FalconStor Virtual Tape Library (VTL) and Deduplication Performance

Evaluation report prepared under contract with FalconStor Corporation

Introduction

Virtual tape libraries (VTL) are a key technology that can save a considerable amount of time and storage capacity in the data center by utilizing disk storage that presents itself as a tape library solution. VTLs typically work seamlessly with enterprise backup software and help reduce the amount of time backup operations consume. VTL solutions often bring compression, single-instance store and data deduplication functions that help reduce the total amount of storage capacity required for backup and related activities. VTLs can make a direct positive impact on the backup window as well as recovery time objectives (RTO) and recovery point objectives (RPO).

Some of the key metrics of VTL solutions are their ***sustained*** rate of I/O during backup operations and during data deduplication operations.

To evaluate the performance of its solutions, FalconStor commissioned Demartek to validate the performance of the FalconStor VTL solution and its deduplication solutions, as they come from the factory, without any special tuning adjustments. The ***sustained*** performance of these solutions is impressive, and was measured using conservative data intended to be typical of many enterprises.

During the testing, Demartek measured the speed of the I/O operations and the data savings achieved using the compression features of the FalconStor VTL solution and the deduplication features of the FalconStor SIR solution.

Evaluation Summary

We found that a single-node FalconStor VTL achieved a ***sustained*** I/O rate of approximately 1.4 GB per second driven by two clients, which translates to roughly 5 TB per hour. The dual-node FalconStor SIR achieved a ***sustained*** I/O rate of 974 MB per second for deduplication, which translates into more than 3 TB per hour. We understand that FalconStor provides these solutions with additional nodes, and based on their architecture, would expect linear scaling of these rates.

These are impressive rates and certainly make the FalconStor solution worth considering.

Evaluation Environment

The evaluation was conducted at the FalconStor headquarters in Melville, New York by Demartek using the following test environment. The FalconStor solutions are software-based appliances that can be installed on a variety of name-brand server platforms and can use various brands of storage on the back-end.

FalconStor Virtual Tape Library (VTL) (single node configuration)

- Dell PowerEdge R900 Server
  - Qty. 2 Intel Xeon E7320 processors, 2.13 GHz (8 total cores)
  - 64 GB RAM
  - Qty. 4 QLogic QLE2562 dual-port, 8-Gbps Fibre Channel HBA (in x8 slots)
  - Qty. 3 Hifn Express DR 1050 compression cards (in x4 slots)
- Storage (~50 TB)
  - Head controller unit with qty. 16, 1 TB SATA disk drives configured into 14+1 RAID5 with one hot spare
  - Qty. 3 expansion units with qty. 16, 1 TB SATA disk drives configured into 14+1 RAID5 with one hot spare

FalconStor Single Instance Repository (SIR) (dual node configuration)

- Qty. 2 Dell PowerEdge R900 Server
  - Qty. 4 Intel Xeon E7440 processors, 2.4 GHz (16 total cores)
  - 256 GB RAM
  - Qty. 4 QLogic QLE2562 dual-port, 8-Gbps Fibre Channel HBA (in x8 slots)
- Storage (~50 TB)
  - Head controller unit with qty. 16, 1 TB SATA disk drives configured into 13+2 RAID6 with one global hot spare
  - Qty. 3 expansion units with qty. 16, 1 TB SATA disk drives configured into 14+2 RAID6

The combined FalconStor VTL and SIR solution includes a Brocade 300 8-Gbps FC switch and a Gigabit Ethernet switch so that the entire unit is a complete solution.

Clients

- Qty. 2 Dell PowerEdge R200 servers
  - Intel Xeon X3210 processor, 2.13 GHz (4 total cores)
  - 8 GB RAM
  - Windows Server 2003 R2 Enterprise x64 edition

Because the FalconStor VTL can emulate many different types of tape drives, and to take advantage of existing drivers installed with the clients, the tape drives appeared to the clients as generic Dell PowerVault 110T LTO2 tape drives.

Test Process

The test consisted of running a backup workload to the VTL simultaneously from two clients, using eight virtual tape drives for each client, followed by moving the data from the VTL into the SIR solution, which performs data deduplication. The key data movement metrics were the ***sustained*** I/O rate, the observed data compression rate, and the observed data deduplication ratio. We noticed that the FalconStor VTL emulates almost all available physical tape libraries.

To perform the backup workload, a simple utility called "tapewrite" was used. This utility performs direct tape I/O with specific parameters that set the compression ratio and redundancy factor (deduplication rate), so that predictable results can be produced. Each of the eight tapes written by the clients contained 250 GB of data, with a 25% randomness ratio (yielding approximately 4:1 compression) and a redundancy ratio yielding an 8:1 data deduplication ratio.

VTL *Sustained* I/O Rate

The observed ***sustained*** I/O rate for two clients writing to the VTL simultaneously was approximately **1440 MB per second** (approximately 720 MB/sec for each client). The following screen shots show the measurements taken.

The first screen shot is taken from the VTL command-line console, showing I/O rates from the two clients during a sixty-second period in approximately the middle of the test run. This rate fluctuated slightly during each sixty-second period, but was observed over the duration of the test run. It should be noted that the VTL test ran for approximately 47 minutes and maintained essentially the same I/O rate for the duration of the run.

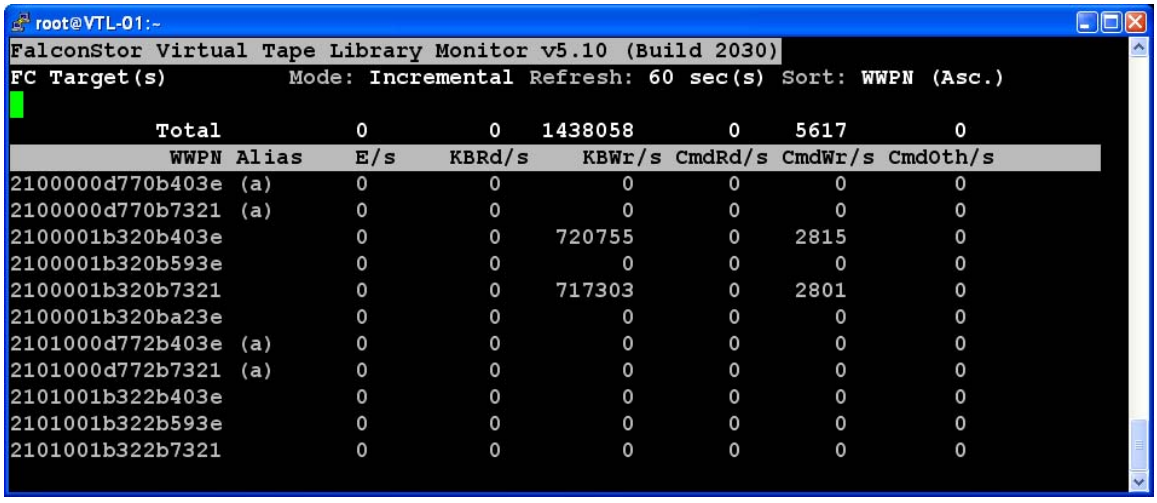

Figure 1 - VTL combined client I/O rates

The following screen shots were taken from VTL console showing the I/O rate from each client separately. Prior to running the simultaneous workloads on both clients, a short workload was run on the first client that achieved nearly 800 MB per second. Note that during the tests, the I/O rates from each client achieved their peak I/O fairly quickly, and ***sustained*** that rate for the duration of

the run. Our suggestion to FalconStor was to provide an option to show a fixed scale on the right side of the graph, so that more precise and consistent comparisons can be shown.

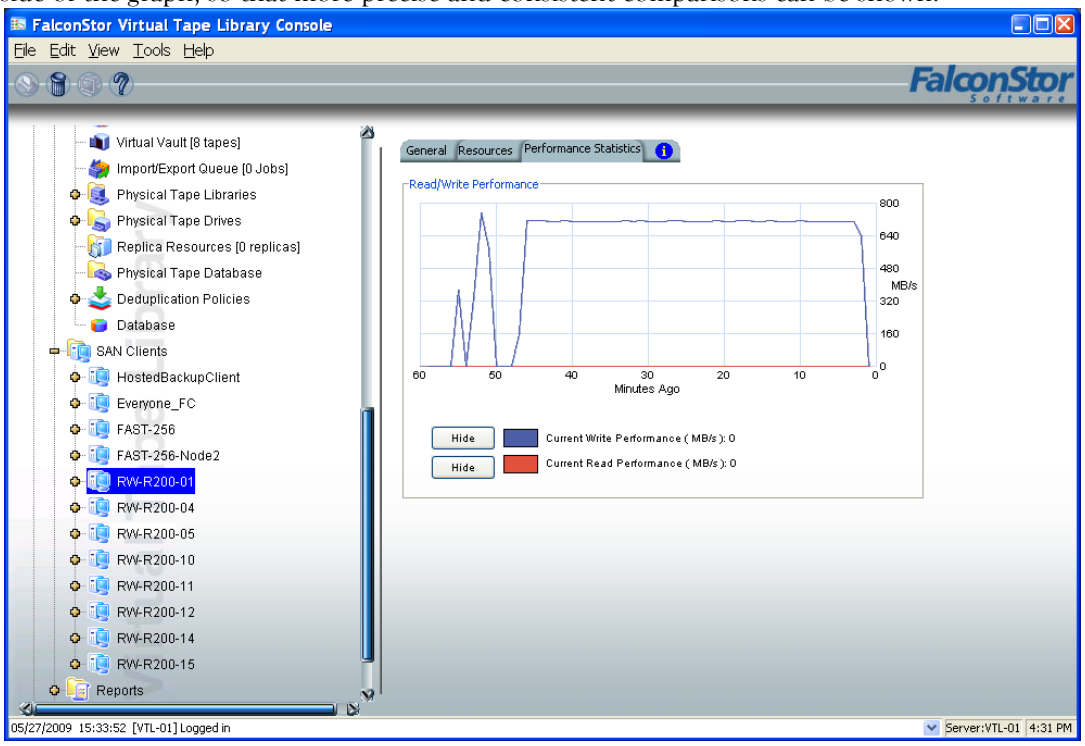

Figure 2 Client 1 VTL I/O Rate

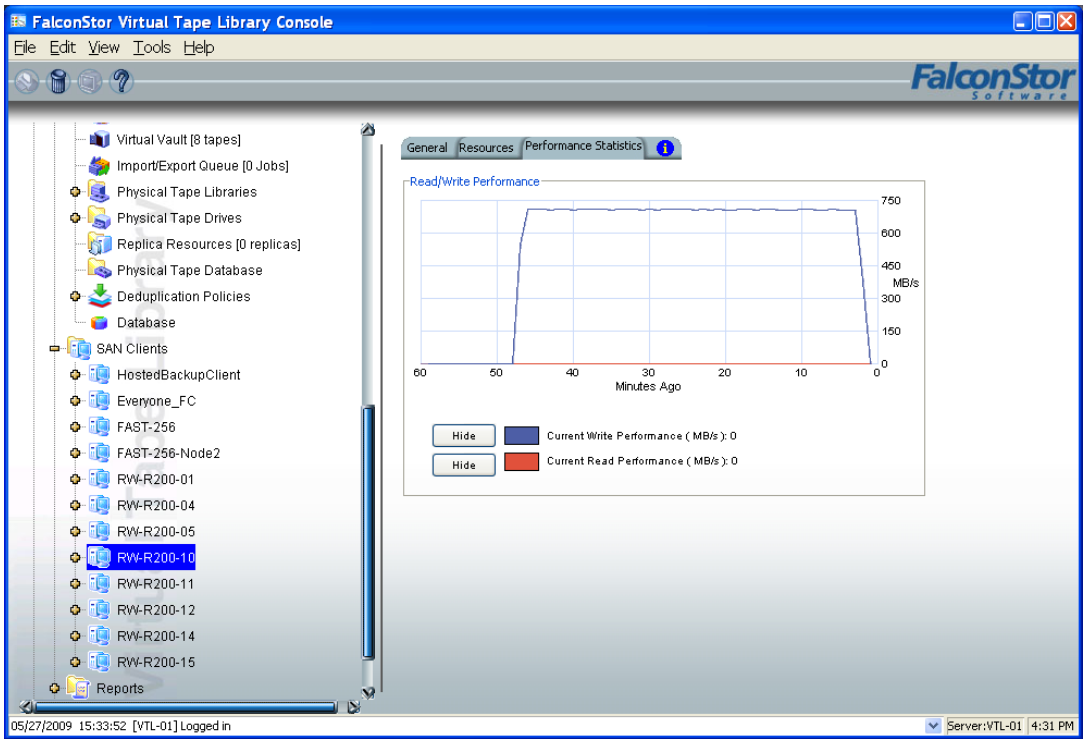

Figure 3 Client 2 VTL I/O Rate

During this test, the VTL CPU cores were running at approximately 26% utilization, suggesting that this VTL could have taken on heavier workloads.

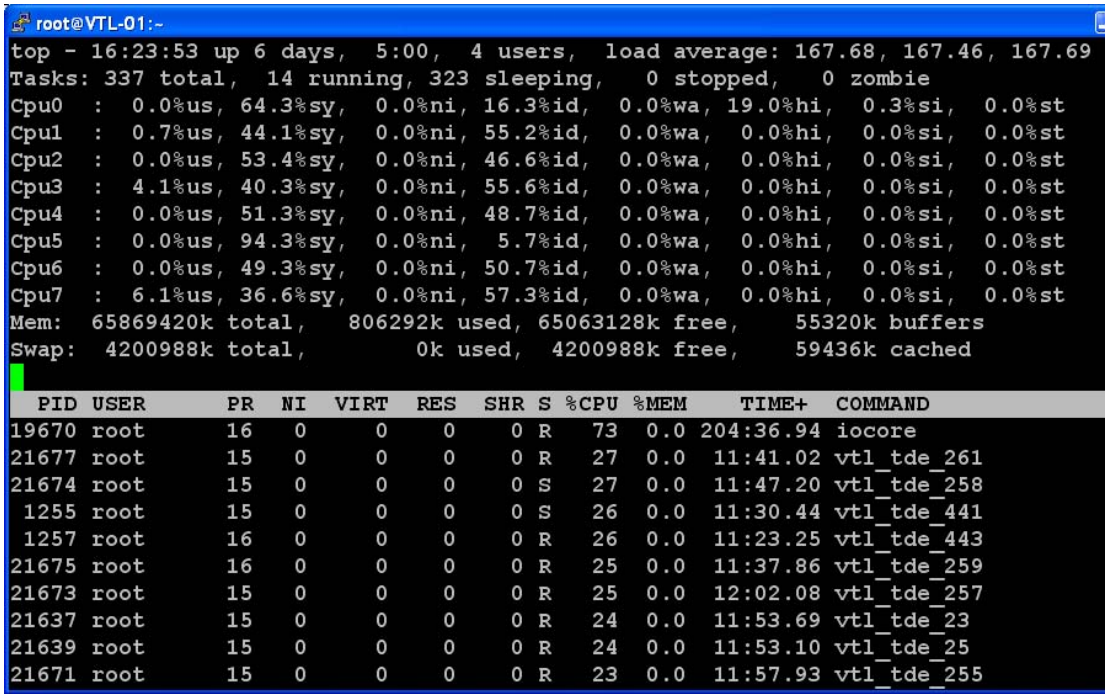

Figure 4 VTL CPU utilization rates

With the ability to sustain high data throughput rates, FalconStor VTL allows backup environments to operate more efficiently while reducing total backup time. Up to 1024 virtual tape drives can be created per VTL node, allowing backup jobs to be distributed across multiple tape targets, rather than having to share a limited number of physical tape drives. This means that simultaneous backup jobs can scale in number until the total data ingest capacity of the VTL system is reached. Since backup jobs won't have to queue up while waiting for a free tape drive, the total backup time across all jobs can be significantly reduced.

SIR (deduplication) *Sustained* I/O Rate

Following the backup to the VTL, the data was then moved to the single instance repository (SIR) using the FalconStor deduplication functions. The data in the VTL was deduplicated, but the metadata in the VTL was not, since the metadata was unique.

The data deduplication function was initiated taking the data that had been backed up to the VTL in the first phase of this test and deduplicating it. The data was created with a predictable, reasonably conservative, duplication ratio as mentioned previously.

The ***sustained*** data deduplication rate for this test was approximately **974 MB per second**, as shown below in Figure 5. We believe that this rate could have been higher, at least 1100 MB per second, if there were more streams running. Note that in Figure 6 the ***sustained*** rate for one node was well over 550 MB per second. As with the VTL solution, the data deduplication I/O rate peaked relatively quickly and ***sustained*** that rate for the duration of the test, with some minor fluctuation.

The SIR configuration assigns tasks corresponding to each virtual tape drive to each of the nodes in the configuration; in this case there were two nodes. In this particular case, the eight VTL virtual drives were assigned to the two SIR nodes with five drives assigned to one node and three assigned to the other. This is reflected in the performance graphs, with one node handling approximately 62% of the workload and the other handling the remaining workload. Due to this particular assignment, one node finished its work a bit sooner than the other, resulting in a spike in throughput in the remaining node.

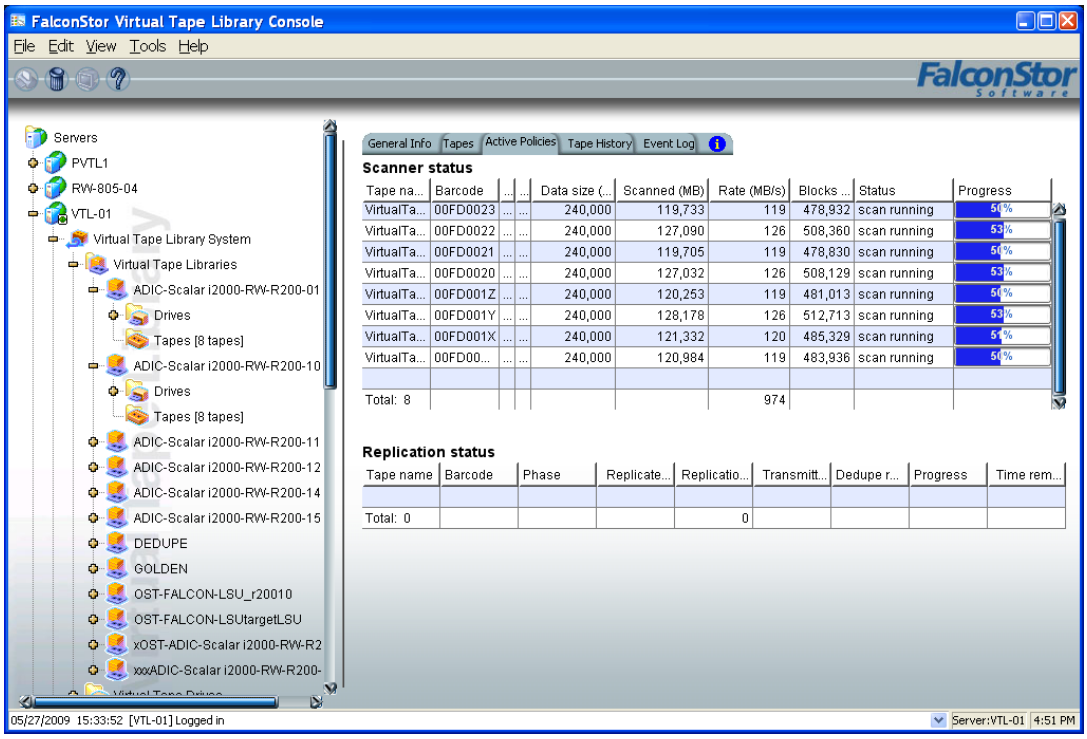

Figure 5 Data Deduplication I/O rate

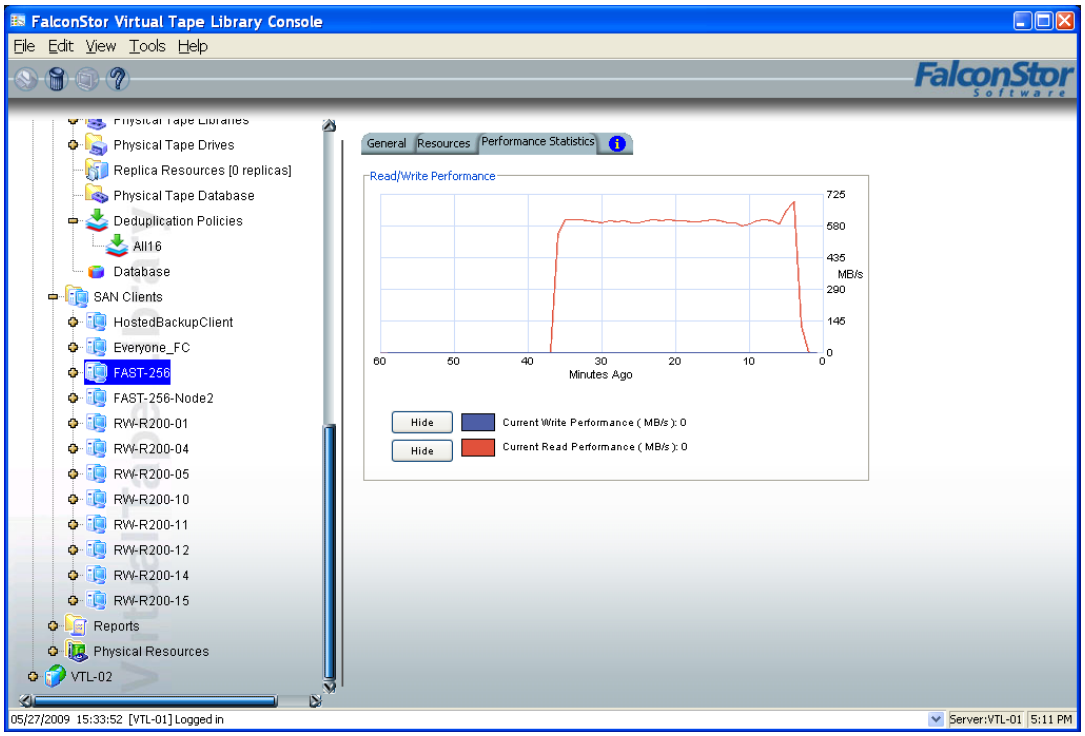

Figure 6 Node 1 Deduplication rate

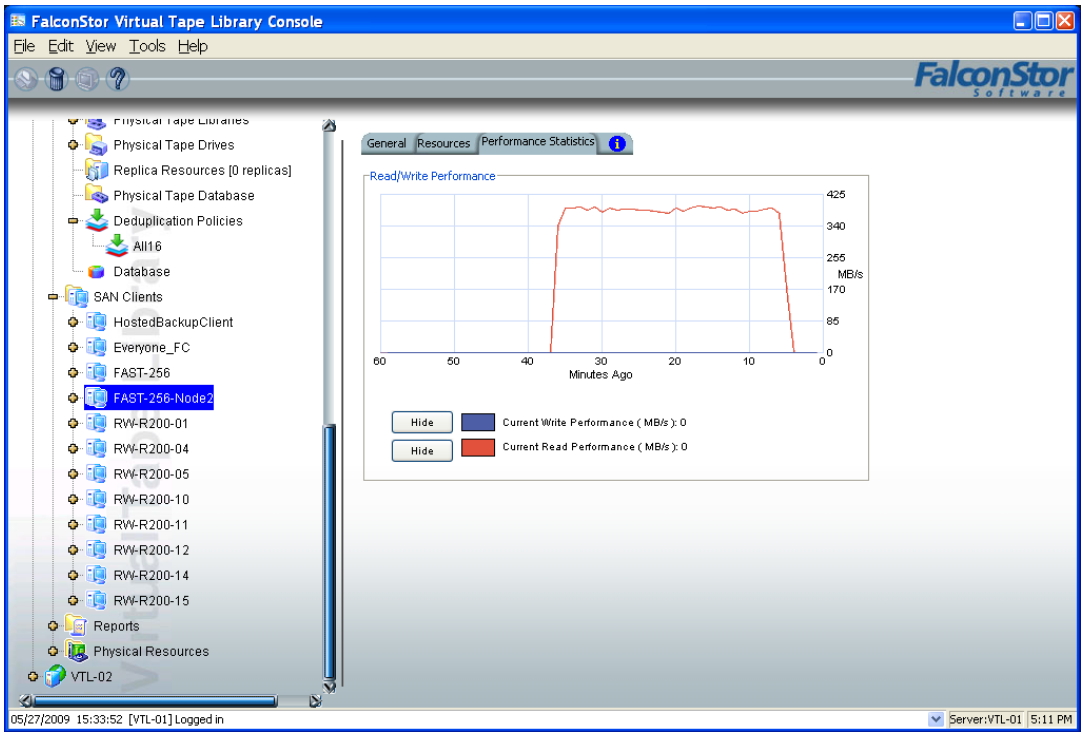

Figure 7 Node 2 Deduplication rate

One final statistic is the deduplication ratio. In this case, approximately 1875 GB of data were originally written, but this data only consumed 59GB in the SIR, resulting in a deduplication rate of 32:1. In other words, nearly 2 TB of logical backup data filled only 59GB of physical disk space, for a storage savings rate of 97%. While this reduction rate is highly dependent on the data and will vary by application and other factors, deduplication technology dramatically extends the amount of data that can be retained on a disk system, lowering storage purchase costs and the associated power and cooling needs.

Figure 8 Deduplication Ratio