the surviving appliance is delivering excellent aggregate performance of 867 MB/sec.

**FIGURE 12. PROTECTIER SYSTEM STATUS AFTER A NODE SHUTDOWN**



As expected, the jobs running on the surviving node completed without error. Jobs that were being serviced by the shut down node hung and eventually timed out. Veritas NetBackup reported that it “cannot write image to media” due to an input/output error. While the node was shutdown, one of the failed jobs was restarted and completed without error. ProtecTIER Manager was used to confirm that the virtual tape library and cartridge used by the restarted job had been automatically migrated to the surviving node.

### Power Failure

A power failure was tested to verify that backup jobs being serviced by a failed node can be reassigned, discovered, and restarted. The node power failure was induced as ten backup jobs were running on one of the NetBackup media servers. Before the power failure, it was noted that jobs were balanced across both nodes in the controller. Jobs running on the surviving node completed without error. ProtecTier Manager was used to unload and move one of the virtual tape cartridges being used by the failing node at the time of the power failure. The virtual tape drives cartridge was reassigned to the surviving node. A NetBackup inventory job discovered the cartridge and a NetBackup verify of the virtual tape cartridge completed without error.

## Why This Matters

Midrange and enterprise-class organizations with mission critical applications that can tolerate little or no downtime cannot afford the risk of their data protection infrastructure being unavailable. ESG Lab has confirmed that ProtecTIER clustering technology can be used to ensure that backup and recovery services remain available in the unlikely event of an IBM TS7650G hardware failure.

## ESG Lab Validation Highlights

- ☑ From a backup administrator's perspective, managing backup jobs running on a virtual tape library presented by ProtecTIER felt exactly the same as a physical tape library.
- ☑ ESG Lab noted that the ProtecTIER GUI and documentation are benefitting from the use of IBM standards and guidelines.
- ☑ Excellent single stream backup performance of 125 MB/sec was recorded.
- ☑ A dual-node cluster reached a peak aggregate backup performance up to 1,371 MB/sec.
- ☑ A peak aggregate restore rate of 1,965 MB/sec was measured.
- ☑ Inline data de-duplication capacity savings of 20 to 1 were confirmed using a daily full backup policy, a three percent daily change rate, and a retention period of 30 days.
- ☑ Availability of backup and recovery services after an appliance reboot, shutdown and power failure was confirmed.

## Issues to Consider

- ☑ ESG Lab believes that extremely large enterprise-class data center environments would benefit from the support of more than two nodes per cluster to further improve performance and data consolidation. IBM has advised ESG that multiple node cluster support is planned for a future release. Of the TS7600 line of offerings
- ☑ While optional two-node clustered support increase the fault tolerance of a TS7650G solution, multi-path failover drivers and integration with industry leading software would provide even greater levels of non-stop reliability. IBM has advised ESG Lab this is planned for a future release.
- ☑ The results presented in this document were obtained in a controlled test environment which was designed to stress the performance limits of TS7650G solutions. Performance in production environments will vary due to a number of factors including the performance of file systems, primary storage systems and networks.

## ESG Lab's View

IT managers have been struggling for decades to keep up with the relentless growth in the volume of data that needs to be protected. The challenges are particularly acute for IT managers responsible for the protection of vital information assets within medium to enterprise-class data centers. Traditional backup methods are straining under the load. Backup windows are shrinking. The number of recovery requests is increasing. Budgets are being stressed. The frequency of legal and regulatory discovery requests is rising. Risk levels and blood pressures are rising as a growing number of IT managers wonder if data can be recovered quickly.

A growing number of organizations are turning to disk-based backup methods to improve the speed and reliability of backup and recovery operations. As a matter of fact, ESG research indicates that 73 percent of enterprise-class organizations have already deployed some form of disk-based backup solution.<sup>10</sup> While disk can be used to accelerate performance, the costs are high compared to legacy tape solutions. Data de-duplication addresses the cost issue by reducing disk capacity requirements. As a result, a growing number of organizations (22 percent according to a recent ESG survey) have deployed a disk-based data de-duplication system.

ESG Lab was impressed with the disk-based ProtecTIER inline de-duplication architecture when it was first tested in 2006. Initial deployment using existing backup software and processes was easy. Enterprise-class levels of performance were measured (266 MB/sec for a single server running ProtecTIER inline de-duplication software). Capacity was reduced by a factor of 28 to 1 for a month's worth of retained full backups.

Since then, ProtecTIER solutions have been adopted by more than 200 world-wide customers, many with multiple deployments. More than 30 PB of managed capacity has been deployed in some of the largest enterprise-class organizations in the world including three of the top ten global telecommunications firms. More recently, ProtecTIER technology was acquired by IBM and version 2.1 software with optional two-node cluster support running on a purpose-built IBM Gateway server was released.

IBM TS7650 gateways and two-node cluster support dramatically increased the capacity and performance scalability of a ProtecTIER solution. ESG Lab hands-on testing in 2008 confirmed backup performance of up to 1,371 MB/sec which is well in excess of IBM's conservative claims of 900 MB/sec. Aggregate restore performance of 1,965 MB/sec (7 TB/hour) was measured. ESG Lab's experience with a number of backup to disk solutions indicate that these are extremely impressive levels of performance—especially given the fact that de-duplication is being done in real-time over a logical pool of retained backup capacity of 20 PB, or more, in size. This is a crucial consideration for enterprise-class organizations that need to back up more than 10 TB or more of data nightly. One logical pool is easy to purchase, deploy and manage. One logical pool increases the cost and energy savings that can be achieved with de-duplication. And one logical pool with optional two-node clustered support increases the availability of backup and recovery services as validated by ESG Lab.

With high-speed inline data de-duplication that reduces the cost of retained disk capacity, optional two-node cluster support, up to one petabyte (1PB) of physical capacity, and VTL technology that is easy to deploy within an existing backup infrastructure, ESG Lab has confirmed that IBM System Storage TS7650G ProtecTIER De-duplication Gateways are designed to meet the disk-based data protection needs of the enterprise-class data center.

---

<sup>10</sup> Source: ESG Research Report, *Data Protection Trends*, January 2008

# Appendix

**TABLE 3. TEST CONFIGURATION**

ProtectTIER Software	Version 2.1
ProtectTIER Gateway Servers	Two IBM TS7650G gateways based on an IBM x3850 server each with a quad core 2.9 GHz Xeon processor, 32 GB of RAM and a pair of 4 Gbps FC HBA's.
ProtectTIER Storage	IBM DS8300 with 256 300 GB 15K RPM FC drives configured with 36 RAID-5 LUNs for 50 TB of backup data (20 7+1, 8 6+1) and four RAID-10 LUNs for meta data (4+4).
Fibre Channel Switch	QLogic SANbox 5200
Backup Software	Veritas NetBackup, Version 6, SP 5
Media Server #1	Four 1 GHz Opteron cores, 16 GB RAM, Red Hat Linux version 4 update 5, one four port Emulex 4 Gbps PCI-Express HBA, driver 8.0.16.27
Media Server #2	Four 2.3 GHz Xeon CPU cores, 2 GB RAM, Red Hat Linux version 4 update 5, one four port QLogic 4 Gbps PCI-Express HBA, driver 8.01.4-08
Media Server #3	Four 1 GHz Opteron CPU cores, 16 GB RAM, Red Hat Linux version 4 update 5, one four port Emulex 4 Gbps PCI-Express HBA, driver 8.0.16.27



20 Asylum Street  
 Milford, MA 01757  
 Tel: 508-482-0188  
 Fax: 508-482-0218

[www.enterprisestrategygroup.com](http://www.enterprisestrategygroup.com)