

Lab Validation Report

Riverbed Whitewater

Optimizing Data Protection to the Cloud

By Tony Palmer and Ginny Roth

November 2010

Contents

Introduction	3
Background.....	3
The Riverbed Whitewater Appliance.....	4
ESG Lab Validation	5
Getting Started	5
Cost Effective Data Protection.....	9
Data Assurance	14
Enterprise Class Performance and Scalability	16
ESG Lab Validation Highlights	18
Issues to Consider	18
The Bigger Truth	19
Appendix.....	20

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 Riverbed.

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

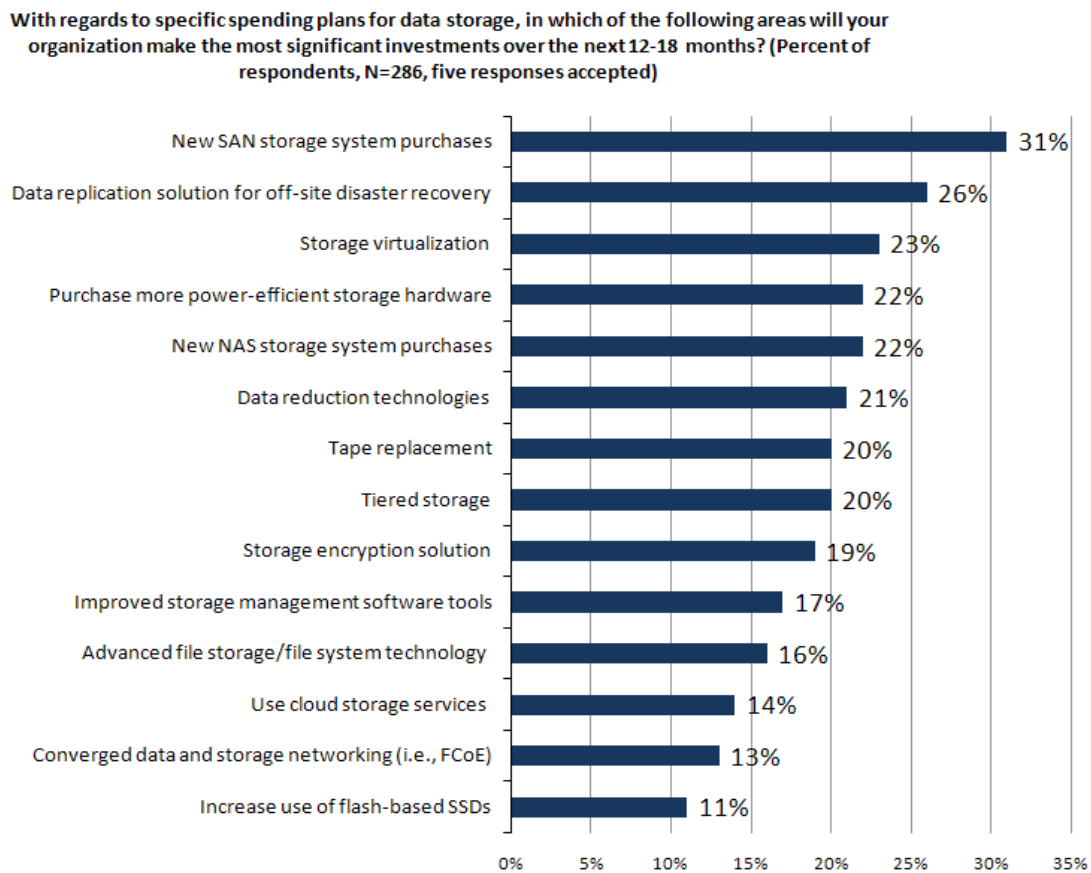
This ESG Lab report documents hands-on testing of the [Riverbed Whitewater](#) appliance with an emphasis on ease of use, cost effective recoverability, data assurance, performance, and scalability.

Background

Many organizations are grappling with the explosion in data growth in their IT environments. Forty-seven percent of midmarket respondents now report more than 10 TB of total data volume. Factor in annual double-digit growth and the sheer volume of data presents a challenge for data protection strategies. Improving disaster recovery capabilities continues to be a concern for many companies, as 35% of customers surveyed by ESG cited disaster recovery as their top area for data protection investment.¹

As shown in Figure 1, 26% of surveyed customers stated that data replication solutions for off-site disaster recovery would be a significant investment for their organizations the next 12-18 months.²

Figure 1. Most Significant Storage Investments



Source: Enterprise Strategy Group, 2010.

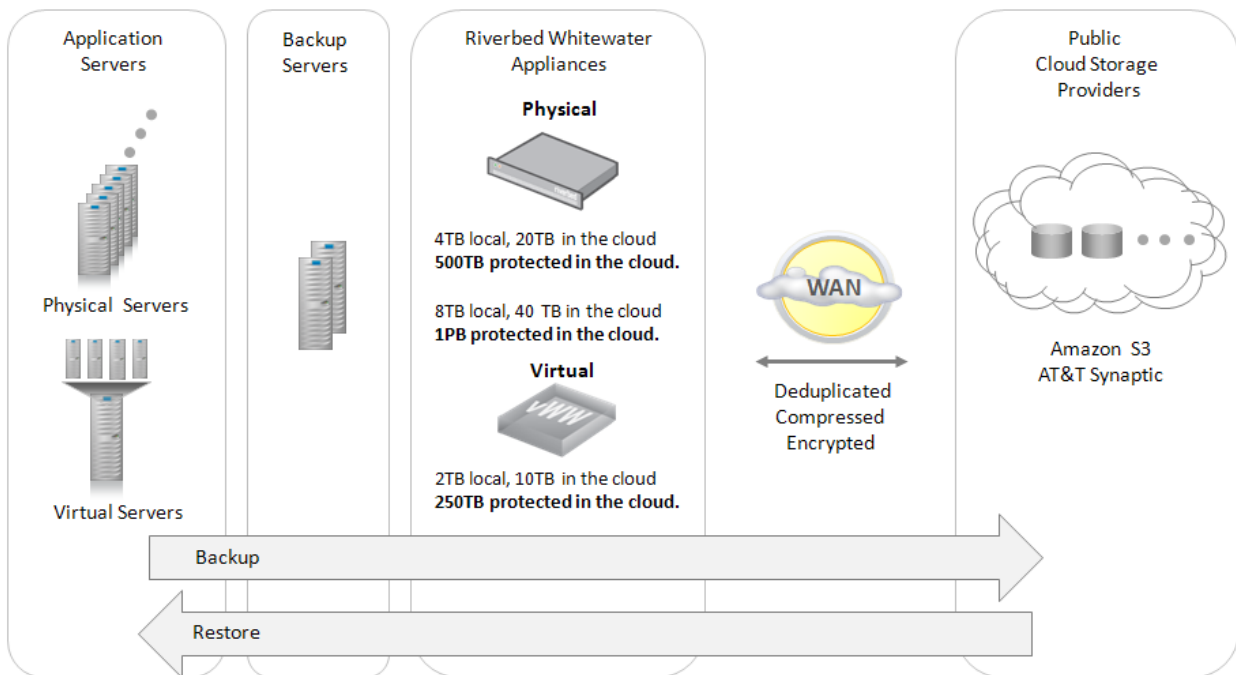
¹ Source: ESG Research Report, [2010 Data Protection Trends](#), April 2010

² Source: ESG Research Report, [2010 IT Spending Intentions Survey](#), January, 2010.

The Riverbed Whitewater Appliance

The Riverbed Whitewater cloud storage accelerator leverages the WAN optimization technology in existing Riverbed offerings to provide a complete data protection service to the cloud. The appliance-based solution is designed to integrate seamlessly with existing backup technologies and cloud storage provider APIs to provide rapid replication of data to the cloud for off-site storage and rapid retrieval for disaster recovery.

Figure 2. Riverbed Whitewater



The Riverbed Whitewater appliance provides several key components that help provide cost-effective data protection to the cloud.

- **Ease of Use:** Management of the appliance is achieved with a simple GUI interface accessed directly from the appliance.
- **Interoperability:** The appliance is designed to drop in to the customer's existing backup environment seamlessly, as a standard network-attached storage target.
- **WAN Optimization:** Using the compression and deduplication technologies that are the cornerstone of current Riverbed solutions, WAN optimization provides performance gains when replicating data to the cloud.
- **Security:** Using strong encryption standards, data is protected both at rest and in transit.
- **Stateless Appliance:** The appliance can rebuild recent backups locally, all of which can be retrieved from the cloud.
- **Near-infinite scalability:** With each appliance able to address 1 PB of backup capacity, even the largest enterprises can achieve offsite data protection for all their data with just a few appliances.

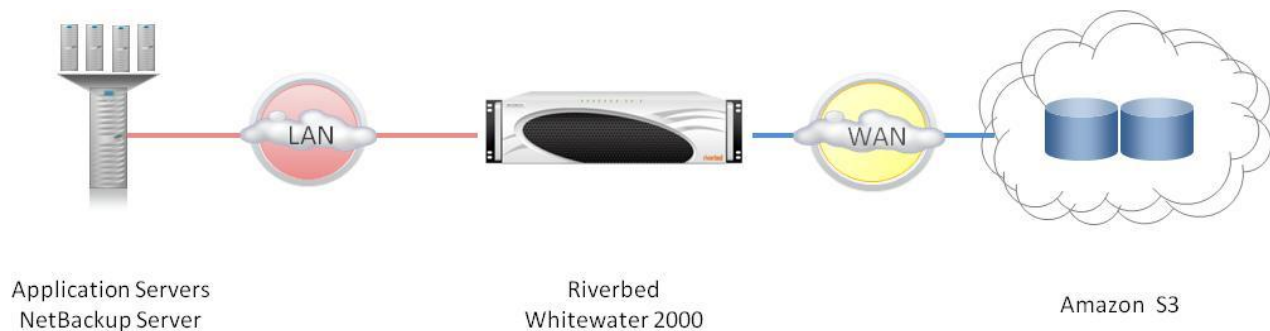
ESG Lab Validation

ESG Lab performed hands-on evaluation and testing of the Whitewater appliance at Riverbed's facilities in Sunnyvale, CA. Testing was designed to demonstrate ease of use and cost-effective backup and recovery, as well as enterprise class performance and scalability.

Getting Started

The Riverbed Whitewater appliance comes ready to deploy in a few simple steps. The appliance consists of two 1 GB management interfaces—one primary and one auxiliary—that are used not only for management functions, but also to replicate data to the cloud provider. Four 1 GB interfaces are used for data ingest of backups to the appliance. As shown in Figure 3, the test lab consisted of multiple virtual servers used as backup clients in addition to a NetBackup server that contained the management software for the backup application. The Whitewater appliance was installed in the network to serve as the target for backups, and also attached to the Internet for replication of data to the storage service provided by Amazon S3.

Figure 3. The Riverbed Whitewater Test Bed

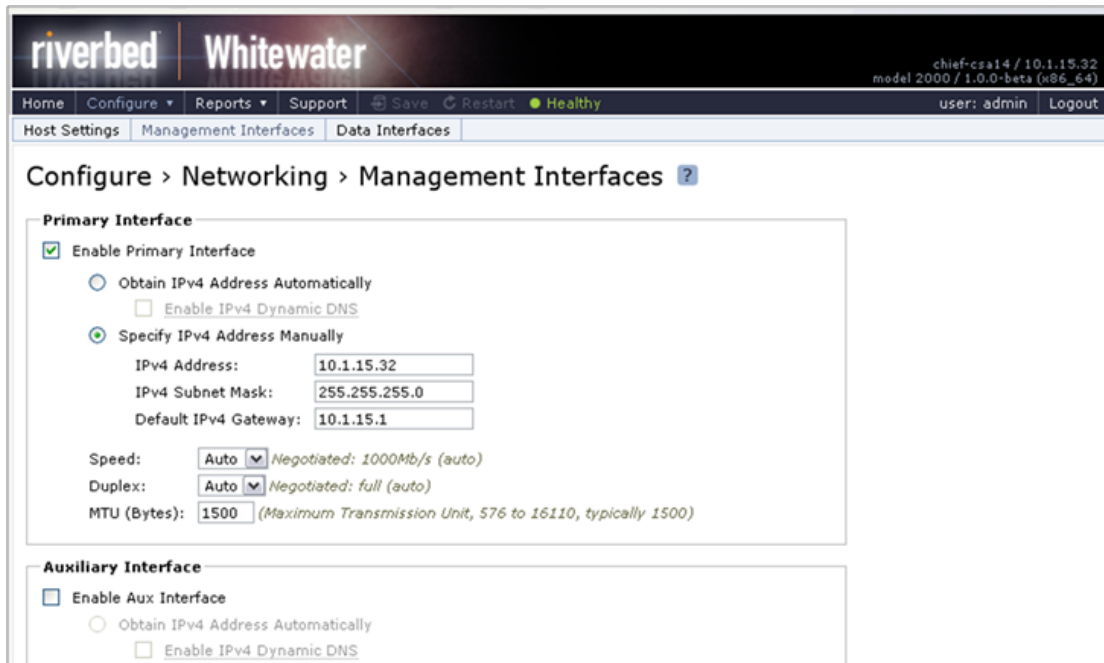


ESG Lab Testing

ESG Lab began testing configuration of the Whitewater appliance by starting the management interface from a client laptop, which is a web-based GUI accessed by entering the IP address of the management Ethernet interface on the appliance. A few simple steps are required to configure the appliance to accept backups and start data replication to the cloud. ESG Lab started by navigating to the Configure tab on the home page and choosing Networking to configure basic network settings to allow the device to talk on the existing network. From the Networking tab, ESG Lab accessed the Host Settings page and set the host name for the appliance, DNS servers, and NTP servers for time synchronization.

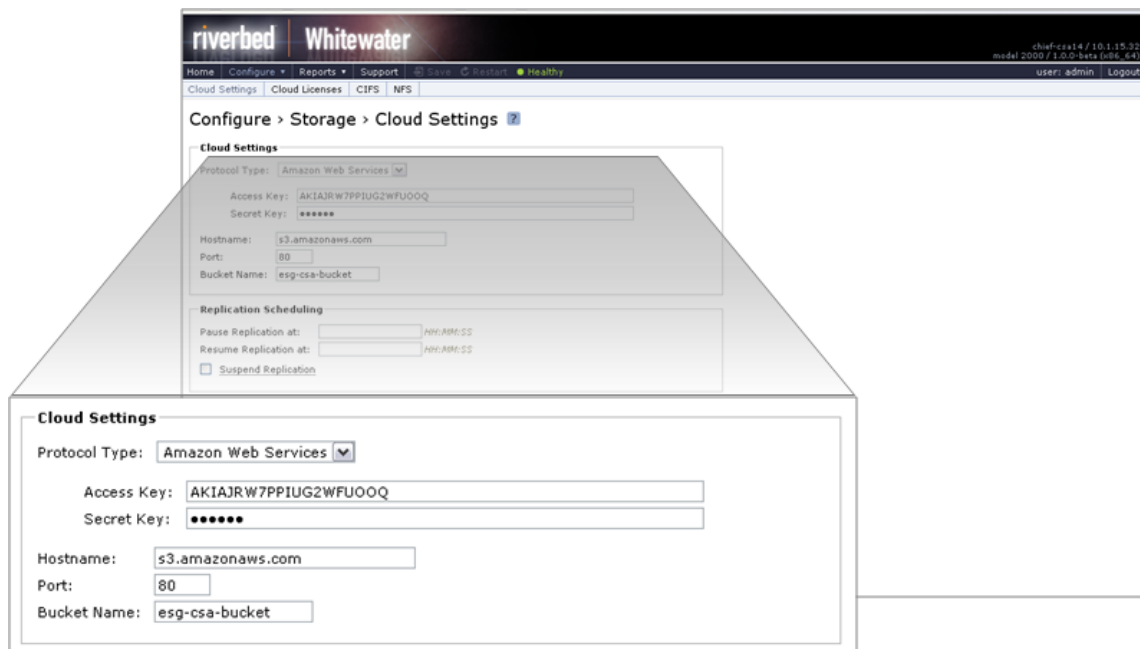
Next, ESG Lab chose the Data Interfaces page which shows the four interfaces used to receive backup data from sources on the network. For testing purposes, only one Ethernet interface was required, which ESG Lab configured with an internal IP address. The final step to complete the networking settings was to configure the management interface. After navigating to the Management Interfaces page, ESG Lab had the option to configure not only the primary interface, but an auxiliary if needed for replication to the cloud. As shown in Figure 4, ESG Lab configured an IP address for the primary interface only.

Figure 4. Configure Management Interface



ESG Lab next navigated to the Storage tab and chose the Cloud Settings page to configure the cloud service to which the appliance will replicate backup data. The service for this lab was Amazon S3, which was available in the drop-down box associated with the Protocol Type. Along with the host name and port for the cloud service, ESG Lab entered an Access Key and Secret Key provided by Amazon. These settings allow the appliance to authenticate to the Amazon service and enable data encryption. Finally, ESG Lab set the Bucket Name, which delineates the storage space that is set aside in the cloud, to esg-csa-bucket.

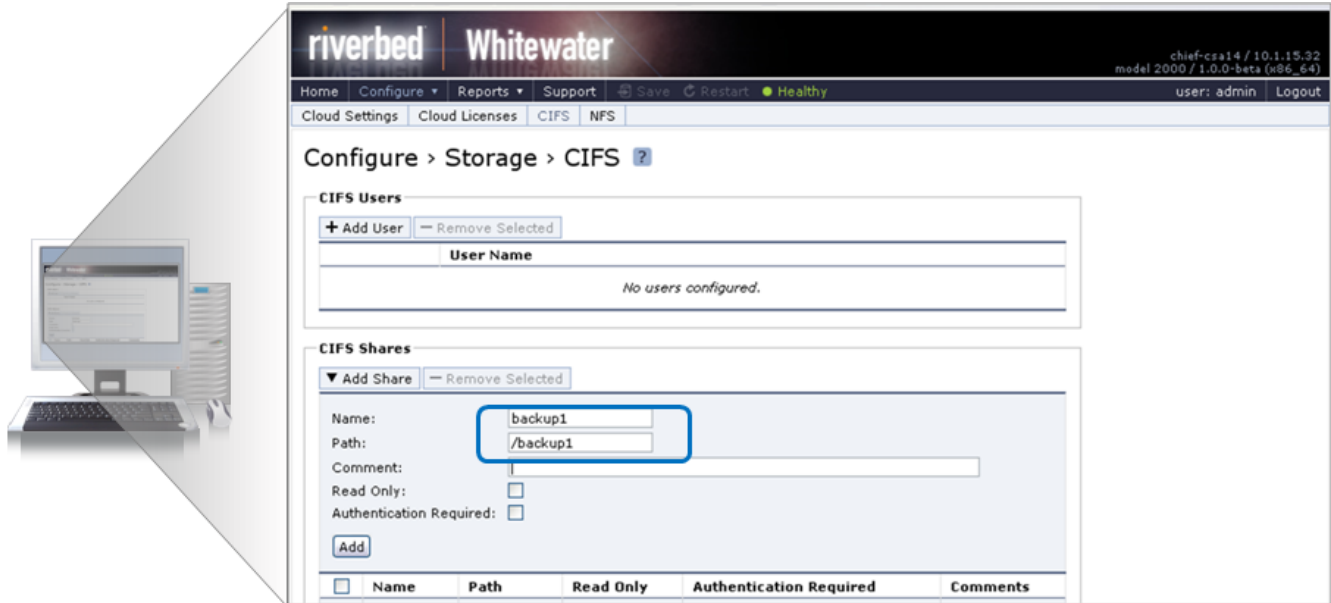
Figure 5. Configure Cloud Settings



At this point, ESG Lab exported the encryption key, which ensures that a customer can recover their data from the cloud should an appliance need to be replaced.

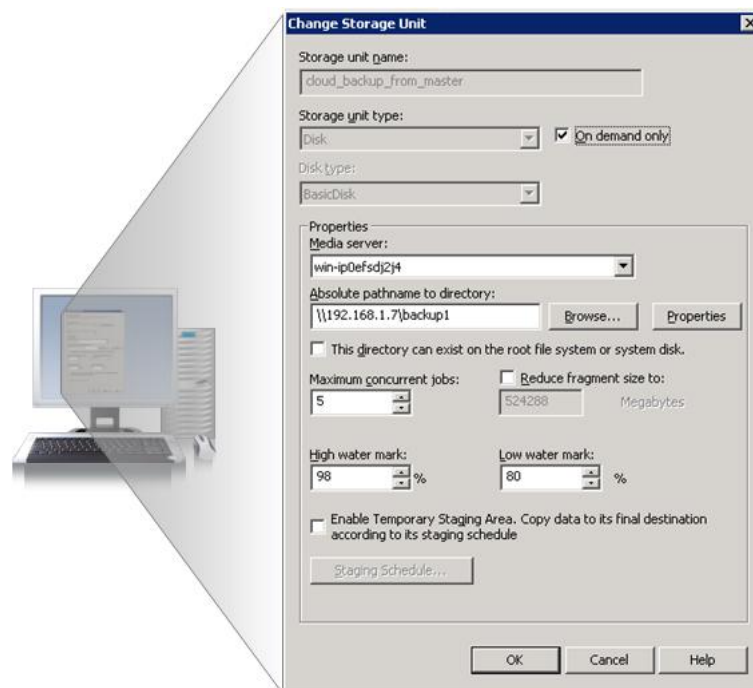
In order to create a target for backup jobs, a share needed to be created. Since the test backup software, NetBackup, was installed on a Windows server, ESG Lab chose to create a CIFS share to allow the Windows server to map a drive to the appliance. ESG Lab created a CIFS share simply by choosing the CIFS page and entering a new share name with a data path for mapping the share to, as seen in Figure 6.

Figure 6. Configure CIFS Share



Once the share was created, the appliance was available as a target for the backup software. ESG Lab tested the ability to use the appliance for backups by opening the NetBackup Administration Console and storage unit configuration. The “Absolute pathname to directory” field allows the administrator to set the target for the storage unit for each server in the data protection group. ESG Lab set the pathname to the CIFS share just created on the Whitewater appliance. NetBackup accepted the new mapping without any errors.

Figure 7. NetBackup Storage Unit Configuration



ESG Lab was able to configure a new Whitewater appliance to talk to the internal network and the cloud service and map a backup application to the appliance in eight simple steps. Setup—from first keystroke to a client backing up to the appliance—took approximately 10 minutes.

Why This Matters

Unrelenting capacity growth and shrinking backup windows are driving a growing number of businesses to deploy backup to disk solutions. As customers consider replacing tape backup infrastructure with disk and off-site storage for disaster recovery, they need a solution that can leverage their backup software and investments easily and seamlessly. Organizations of all sizes struggle with limited IT resources and need data protection solutions that can be integrated into their backup environment with minimal effort. Riverbed's Whitewater appliance is designed to drop into an existing environment with minimal disruption.

ESG Lab installed and configured a new Whitewater appliance to integrate into an existing NetBackup environment with just a few simple steps. Backup administrators don't have to learn anything new to leverage the cloud for data protection.

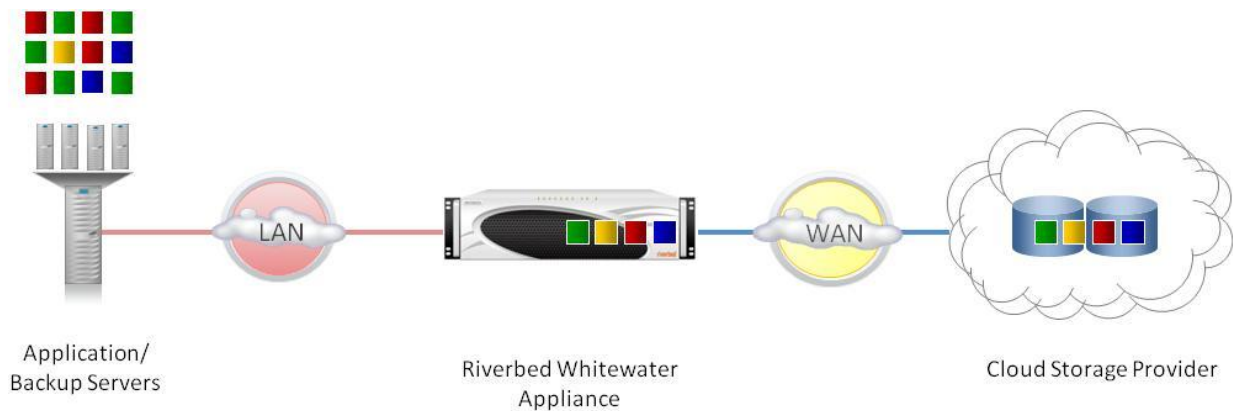
Cost Effective Data Protection

Riverbed provides comprehensive WAN optimization solutions helping organizations share applications and data across global wide-area networks. Riverbed's WAN optimization solutions have been proven in the field to give businesses an order-of-magnitude increase in the performance and value of their existing IT infrastructure and mission-critical applications, including file sharing, e-mail, backup, document management systems, IT tools, and ERP and CRM solutions.

With the Whitewater appliance, Riverbed has applied their WAN optimization technology to enable businesses to utilize public cloud infrastructure to store backup data for routine restores and disaster recovery with no negative impact to users.

Data deduplication 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 on disk, which, in turn can deliver significant economic benefits. Riverbed Whitewater appliances use an inline deduplication method, which eliminates duplicate data as it is being backed up. Whitewater appliances use an efficient algorithm designed to eliminate the risk of hash collisions, ensuring that deduplicated data is always available and correctly identified.

Figure 8. Riverbed Whitewater Data Deduplication



The level of disk capacity savings that can be achieved with deduplication varies according to the backup policy in use, the number of backup images retained on disk, (a.k.a. retention policy), whether the data is structured or unstructured, the rate at which data is changing, and the amount of duplicate data found within an organization. The ratio of capacity backed up versus capacity stored on disk after deduplication is generally known in the industry as the data deduplication factor. An ESG survey³ of current users of deduplication solutions indicates that factors between 10:1 and 20:1 are achieved by nearly 60% of users and 11% of users see rates above 20:1. A deduplication ratio of 10:1 reduces disk capacity requirements by 90%.

ESG Lab Testing

ESG Lab used NetBackup 6.5 to perform seven full backups and recorded the capacity backed up and stored on disk after each backup. An 11 GB collection of file data designed to mimic the contents of typical knowledge workers home directories (e.g., documents, spreadsheets, presentations, video and compressed files) was used during this round of ESG Lab testing. Daily changes were emulated using the `hpcreatedata` utility to add 1% of new data after each backup job had completed.⁴

³ Source: ESG Research Report, [2010 Data Protection Trends](#), April 2010

⁴ Details of the data type and distribution of files can be found in the Appendix.

Figure 9. Throughput From the Client to the Cloud

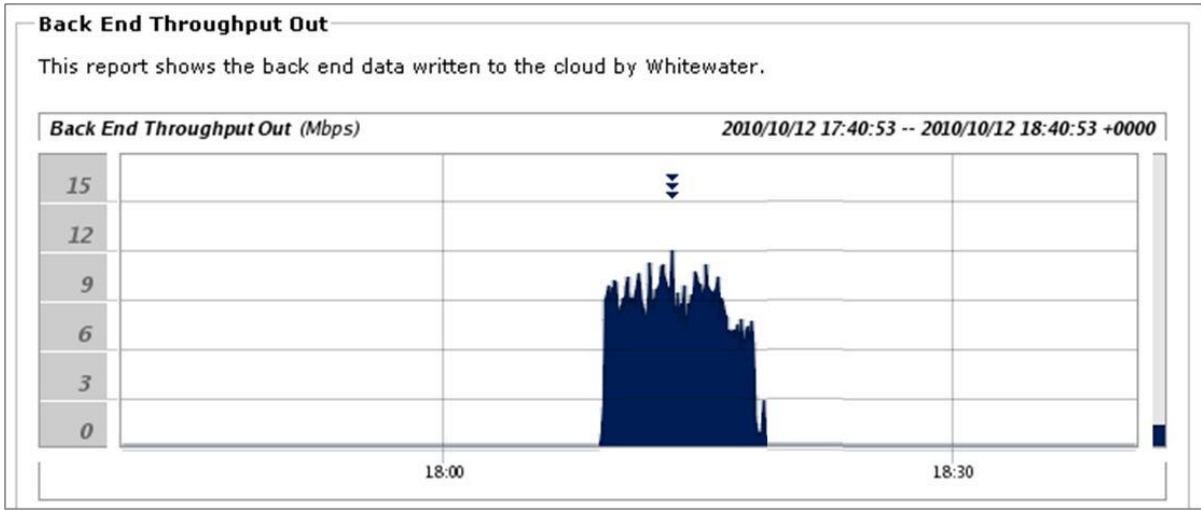


Figure 9 shows the Whitewater Storage Optimization reports for back end throughput to the cloud while the first full backup was being run to the appliance. While NetBackup was writing raw data to the appliance at an average of 55 MB/sec, and completed the backup in just over 3.5 minutes, the Whitewater appliance was writing deduplicated data to the cloud and completed transferring the entire 11GB backup image to the cloud in just over 10 minutes.

Figure 10. Deduplication Capacity Savings

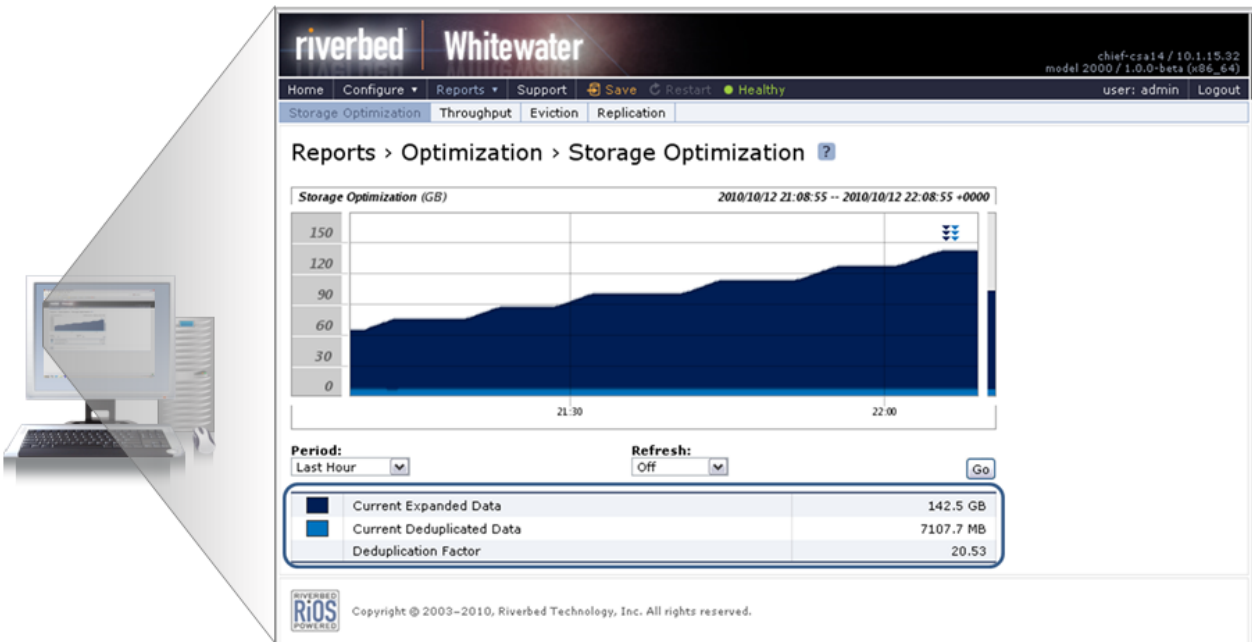
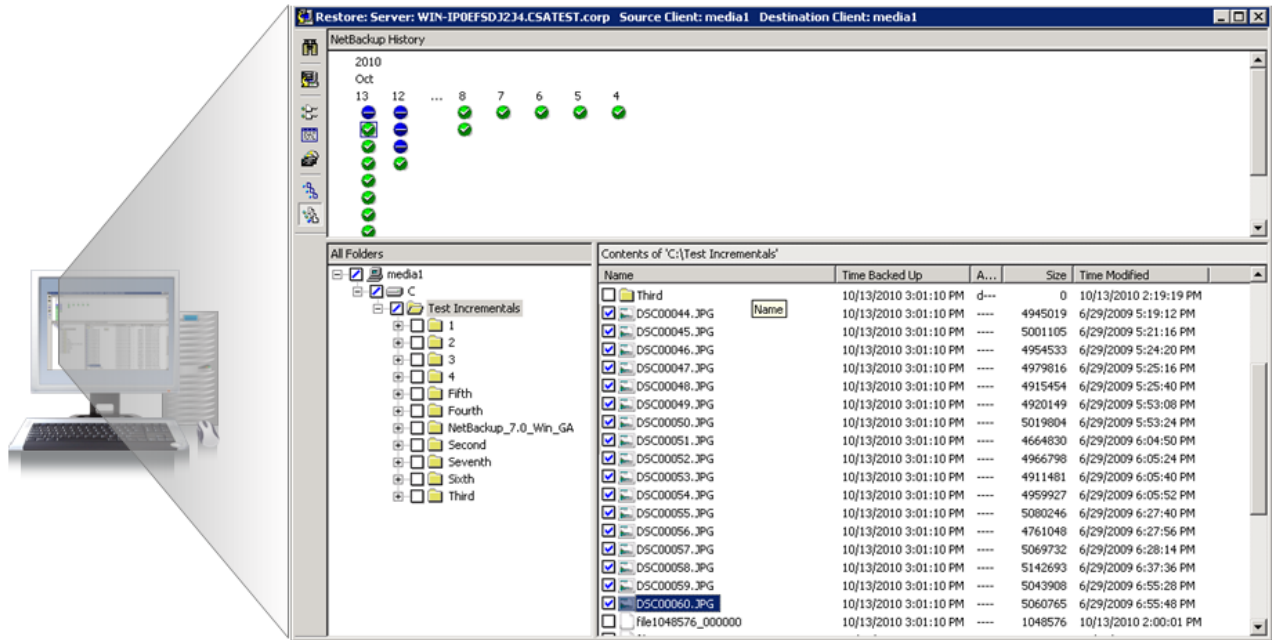


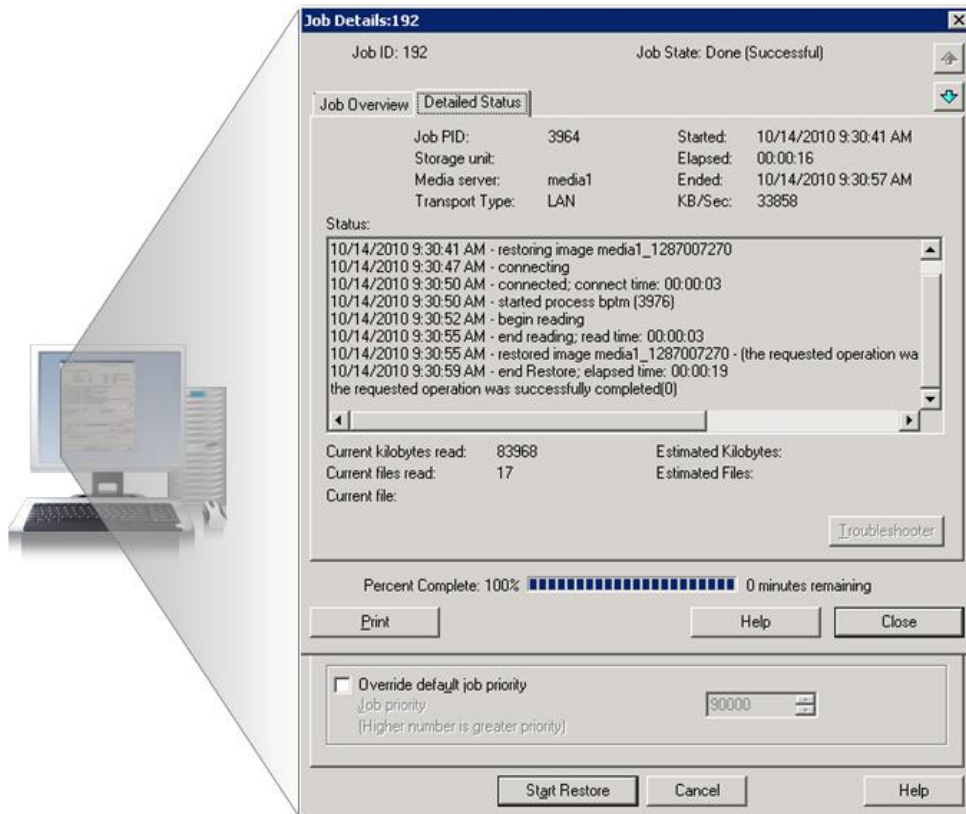
Figure 10 shows actual deduplication capacity savings over seven full backups. The capacity savings over seven full backups was greater than 20:1 or more than a 95% data reduction.

Figure 11. Selecting Files For Restore



Next, several files were deleted from the client and the files were restored using NetBackup, as seen in Figure 11. Figure 12 shows the NetBackup activity monitor for the restore. ESG Lab confirmed that the files were successfully restored by opening and viewing them on the client system.

Figure 12. Successful Restore



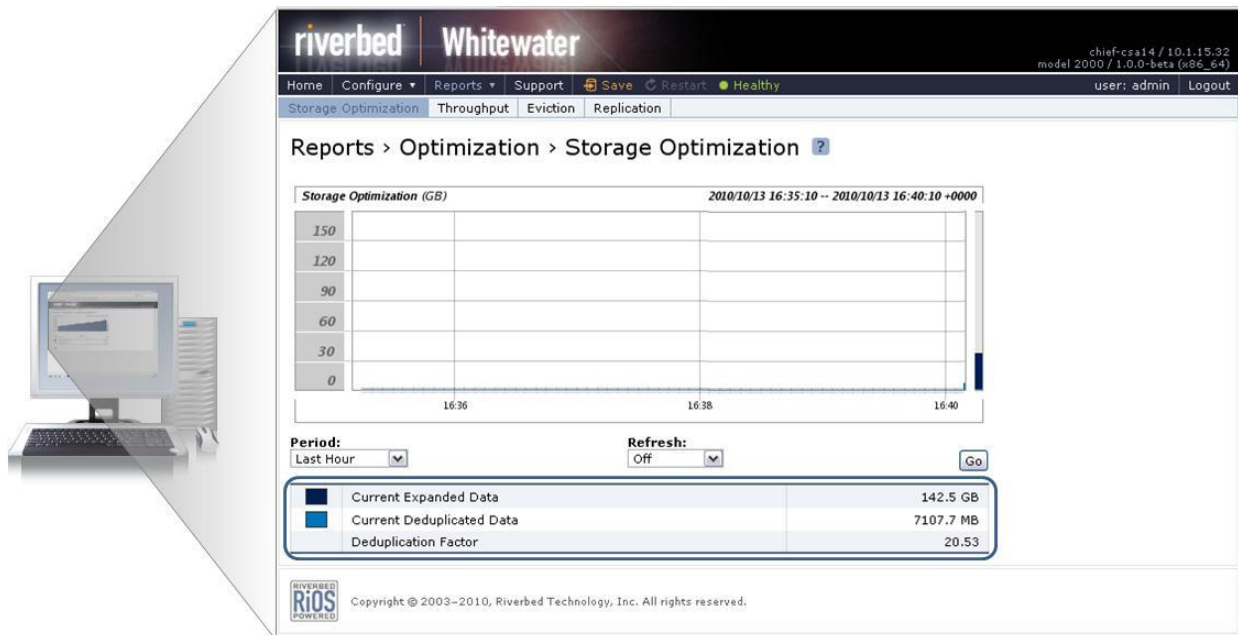
Next, ESG Lab simulated a disaster recovery where an entire site has gone offline and backup data needs to be recovered from the cloud. First, the Whitewater appliance was reset to factory defaults and the internal datastore was wiped clean. ESG Lab verified that all data had in fact been erased from the appliance.

Figure 13. Importing the Encryption Key



ESG Lab walked through the setup procedures described in the Getting Started section, including entering the cloud provider settings shown in Figure 5, and importing the encryption key, as seen in Figure 13. Figure 14 shows the storage optimization screen after the appliance was re-connected to the cloud.

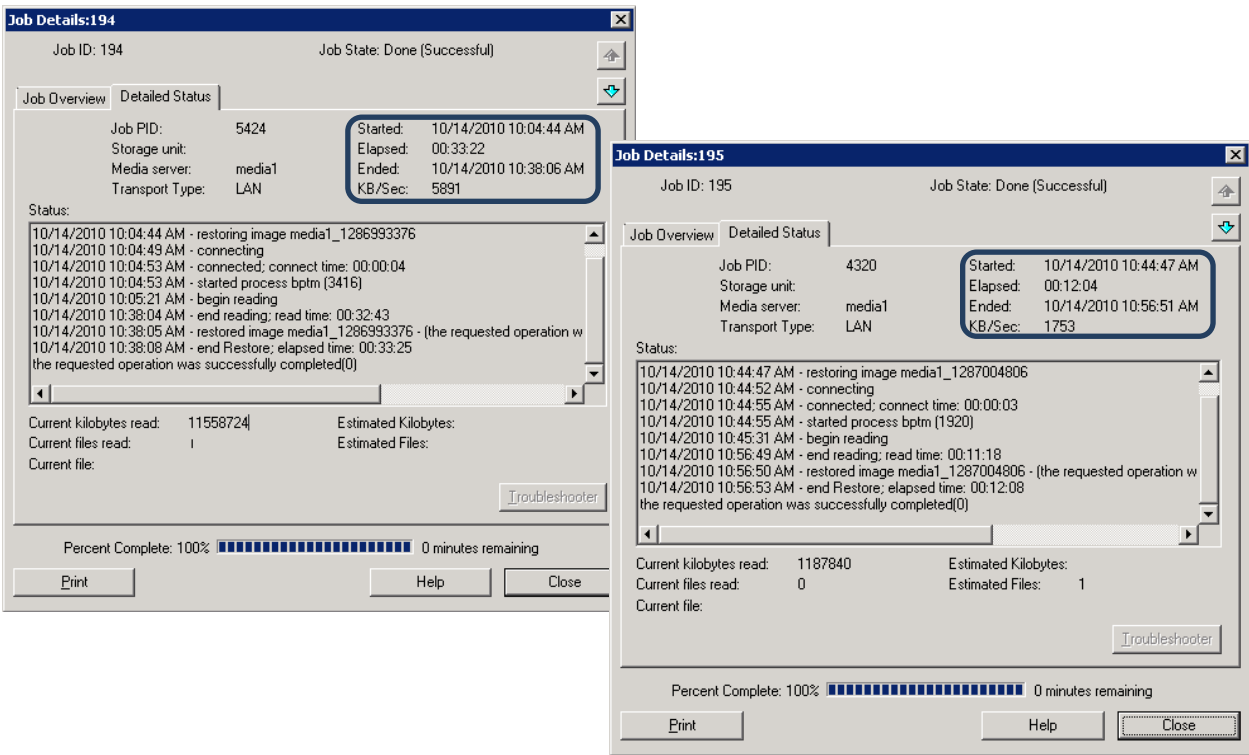
Figure 14. Recovering an Appliance



ESG Lab confirmed that, after connecting to the cloud and importing the encryption key, the appliance reported exactly the same amount of deduplicated and expanded data that it had before the reformat and reset, shown in Figure 10. Using the engineering CLI, ESG Lab confirmed that no data had yet been transferred back to the appliance, and everything that had been backed up in previous tests was now residing solely in the cloud.

Finally, ESG Lab selected two files to restore from the cloud, a 1 GB highly compressed ZIP file and an 11 GB uncompressed VMDK file. The NetBackup activity monitor for both restores is seen in Figure 15.

Figure 15. Restoring From the Cloud



The 1 GB ZIP file restored in approximately 12 minutes, with an average throughput of 17.5 Mb/sec to restore from the cloud. The 11GB VMDK file restored in about 33 minutes, with 55.9 Mb/sec of achieved bandwidth. It’s important to note that both of these files were restored from the cloud to an empty appliance, and the difference in observed speed of restore is due to Riverbed’s WAN optimization algorithm.

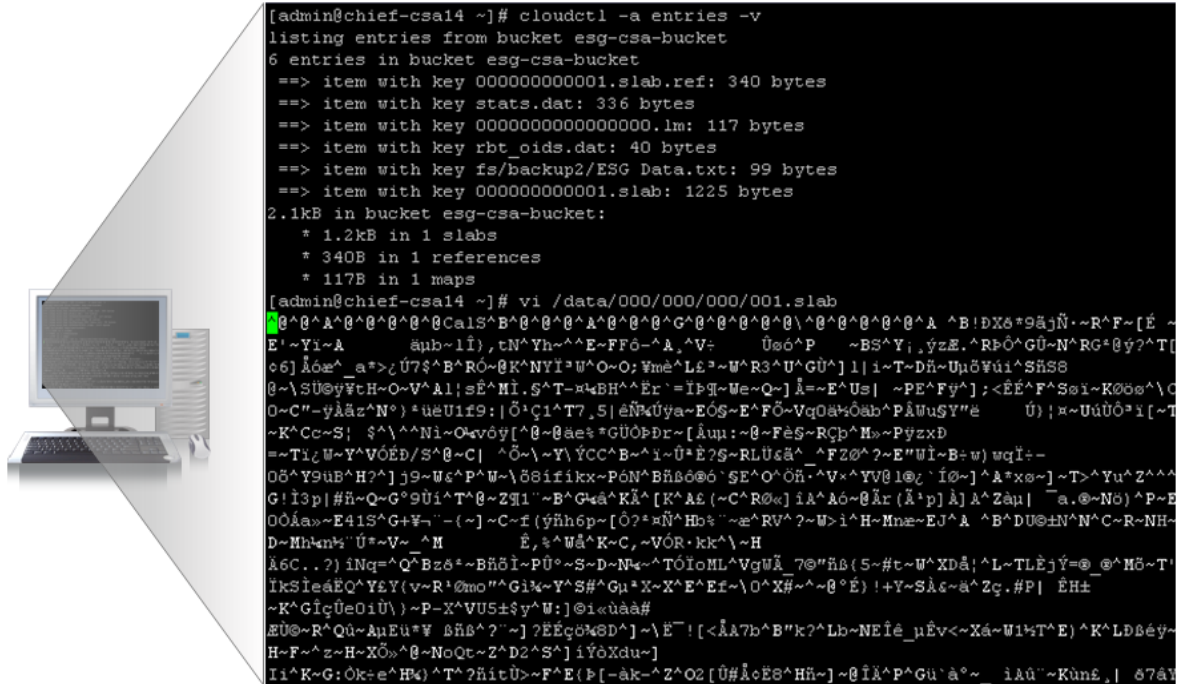
Why This Matters

ESG research indicates that nearly half (48%) of non-adopters cite cost as the leading reason they have not yet embraced a disk-based backup strategy.⁵ Data deduplication with replication to the cloud addresses the cost issue by reducing the disk capacity required to maintain multiple generations of backup data for quick and reliable restores and eliminating secondary target appliances. ESG Lab has confirmed that Riverbed Whitewater inline deduplication and WAN optimization can be used to reduce retained backup to disk capacity requirements by a factor of 20 to 1 or higher while providing local disk performance levels for backup, restore, and disaster recovery from the cloud.

⁵ Source: ESG Research Report, [2010 Data Protection Trends](#), April 2010

After the file was copied, ESG Lab connected directly to the Whitewater appliance via an engineering CLI to examine the file. When ESG Lab examined the contents using the vi editor, none of the clear text was visible, as illustrated in Figure 18.

Figure 18. Encrypted File



Why This Matters

Security is often cited as the biggest concern when moving data and applications to the cloud. Not only are companies concerned about data in transit, but are how that data is protected at the cloud provider’s location is critical. Riverbed’s Whitewater appliance provides strong data encryption in transit over the Internet, and at rest both in the appliance and in the cloud provider’s storage.

ESG Lab confirmed that data was encrypted by the appliance before being written to the data store or replicated to the cloud, automatically, with no administrator actions required.

Enterprise Class Performance and Scalability

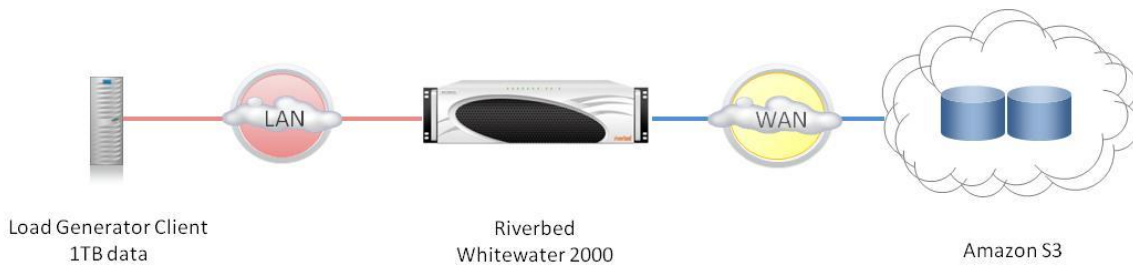
Midrange and enterprise-class data centers backing up multiple terabytes of data nightly are presented with a number of potentially conflicting challenges. The backup target needs to be fast to avoid a missed backup window, while at the same time able to store and retain hundreds of terabytes of backup data for quick and reliable restores. And the cost of the system has to fit within stagnant or shrinking budgets. While inline deduplication can drastically reduce the amount of physical capacity required (and hence the cost), it still needs to support a large pool of capacity to fully realize the benefits of deduplication.

The deduplication and optimization technology at the heart of the Riverbed Whitewater solution was designed to meet the performance and capacity needs of enterprise-class data center environments, while leveraging cost-effective cloud storage to meet both capacity and offsite requirements. The Whitewater appliance uses the same robust optimization technology originally developed for Riverbed’s Steelhead WAN optimization solutions. Whitewater appliances implement deduplication across local physical storage and cloud-based capacity to offer up to one petabyte (1 PB) of retained backup capacity.

ESG Lab Testing

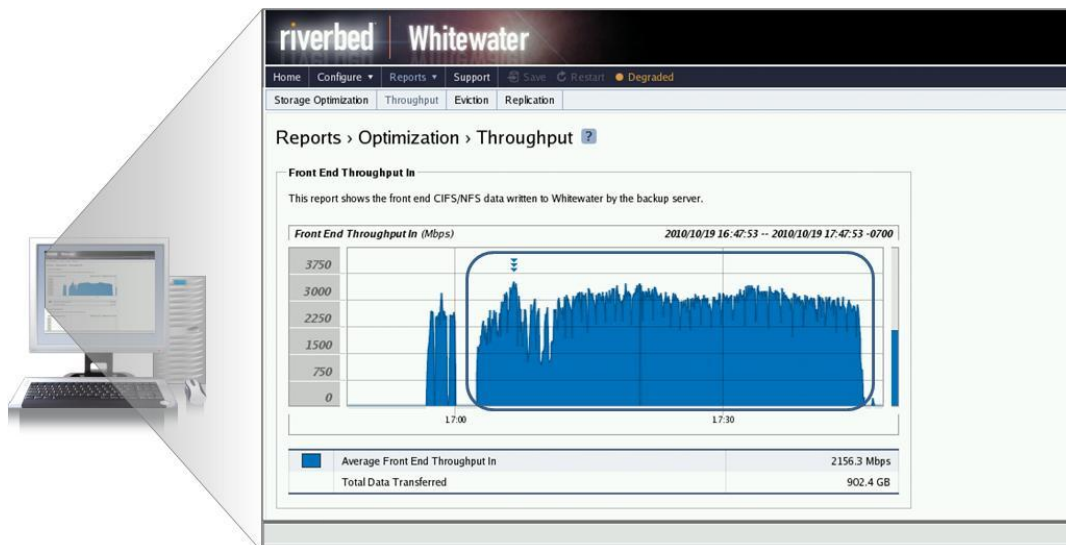
To test the performance of the Whitewater 2000 appliance, a single Intel Xeon server running CentOS 64-bit Linux OS used Cloud Backup Test (CBT), a Riverbed developed tool, to create 341 3GB data files with 3:1 compression and 3:1 deduplication factors. Multiple process loops were run on each of four 1Gbps NICs in the load generator to use smbclient to write a randomly selected 3GB file to a CIFS share on the Whitewater appliance.

Figure 19. The Whitewater 2000 Performance Test Bed



The Whitewater front end throughput report, shown in Figure 20, displayed the rate of data being processed by the appliance as averaging nearly 3,000 Mbps over the course of the 1 hour test or more than 1 TB per hour.

Figure 20. 1 TB per Hour Sustained Ingest



ESG Lab validated the reported throughput using the Linux iftop utility to examine the throughput on each Gigabit ethernet interface writing data to the Whitewater appliance. The results are detailed in Table 1.

Table 1: Client Throughput by Interface

Interface	Source IP address	Target IP address	Transmit rate (Mbps)	Transmit Rate (GB/hour)
Eth0	10.12.2.46	10.12.2.92	725	261
Eth1	10.12.2.56	10.12.2.93	721	259.6
Eth2	10.12.2.66	10.12.2.91	722	259.9
Eth3	10.12.2.76	10.12.2.89	724	260.6
Totals			2,891	1,041.1

What the Numbers Mean

- A single Whitewater appliance tested by ESG Lab was able to back up multiple streams of moderately compressible and deduplicatable data at a sustained rate of 1TB per hour.
- Riverbed's inline deduplication and WAN optimization enabled data to be written to the cloud at 10x wire speed.

Why This Matters

ESG research⁶ has found that the top challenges enterprises report with their data protection processes and technologies is keeping pace with the capacity of data to protect, the need to reduce backup times, and the costs associated with backup solutions. With tape media budgets running into the tens of millions of dollars for many enterprise-class organizations, managers are looking for ways to reduce backup and DR costs without putting their businesses at increased risk. By combining deduplication and WAN-optimized cloud storage into one package, Riverbed can significantly reduce the cost of data protection by reducing the amount of local and cloud storage capacity and bandwidth that is ultimately required during the backup and DR process.

ESG Lab has confirmed via hands-on testing that the Whitewater appliance can sustain an ingest rate of 1 TB per hour while deduplicating and replicating to the cloud. Based on experience testing a number of backup to disk solutions, ESG Lab is extremely impressed with the performance of the Whitewater appliance— especially given the fact that it performs inline deduplication to reduce capacity requirements and WAN-optimized cloud replication to provide offsite data protection while servicing backups at 1 TB per hour.

⁶ Source: ESG Research Report, [2010 Data Protection Trends](#), April 2010

ESG Lab Validation Highlights

- ☑ ESG Lab installed and configured a new Whitewater appliance to integrate into an existing NetBackup environment with just a few simple steps, in just 10 minutes.
- ☑ Riverbed Whitewater inline deduplication and WAN optimization were used to reduce retained backup to disk capacity requirements by a factor of more than 20 to 1 while providing local disk performance levels for backup, restore, and recoveries from the cloud.
- ☑ Riverbed's Whitewater appliance provided strong data encryption in transit across the internet, and at rest both in the appliance and in the cloud provider's storage.
- ☑ ESG Lab has confirmed via hands-on testing that the Whitewater appliance can sustain an ingest rate of 1 TB per hour while deduplicating inline and replicating to the cloud for offsite protection.

Issues to Consider

- ☑ Access to NFS exports on the appliance can be restricted by IP address, but currently no user login is required for access. CIFS shares support a user login but cannot be restricted by shares. One login can access all shares mapped on the appliance. User access to both NFS and CIFS are being addressed in later versions of the product.
- ☑ Recovery of an appliance pulls back the namespace for customer data in the cloud. While customers can browse the Whitewater share using a graphical user interface or a command line interface, slower response times may be experienced while browsing the Whitewater remote share in comparison to a remote share provided by a NAS filer, as the product is primarily optimized for backup and recovery operations.

The Bigger Truth

Data growth over the years has increased at a rapid pace and shows no signs of slowing down. This creates a special challenge for IT administrators responsible for protecting critical company assets. The cost to provide adequate capital resources for data protection keeps rising and the window for backup and recovery of important data continues to shrink, creating a strain on IT budgets. Customers are now starting to see limitations with current strategies for disaster recovery and are looking for cost-effective solutions. In fact, more than half of companies surveyed by ESG cited keeping pace with capacity of data to protect and the need to reduce backup and recovery times as top concerns for overall data protection strategies.⁷

As customers start to look at cloud storage services as a low cost alternative to maintaining their own storage infrastructure, there are clear benefits offered by the Riverbed solution. Off-site storage of company data in the cloud for disaster recovery using strong, standards-based encryption is attractive since it doesn't require maintenance of hardware or data center footprint, and is only required to be accessed for retrieval of lost data.

Cloud storage services optimized by Riverbed Whitewater appliances provide a cost-effective option for customers looking for a new strategy to replace expensive, dedicated disaster recovery technology for businesses sensitive to capital outlays associated with off-site storage.

Riverbed provides comprehensive WAN optimization solutions helping organizations share applications and data across global wide-area networks. Riverbed's WAN optimization solutions have been proven in the field to give businesses order-of-magnitude increases in the performance and value of their existing IT infrastructure and mission-critical applications, including file sharing, email, backup, document management systems, IT tools, and ERP and CRM solutions.

Riverbed has applied their field-proven WAN optimization technology to provide similar performance gains for data protection extended to the cloud. Achieving data reduction of more than 20 to 1 with inline deduplication, the Whitewater appliance not only reduces the amount of bandwidth needed to replicate to off-site storage, but shrinks the data capacity requirements to store the data in the cloud, maximizing investments with cloud storage providers. Whitewater bridges the gap for off-site data storage by providing an easy to deploy appliance that integrates with existing customer investments in backup software and secures that data as it leaves the customer premises, easing a primary concern of customers for data stored in the cloud.

With high-speed inline data deduplication that reduces the cost of retained disk capacity, WAN optimization capabilities that extend deduplication to offsite storage, and a drop in place appliance that is easy to deploy within an existing backup infrastructure, ESG Lab has confirmed that Riverbed's Whitewater cloud storage appliance is designed to meet the local and off-site data protection needs of the enterprise-class data center, for businesses of all sizes.

⁷ Source: ESG Research Report, [2010 Data Protection Trends](#), April 2010

Appendix

Table 2. ESG Lab Test Bed

Hardware	
Whitewater 2000 Cloud Storage Appliance	8TB Cache CPU: 2 x Six-Core AMD Opteron RAM: 32 GB OS: Linux NIC: 4 x 1Gbps Ethernet for data 2 x 1Gbps Ethernet for management/replication
Load Generator Client	CPU: 2 x Quad core Intel Xeon E5506 2.13GHz RAM: 4 GB OS: Linux CentOS 2.6.18-194.17.1.el5 #1 SMP x86_64 NIC: 4 x 1Gbps Ethernet for data 2 x 1Gbps Ethernet for management Network Protocol: CIFS using Samba: 3.3.9
NetBackup Client and Media Server	1x Dual core Intel Xeon CPU RAM: 4GB OS: Windows Server 2008 NIC: 2 x 1Gbps Ethernet for data and management
Software	
Cloud Backup Test (CBT)	Created 341 3GB files with 3X compression and 3X deduplication factors using consecutive integer seed values
hpcreatedata	Version 1.2.3 Created 320 files with equal distribution from 64KB to 8MB using 2X compression factor
NetBackup	Version 6.5
Cloud Services	
Amazon S3	1TB Reduced Redundancy Storage (Designed for 99.99% Durability)



Enterprise Strategy Group | **Getting to the bigger truth.**