Disaster Recovery White Paper: Reducing the Bandwidth to Keep Remote Sites Constantly Up-to-date

Storage Virtualization and WAN optimization help organizations achieve more stringent Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) without incurring the expense of higher speed inter-site lines.

Author: Mike Beevor – DataCore Software
Co-Authors: Simon Birbeck & Richard O'Connor – Waterstons Ltd.

Although the consolidation from discrete physical servers to virtual servers has made it possible for many organizations to configure an affordable remote Disaster Recovery (DR) site, the inter-site traffic to keep the DR location current soon grows to be a problem. The volume of updates from newly created virtual machines at the primary site can rapidly climb to the point where they exceed the line's capacity to transmit them. The overload causes the DR site to fall behind more and more, exposing companies to unacceptable data loss in the event of a disaster.

Some customers' first thoughts were to upgrade to higher speed lines. However, after seeing how much more that would cost in infrastructure and recurring leased line rentals, there is often the need to investigate other less pricey alternatives to attain a satisfactory Recovery Point Objective (RPO).

In this case study, the IT consultants at W@terstons pondered the scenario and proposed to address these requirements through a clever combination of storage virtualization and WAN optimization. They employed the built-in remote replication function in DataCore SANmelody™ software to relay changes in virtual disk blocks from the primary site to the DR site over an existing 2MBps IP WAN. These inter-site transmissions were compressed, cached and optimized inline by inserting Riverbed Steelhead network appliances on either side of the link. The solution succeeded in meeting the firm's disaster recovery objectives and proved much more cost-effective than upgrading to higher bandwidth service. Details of the configuration, observations and analysis are provided below.

Remote Replication via DataCore Storage Virtualization Software

DataCore offers two replication mechanisms to keep DR sites up to date, depending on distance and available bandwidth. Both functions transparently replicate writes to selected virtual disks issued by any application, operating system or hypervisor.

When sites are separated by less than 10km using high-speed connections like Fibre Channel, the synchronous mirroring option keeps local and remote disks in lock-step. Essentially, every I/O to a designated disk is replicated in real-time to the remote site before it is acknowledged as complete to the host.

As distances stretch further or budgets dictate use of lower costs WANs, synchronously replicated I/Os experience higher latencies causing applications to time out, or response to become inordinately long. In this scenario, the use of DataCore asynchronous remote replication over an IP WAN is necessary.

With asynchronous replication, the primary site acknowledges updates to applications immediately and then makes a best effort to deliver them to the remote site over the available bandwidth.

There is always the risk of some data loss when restoring operations from the DR location because the primary may not have sent all of its changes prior to the disaster. But it is far more effective than relying on shipping frequent backups to restore the DR site and allows for a greater degree of granularity and reduction in RTO.

Terminology

The “source volume” is the local production volume an application server writes to and reads from. With Asynchronous Replication it is understood that the bandwidth between the two sites is limited, and so writes to the local source volume are acknowledged immediately, and replicated when bandwidth permits, or at a specified instant in time to the “destination” volume at the DR facility. The destination volume is usually always in a "catch- up" mode, anywhere from a few writes to a few gigabytes worth of writes behind the source volume. How far the destination falls out of synch (or lags behind) the source volume is determined by the quantity of changes occurring on the source volume, the bandwidth of the link used for the replication between sites, network QoS (Quality of Service) policies and/or replication schedules or administrator imposed throttles.

Clearly, provisions must be made to keep track of changes to the source volume in order to replicate those changes and bring the source and destination as close to being synchronous as time and bandwidth permit. There are a number of ways to do this, ranging from marking changed blocks that need to be replicated, to buffering the changes in a reserved storage area. DataCore implements this feature in the AIM (“Asynchronous IP Mirroring”) option for their SANmelody and SANsymphony™ products. For an in-depth review of asynchronous replication and DR, please refer to the DataCore white paper on “DR and Asynchronous Replication”.

WAN Optimization from Riverbed Steelhead Appliances

Riverbed’s Steelhead WAN optimization appliances are designed to provide the highest performance for IP-based traffic between sites, while at the same time, making it easy to deploy, manage and monitor Wide-Area Data Services (WDS). The underlying software known as RiOS provides an integrated framework for data reduction, TCP optimization, application level optimization, remote-office file services, and management services to provide a comprehensive solution for enterprise WDS. RiOS scales across a range of applications and network topologies. In this case study, one appliance was located at the primary site and another at the DR site at the WAN termination points.

RiOS overcomes the chattiness of transport protocols through transport streamlining. Transport Streamlining is a set of features that reduce the number of roundtrips necessary to transport information across the WAN while maintaining the reliability and resiliency of the transport. This is accomplished through a combination of window scaling, intelligent repacking of payloads, connection management and other protocol optimization techniques. RiOS accomplishes these improvements while still maintaining TCP as a transport protocol. As a result RiOS adapts to network conditions on the fly, responding appropriately to events such as congestion or packet loss.

RiOS also provides Application Streamlining allowing additional Layer 7 acceleration to important (but poorly behaved) protocols through transaction prediction and pre-population features. The application streamlining modules provide additional performance improvement for applications built on particular facilities such as Microsoft Windows file systems better known as Common Internet File Systems or CIFS. DataCore uses CIFS to transfer I/Os buffered at the source. Application streamlining modules eliminate roundtrips that would have been generated by the application protocol. Reducing roundtrips may be necessary even with a very efficient implementation of TCP, because otherwise the inefficiency of the application-level protocol can overwhelm any improvements made by the transport layer. It eliminates up to 98% of the roundtrips taken by specific applications, delivering a significant improvement in throughput, in addition to what data streamlining and transport streamlining already provide.

With particular relation to CIFS, RiOS contains approximately a dozen CIFS optimizations for various operations such as file sharing, folder browsing, accessing files from within other applications, and more. RiOS’ out-of-the-box acceleration of back-up and replication operations already generated significant performance gains for data transfer jobs at this location. The traffic recognition capability identifies a large-scale data transfer flowing through a Steelhead Appliance and applies some specific system optimizations to enhance the throughput and handling of high rate, high volume back-up data sets. These additional optimizations improve disk utilization, while also dramatically applying data reduction and compression algorithms. The resulting throughput enhancement further reduces the lag between the production and DR sites.

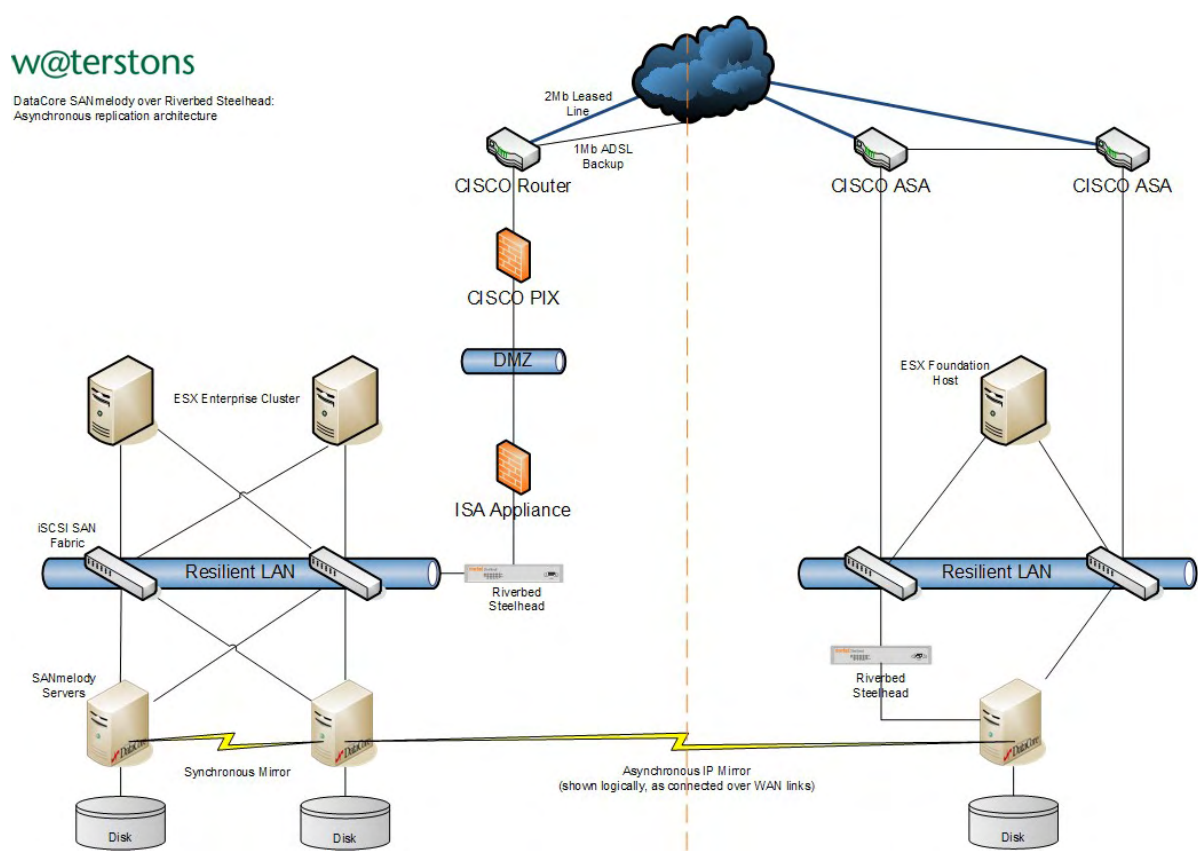
Customer's Environment

All tests were performed in the customer's live environment, with production data operating across the two sites. The customer is a syndicate of salvage companies that specialize in the recovery and safe disposal of vehicles that have been written off in an accident or similar incident. The customer's primary competitive advantage is their ability to cost effectively manage the complete cycle of disposal on behalf of their insurance company clients and to furnish them with comprehensive reports as to progress and overall performance.

Their elevated DR concern arose after undergoing a business continuity analysis, in which recommendations were made to enhance the security and resilience of the technical infrastructure. After many of these changes were made, they experienced performance issues with respect to network and application responsiveness.
For example, client/server response times were affected by scheduled backup jobs which took a long time to complete and negatively impacted their prime business hours. Additionally, the existing Microsoft cluster, hosting SQL and Exchange 2003 servers, provided business continuity, but at a high complexity and cost inefficiency with regards to administrative ownership.

The existing BI (Business Intelligence) information is housed on the clustered SQL server and occupies approximately 60GB. Running the SQL Analysis Services cube takes a number of hours to complete and historical data requires a large degree of management upon its retirement. The customer has a resilient communications environment between the primary and DR site using a 2Mbps Megastream primary circuit and 1Mbps ADSL failover; both routed through a Cisco 2611 router. This is routed from the local LAN via a MSA2020 ISA appliance to a DMZ, then through a Cisco PIX506E to the external facing Cisco 2611.

Current Topology Overview



Methodology

Since the environment was live and production data was being transferred whilst our testing was taking place, it became necessary to ensure that conditions remained identical for all transfers. This was felt to be a important in this scenario since one of the aims of the whitepaper was to demonstrate the improvement in performance in a live environment.

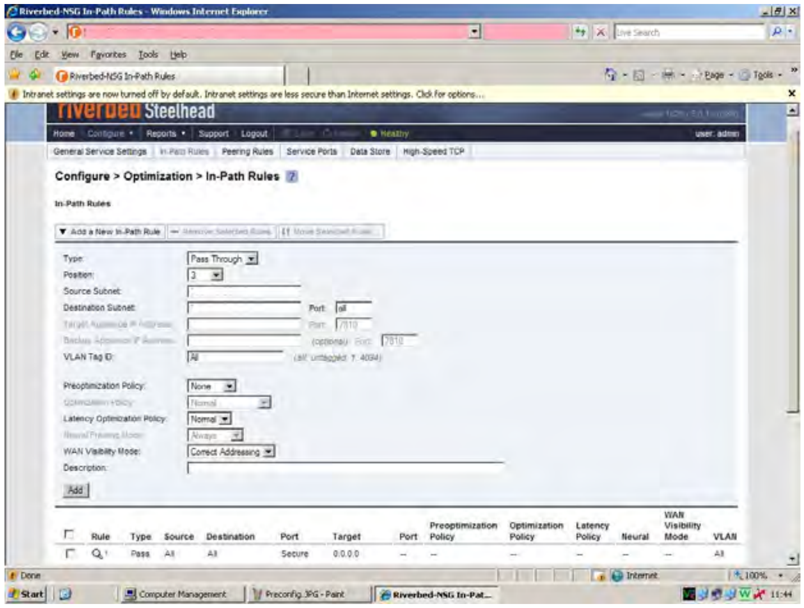
In order to ensure a measure of scientific control, the following methodology was introduced:

- Capture screenshots of current data being transferred across the WAN through DataCore remote replication to record transfer rates and latency times.
- Create 2 Virtual Servers running Windows Server 2003 Enterprise Edition with a size of approximately 12GB.
- Begin replicating these Virtual Machines (VMs) across the link, utilizing a pass-through rule within Riverbed to ensure there is no optimization of the data stream.
- Record a period of screenshots for this period to track transfer rates.
- Create a 3rd Virtual Machine running Windows Server 2003 Enterprise Edition with a size of approximately 12GB
- Begin replicating the newly created Virtual Machine
- Remove the pass-through rule on the Riverbed appliance to optimize the remote replication over the WAN
- Take screen shots of the transfer behaviour over a similar period of time and perform a comparison of throughput rates and latency times.
- Document any change in performance.

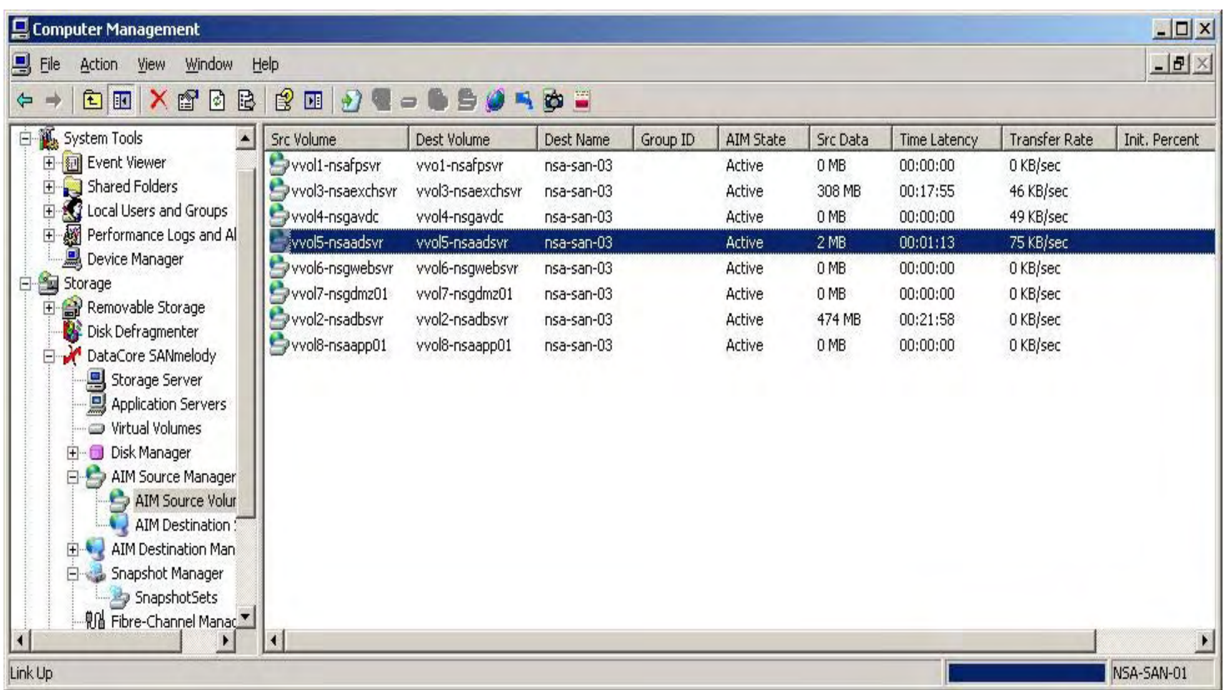
It was important as part of the testing that the Virtual Machines were created from the same clone template, on the same host, although within separate virtual machines, to prevent serious degradation of the end user experience.

Results & Explanations

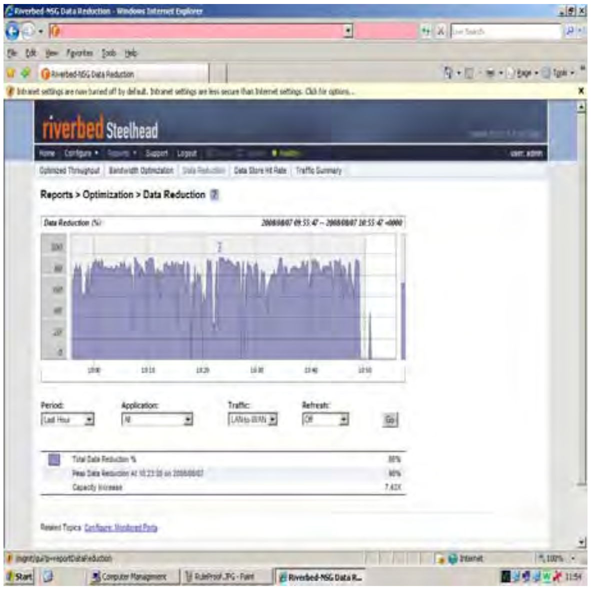
We chose to undertake the testing over the period between 12pm to 2:30pm to reduce the impact on the end users. The pair of Riverbed Steelhead appliances were put in place, but configured in “Pass Through” mode to disable any optimization, as seen in the screenshot below.

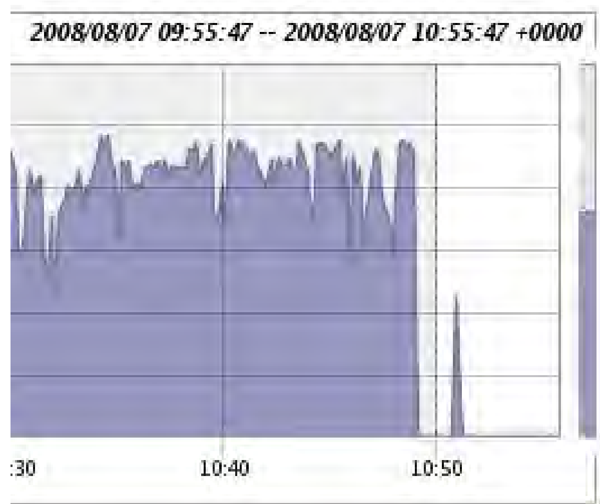


The screenshot below shows the transfer rate being generated by the remote replication software before the templates were created. In theory, a 2Mbps (Megabits per second) Megastream line should carry up to 240 KB/sec (Kilo Bytes per second). You can see that the aggregate transfer rate across the three replicated volumes is 170KB/sec (75KB/sec [highlighted] + 46KB/sec +49KB/sec). The remaining 70 KB/sec is being used by other inter-site traffic sharing the line, including Voice over IP (VoIP) circuits.

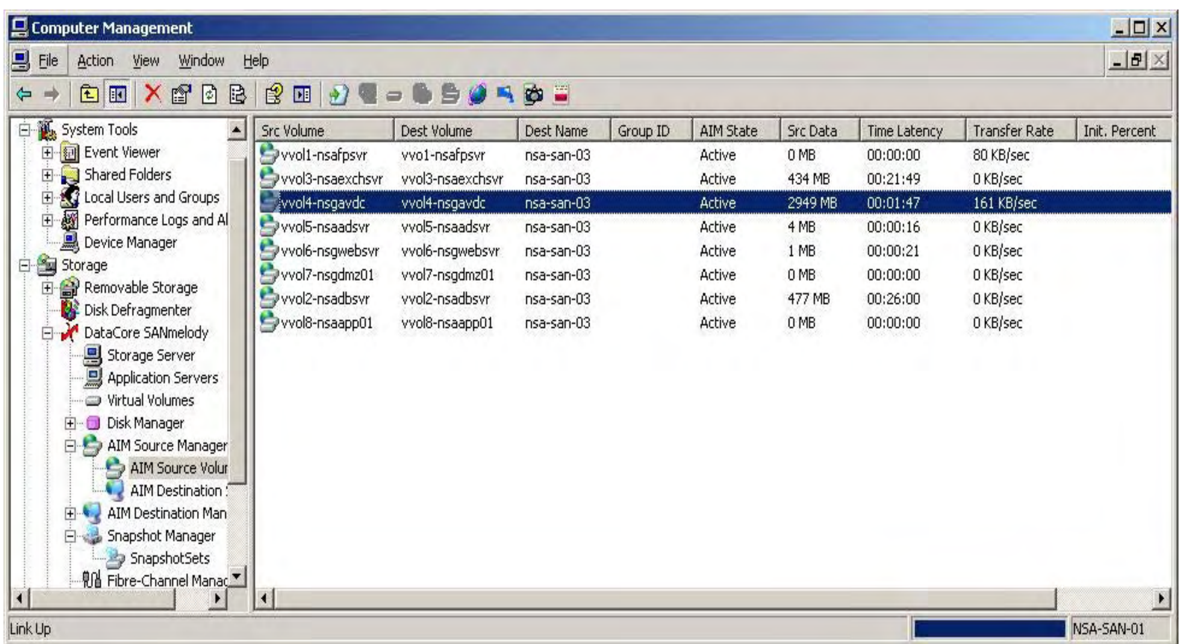


In order to verify that the Pass-Through rule was working correctly, a screenshot was taken of the Riverbed Data-Reduction chart. If there was a lack of data reduction, it proves that no data is passing through the Riverbed appliance for optimization. It is obvious from the flat line in the chart below that this is the case. The chart on the right hand side is an increased magnification of the full chart on the left.

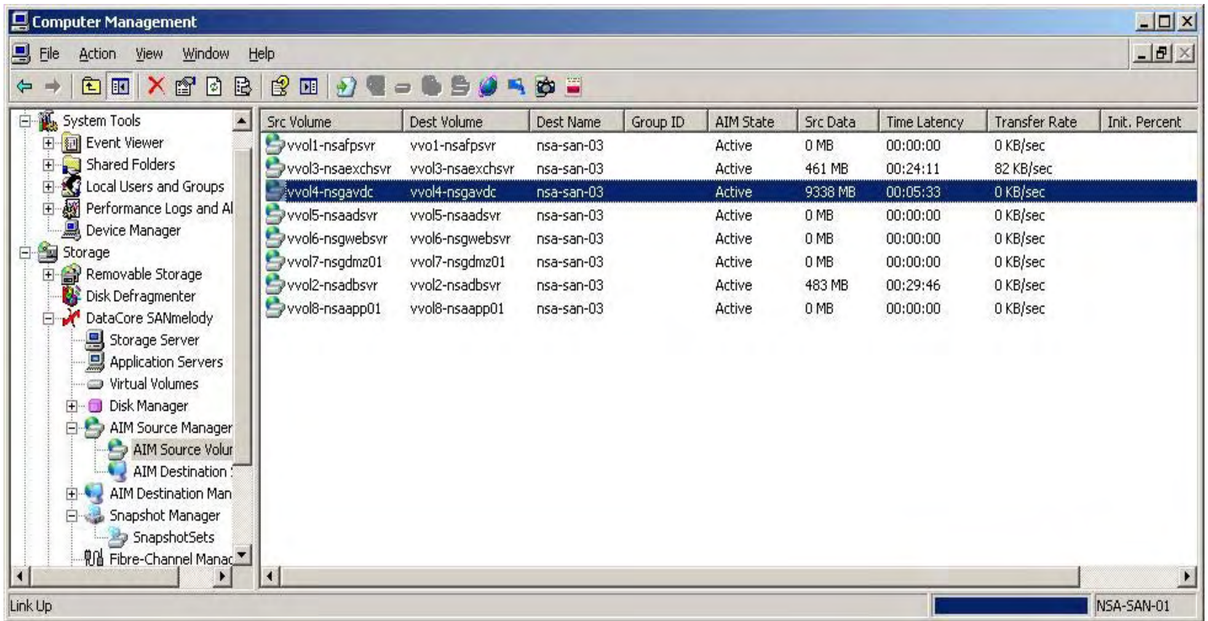


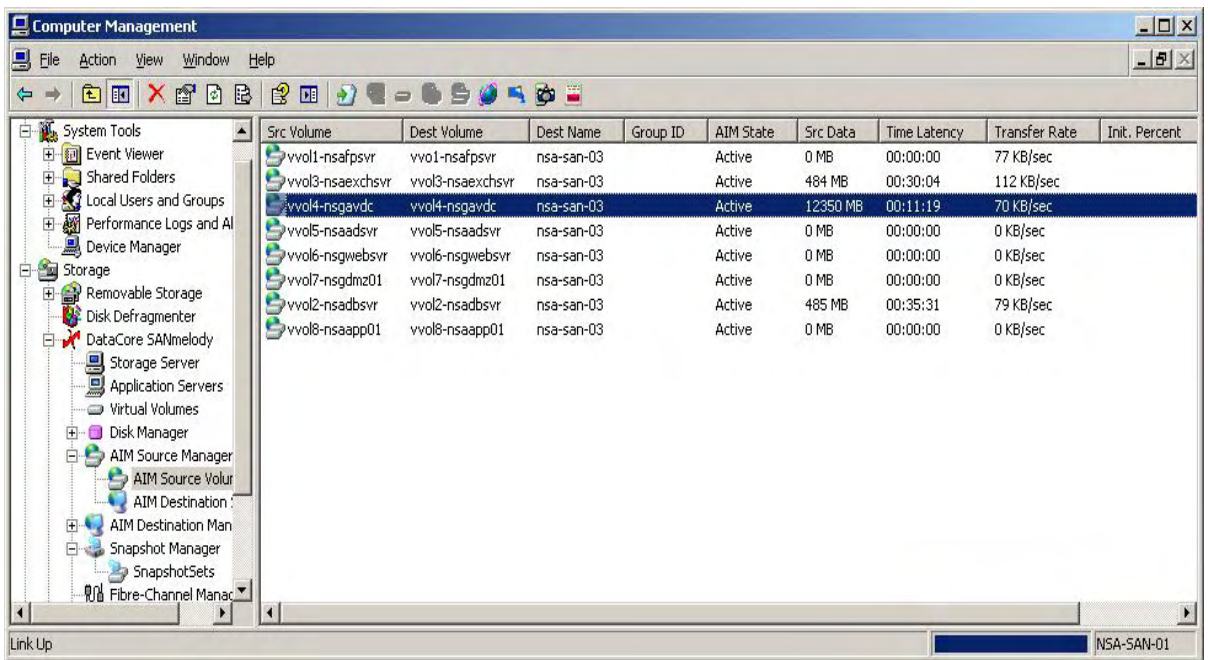


At this point, the first Virtual Machine was created from a template and the DataCore software began to asynchronously replicate it to the DR site. This is shown by the increase in Src Data (AIM Source Data) from 0 to 2949 MB for the third Src Volume in the screen shot below.

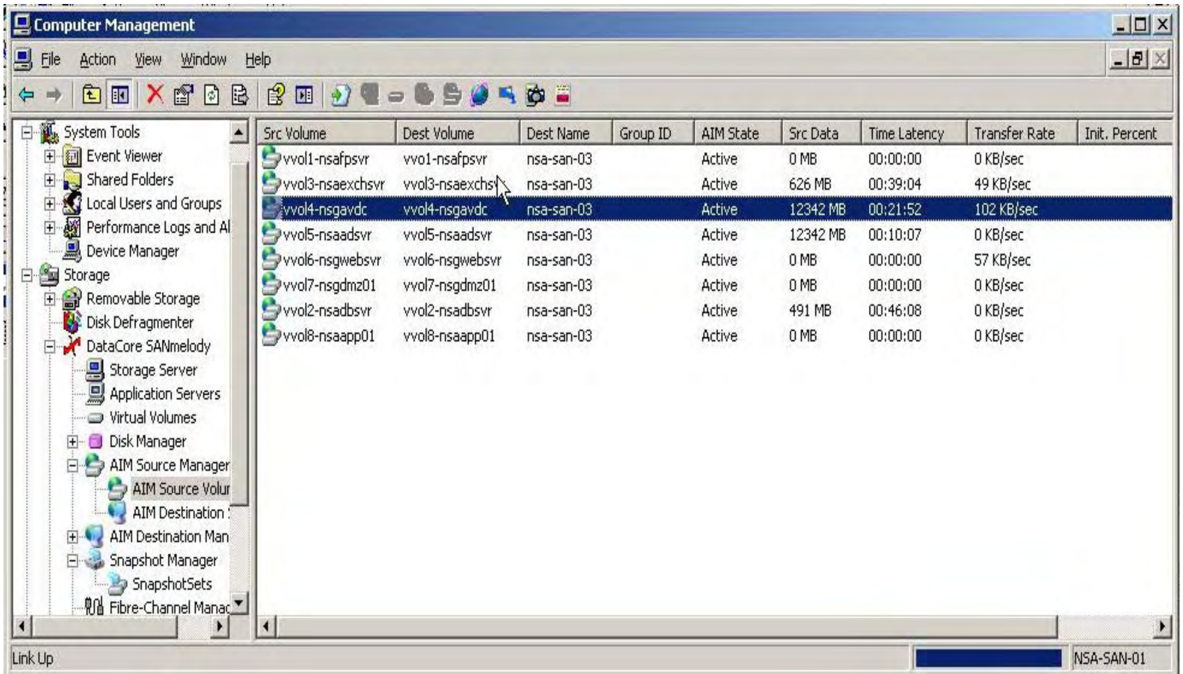


As is evident from the increasing Time Latency below, when competition for the line heats up, it slows down other volumes being remotely replicated. Their latency, which is increasing, measures the disparity between the production and DR site. Were there to be an incident at the production site in this instance, RTO and RPO SLA's could be compromised. This slowing trend continues as the newly created Virtual Machine size continues to grow until it reaches the defined 12GB limit.



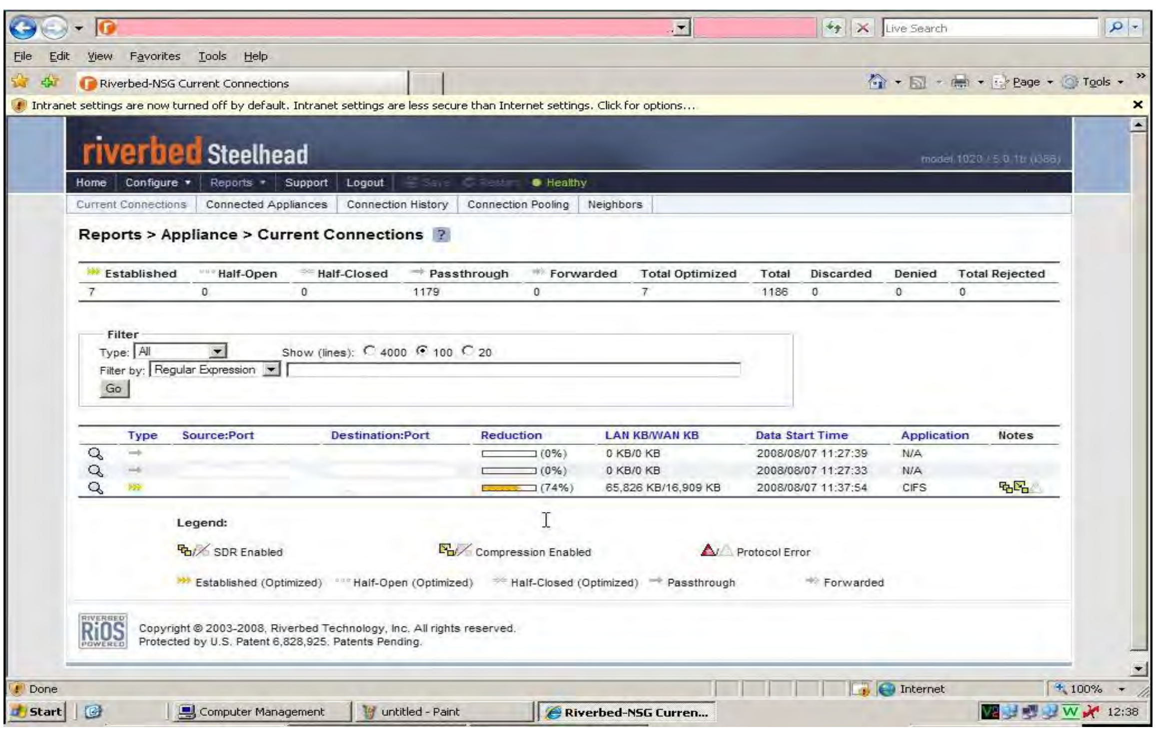


As you can see, by the time the second Virtual Machine has been created and stored, the latency has reached almost 40 minutes, from its initial starting point of 20 minutes and is expected to increase further. Replication traffic for the rest of the production site is also beginning to be affected.

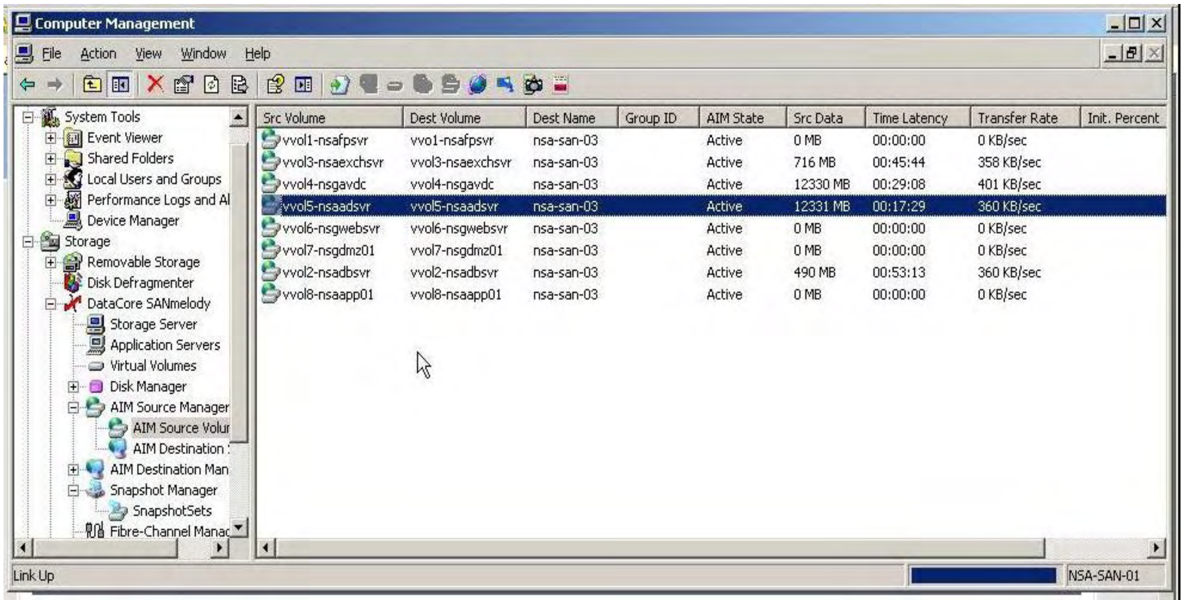


It is also possible to see that the transfer rate has remained relatively constant, implying that since there is more data to be transferred over the link, at a constant rate, the time to catch up will increase. This has potentially serious consequences for the RTO and RPO objectives of the organization, especially in a regulated DR environment.

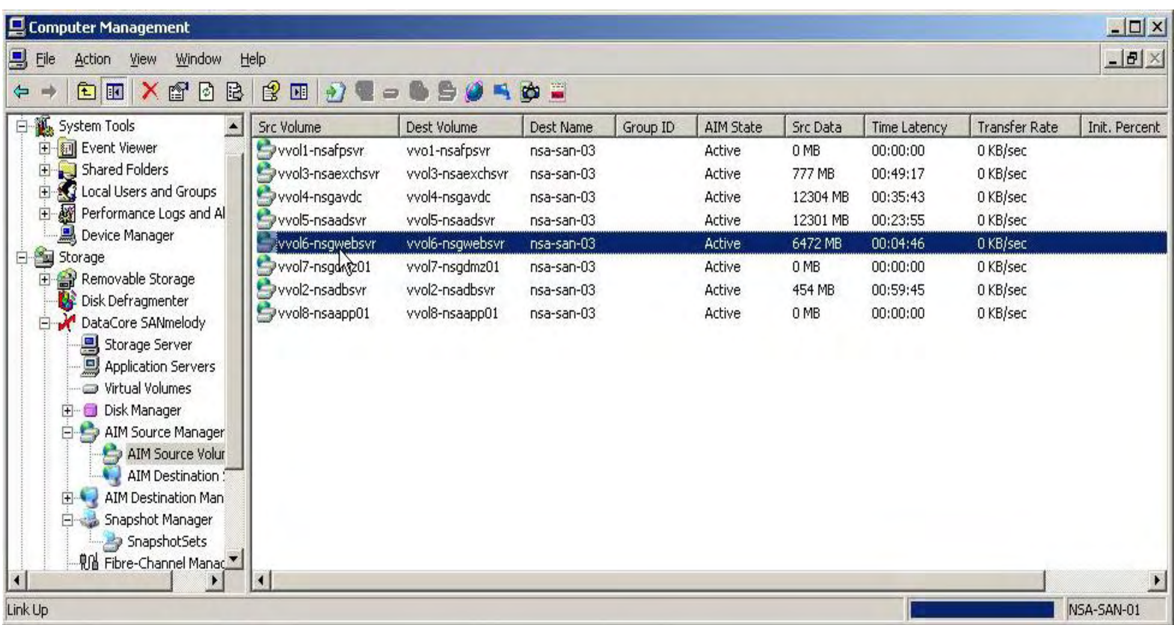
At 12:38 we instate the rule for Riverbed to optimize the CIFS traffic flowing through the appliances. The screenshot below illustrates the changes made.



The increase in throughput to 1479 KB/sec is evident immediately. We witnessed an 8x improvement in throughput instantly compared with the final screenshot from the previously non-optimized replication streams.

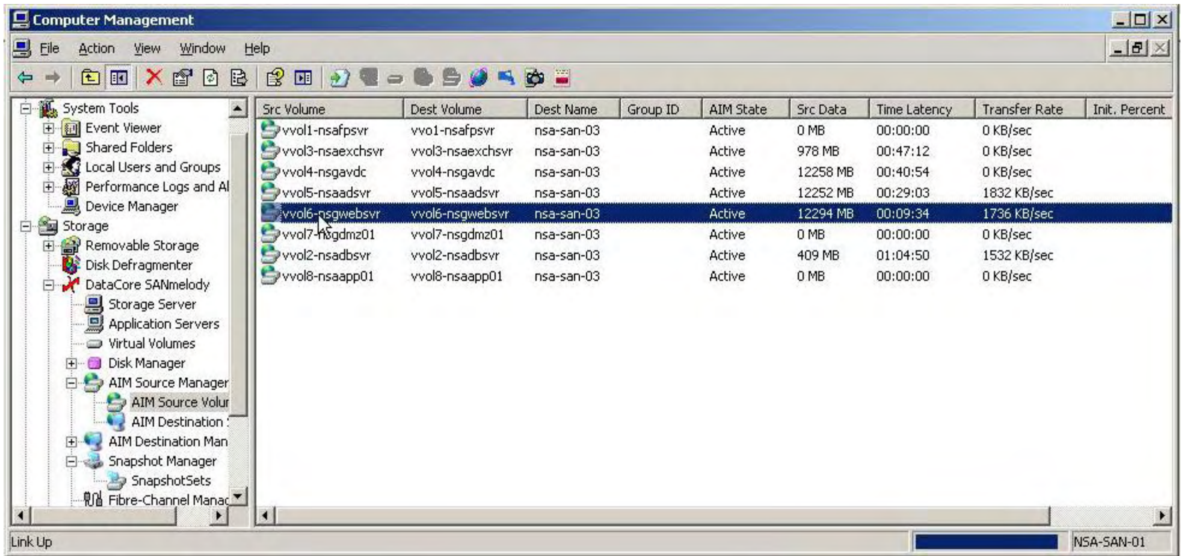


In order to further highlight the impressive performance that DataCore and Riverbed provide, we cloned a 3rd VM from the template and set it to replicate over the link. The screenshot below shows that progress, approximately halfway through.

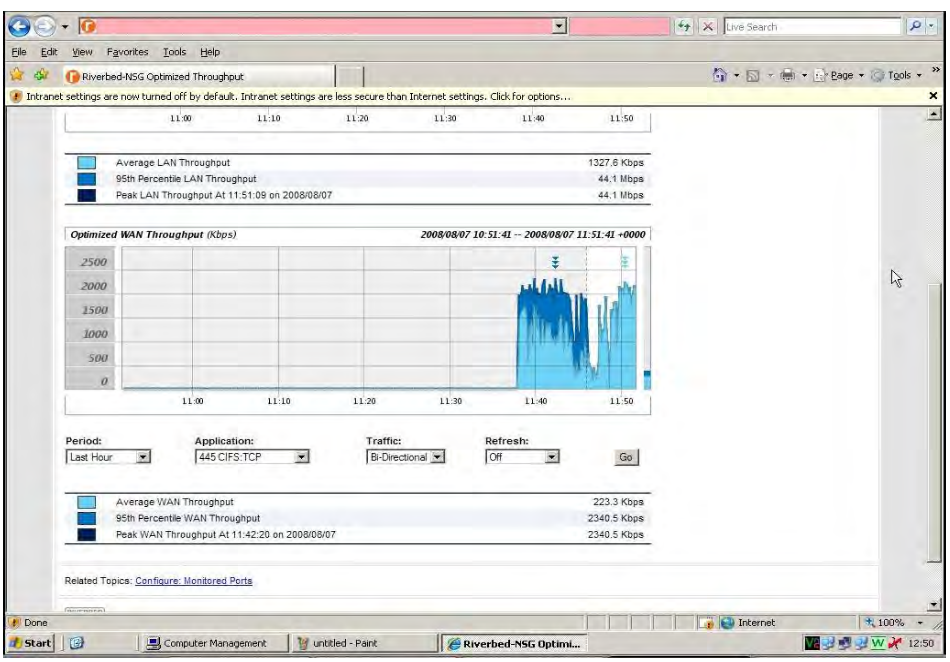


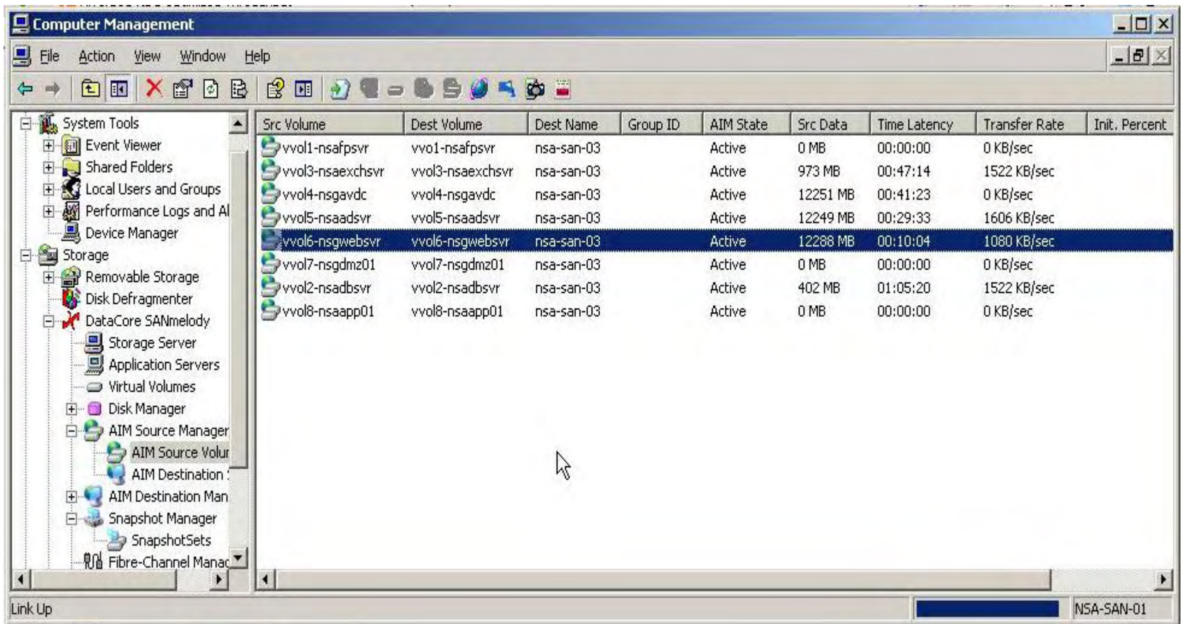
Even with the addition of an extra 6GB to the AIM buffer, the latencies are remaining roughly similar.
It is also relevant to consider that the Riverbed appliance takes a short while to “come up to speed” and since this was a mere 6 minutes after initializing the CIFS pass-through command, full speed has not yet been achieved.

By 12:51 the clone had completed and the Riverbed appliance was beginning to accelerate the CIFS data well. This is shown by an aggregate throughput of 5100KB/Sec, well in excess of 20x the theoretical maximum achievable previously.

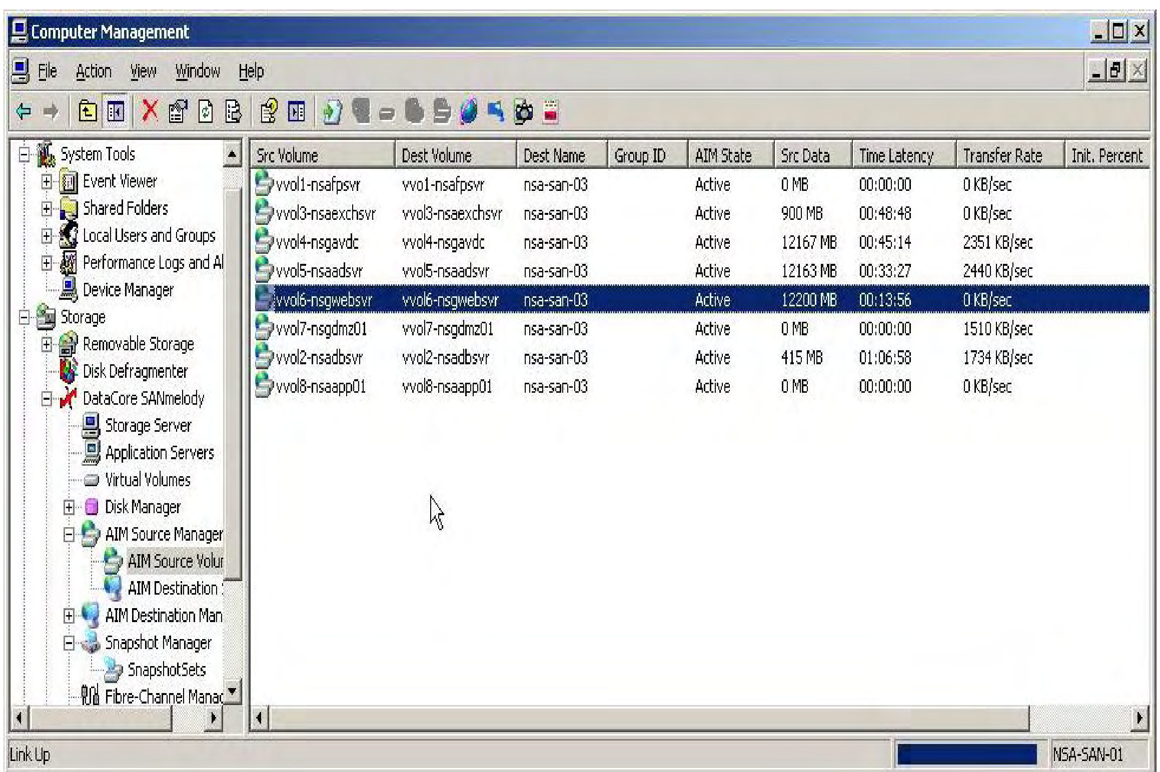


The Riverbed Appliance continued to accelerate the AIM transfer rate, and proof of this is displayed in the following screenshots. The first shows Riverbed passing through the CIFS traffic, and the subsequent DataCore screenshots show the transfer rate improvements as the Riverbed reaches full acceleration, as well as the reduction in source buffer size and a decrease in time latency.

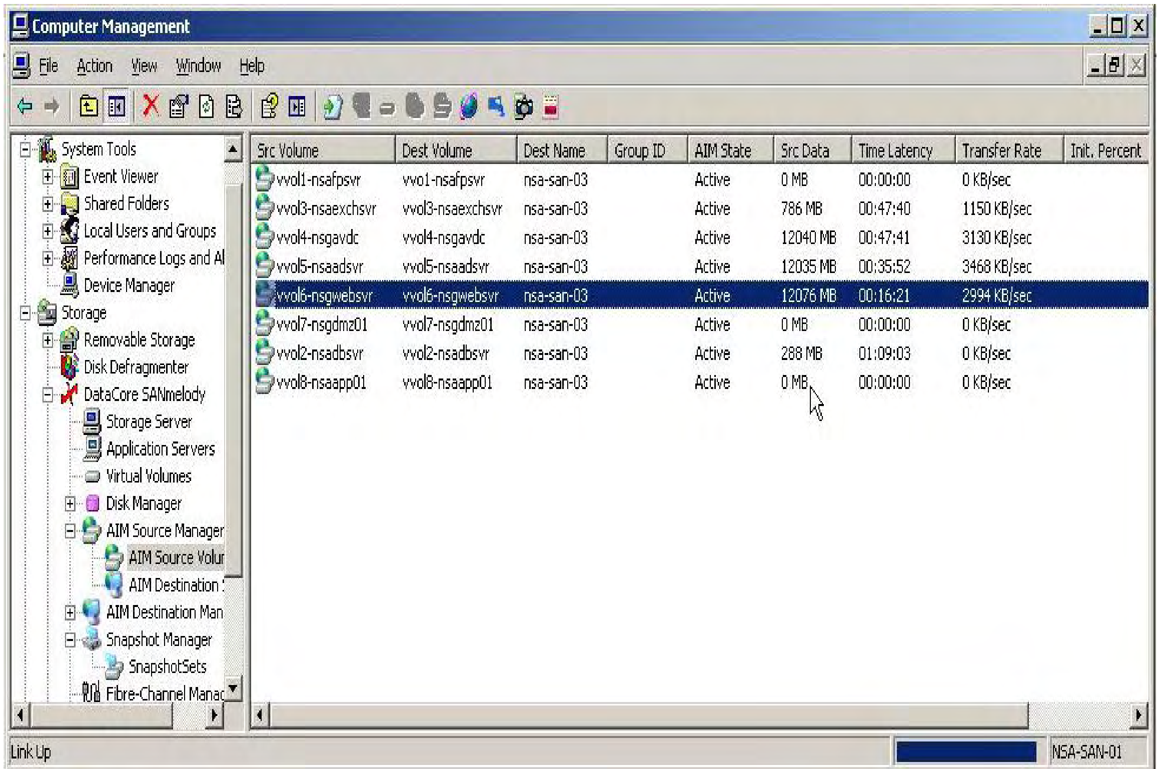




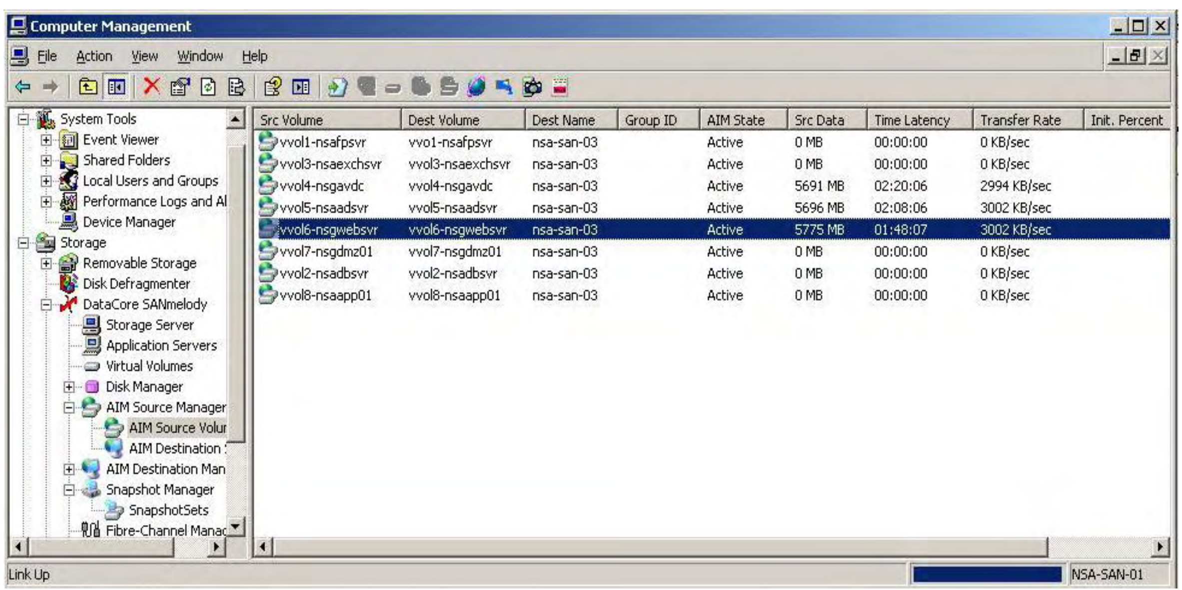
Src Volume	Dest Volume	Dest Name	Group ID	AIM State	Src Data	Time Latency	Transfer Rate	Init. Percent
vvol1-nsafpsvr	vvol1-nsafpsvr	nsa-san-03		Active	0 MB	00:00:00	0 KB/sec	
vvol3-nsaexchsvr	vvol3-nsaexchsvr	nsa-san-03		Active	973 MB	00:47:14	1522 KB/sec	
vvol4-nsgavdc	vvol4-nsgavdc	nsa-san-03		Active	12251 MB	00:41:23	0 KB/sec	
vvol5-nsaadsvr	vvol5-nsaadsvr	nsa-san-03		Active	12249 MB	00:29:33	1606 KB/sec	
vvol6-nsgwebsvr	vvol6-nsgwebsvr	nsa-san-03		Active	12288 MB	00:10:04	1080 KB/sec	
vvol7-nsgdmz01	vvol7-nsgdmz01	nsa-san-03		Active	0 MB	00:00:00	0 KB/sec	
vvol2-nsadbsvr	vvol2-nsadbsvr	nsa-san-03		Active	402 MB	01:05:20	1522 KB/sec	
vvol8-nsaapp01	vvol8-nsaapp01	nsa-san-03		Active	0 MB	00:00:00	0 KB/sec	



Src Volume	Dest Volume	Dest Name	Group ID	AIM State	Src Data	Time Latency	Transfer Rate	Init. Percent
vvol1-nsafpsvr	vvol1-nsafpsvr	nsa-san-03		Active	0 MB	00:00:00	0 KB/sec	
vvol3-nsaexchsvr	vvol3-nsaexchsvr	nsa-san-03		Active	900 MB	00:48:48	0 KB/sec	
vvol4-nsgavdc	vvol4-nsgavdc	nsa-san-03		Active	12167 MB	00:45:14	2351 KB/sec	
vvol5-nsaadsvr	vvol5-nsaadsvr	nsa-san-03		Active	12163 MB	00:33:27	2440 KB/sec	
vvol6-nsgwebsvr	vvol6-nsgwebsvr	nsa-san-03		Active	12200 MB	00:13:56	0 KB/sec	
vvol7-nsgdmz01	vvol7-nsgdmz01	nsa-san-03		Active	0 MB	00:00:00	1510 KB/sec	
vvol2-nsadbsvr	vvol2-nsadbsvr	nsa-san-03		Active	415 MB	01:06:58	1734 KB/sec	
vvol8-nsaapp01	vvol8-nsaapp01	nsa-san-03		Active	0 MB	00:00:00	0 KB/sec	



The corresponding DataCore screenshots below further serve to illustrate the excellent performance of this solution.



The previous screenshot was taken at 14:33 and the final VM Clone completed its buffering at 12:51. If we subtract the total AIM Src Data figures, to find the volume of data transferred across the link in the 102 minutes we left it processing, we arrive at a figure of 20,063MB or 20.063GB of data transferred in 102 minutes. The full theoretical calculation and the actual transfer calculations are shown below.

Theoretical Time to Replicate without Optimization

A 2MBps line is capable of transferring 240KB/sec without optimization and therefore, to initially replicate three 12GB VM clones (36 GB total) with no other competing traffic would take over 42 hours:

36 GB = 36 x 1024 X 1000 = 36,864,000 KB
36,864,000 KB ÷ 240 KB/sec = 153,600 sec
153,600 ÷ 3600 sec/hour = 42.7 hours

While it is unlikely that you will transfer this volume of data on a regular basis, it is important to recognize that VM templates are being created daily in a large number of organizations. This also excludes the traffic created by the rest of the network in its day to day operation.

14 X faster after WAN Optimization

We have demonstrated that we can transfer 20 GB of data in 102 minutes using Riverbed's WAN optimization. This gives us an effective 3425 KB/sec transfer rate, or roughly 14x faster than the theoretical maximum of 240KB/sec that the line can carry.

20 GB / 102 min = 0.196 GB/min = 3.345 MB/sec = 3425 KB/sec
3425 KB/sec ÷ 240 KB/sec = 14.2 times faster

Conclusions

This live analysis reveals that we can drop the time to remotely replicate our 3 VMs from 42 hours down to 3 hours, substantially improving our achievable Recovery Time Objective.

36 GB / 0.196 GB/min = 183 min = 3.06 hours

Said differently, an organization can now recover more than 2 days worth of data that may have been lost in their prior un-optimized configuration, or would have had to spend substantially more on faster leased lines to minimize the exposure to serious data loss.

Furthermore, DataCore's remote replication software allows customers to map snapshots of the remotely replicated volumes to the ESX hosts at the DR site to dramatically simplify and speed up recovery without painful back-ups and restores.
Equally compelling is the cost savings afforded by these technologies. The quotes received to increase the bandwidth beyond 2 Mbps required very high ongoing expenses:

- 4MBps: Configuration Cost - £5,000 and Annual Rental - £11,000
- 10Mbps via LES10 Configuration Cost - £9,000 and Annual Rental - £22,000

The total cost of the combined and DataCore / Riverbed solution was only:

- £16,800 in hardware plus software with an annual cost of £2,800.

Needless to say, these living ROI finding are convincing reasons to employ DataCore and Riverbed in your DR plans.