Exploring the ROI of VTL Deduplication Solutions: How HP StorageWorks Stands Up Against the Competition

Silverton Consulting
StorInt™ Briefing

Introduction

Particularly in today's economy, costs and return on investment (ROI) often dominate product selection decisions. However, gathering the appropriate information can be a daunting task with changing prices, changing capabilities, etc. In an effort to simplify this task and at least segregate the options worth further consideration, two models were developed and funded by Hewlett Packard. While both models are similar they are necessarily different as one was built to determine the ROI of the HP VLS9000, an enterprise class deduplication system, and some if its unnamed competitors; the other model determines the ROI for the HP D2D4112, a deduplication appliance for the small to medium (SMB) data centers, and some of its unnamed competitors.

HP's VLS9000 and D2D4112 deduplication systems

HP's enterprise class deduplication system, the VLS9000, can be configured as a multi-node system supporting four different storage arrays, a 40TB, 30TB, 10TB or 7.5TB storage array. Unlike some deduplication products, VLS9000 provides inline hardware data compression and software deduplication during post-processing after data has been written. Data compression can reduce a 50TB backup in half or to 25TB of compressed data and deduplication can further reduce this by up to a factor of 20X (20*2=40X total) or 1TB of unique compressed backup data. In addition, the VLS9000 sustains up to 600MB/sec of post-processing deduplication and compression throughput per system node. Thus, with eight nodes per system, a VLS9000 system can sustain over 4.8GB/sec of data throughput. The deduplication capability of this HP offering is a separately licensed option. However for purposes of the model discussed below, this option, because of its significant storage increasing ability, is assumed to be purchased.

For SMB data centers, HP's D2D4112 appliance comes pre-packaged with inline deduplication capabilities that can also reduce the backup data load by a factor of 50 and requires no post-processing. The D2D4112 can sustain up to 175MB/s of deduplication data throughput. HP's D2D4112 supports 12TB or with an optional capacity expansion, 24TB of raw data storage using RAID 6 protection. It can also be configured to support iSCSI or FC virtual tape devices.

Both the D2D4112 and VLS9000 emulate tape devices and as such, are configured from the host or server as virtual tape transports. By emulating tape devices both HP products can transparently support any backup application that uses tape and thus, works with most major backup applications.

Competitive products of both the VLS9000 and the D2D4112 have been chosen based on similar deduplication capabilities, expandability, etc.

The Financial Model

In developing both of the models, simplicity to the client was paramount. Indeed, the only input data required by the client for both models is information concerning the expected workload both current and future. Specifically, the potential client would derive with their HP representative backup workload parameters including:

- Amount of data to be backed up for each full and incremental backup
- Time window to complete a full backup,
- Number of weeks of backup data to retain online,
- Percent of backup data growth
- Full time equivalent (FTE) cost of an administrator

With the input of these workload parameters, the models automatically configure the required HP and competitor storage systems. The models further project changes to the initial configurations for the next five years based on the percentage of backup growth given as input. Two similar but different financial models were derived to accommodate and ease the determination of this initial and subsequent storage configurations necessitated by the very different workload requirements of the enterprise systems versus the SMB systems.

The models also include more static financial information regarding power and cooling costs and facility footprint costs. For both models and all configurations these costs have been fixed as follows:

Fixed parameter	Cost
Power	$0.0937/Kwh
Cooling	70% of power cost
Facility footprint	$40 per square foot.

Table 1 Fixed Model Parameters

Critical assumptions in both models are 1) that all financial information is stated in US dollars and 2) no inflation has been projected. Maintenance costs were based on 4-hour repair for the enterprise and next business day repair for the SMB systems.

After inputting the potential workload requirements, the models each generate 15 charts of financial and system comparison data. Of these, three charts are of particular interest in that they represent consolidated, "rolled-up" information of other generated charts. Specifically, these charts include:

- **Projected Capital Expense** – shows both projected initial acquisition and subsequent

capital outlays over the next five years

- **Loaded Operating Expense** – shows the combined costs of power and cooling, facility footprint, FTE administration, and system maintenance
- **Cost per GB stored** – shows total of cumulative capital costs and year five operating expense divided by end of year five storage requirements in GB.

Additionally, the administration expense chart is shown and explained as administration expense represents such a significant portion of "loaded operating expense."

An Enterprise Deduplication Example

The projected workload of the hypothetical enterprise class backup was input as follows:

- 50TB of full backup per week
- 10% of incremental backup data per day for five days during the week
- 50% annual backup data growth
- 8 weeks of backup data retained online and
- 12 hour full backup window
- $150,000/year FTE cost of administrator

Given these requirements and projected growth, the storage requirements at the end of the next five years are as follows in Table 2:

	Year 1	Year 2	Year 3	Year 4	Year 5
YE full backup (TB)	75	112.5	168.8	253.1	379.7
Retained capacity (TB)	621	931	1,396	2,094	3,142
Required perf. (MB/s)	1,736	2,604	3,906	5,859	8,789

Table 2 Enterprise backup workload requirements

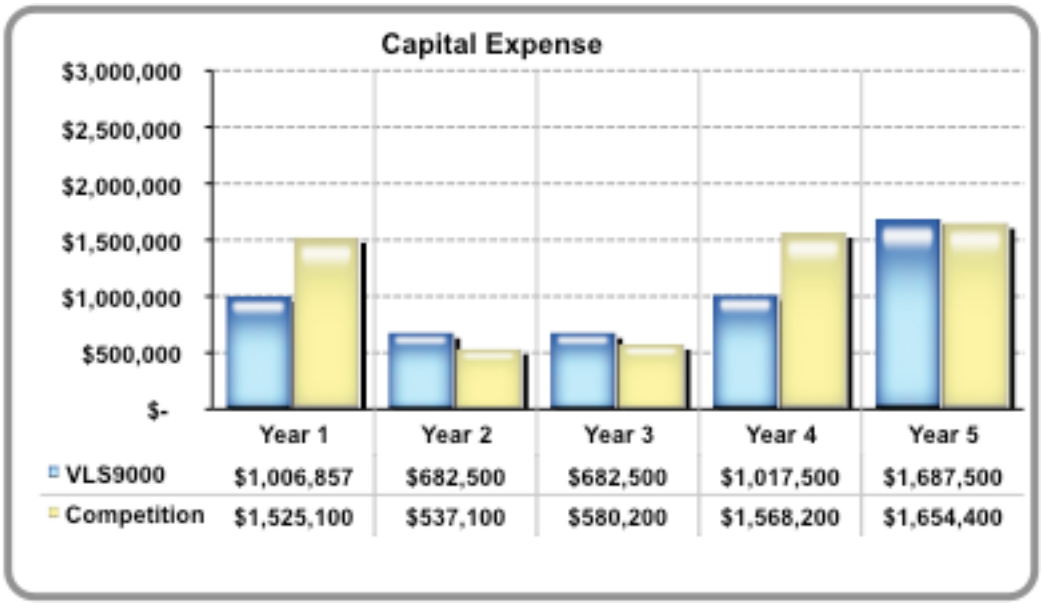
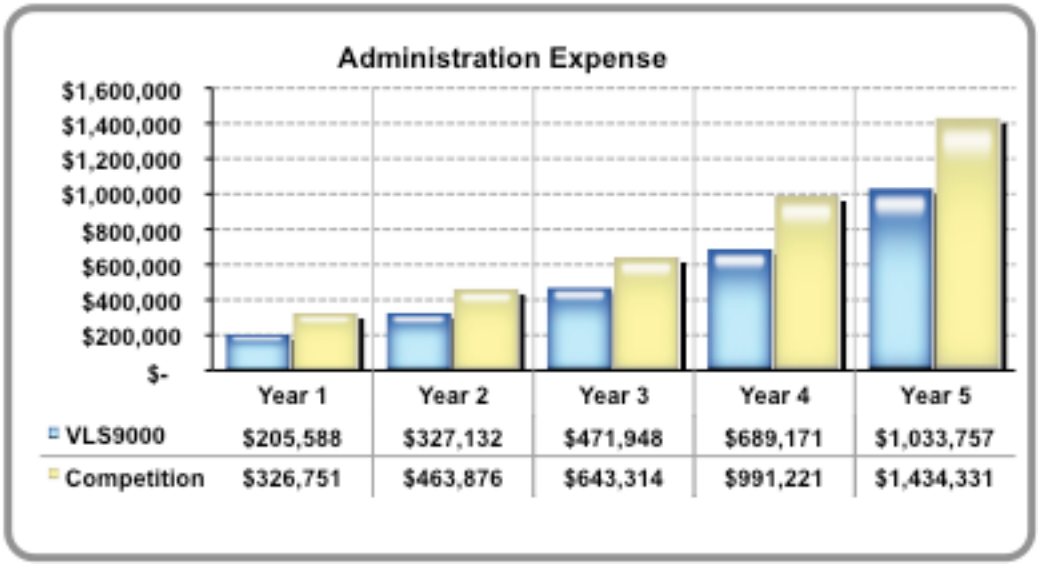


Figure 1 Enterprise Solution – Capital Expense

Silverton Consulting
Strategy, Storage & Systems
P.O. Box 492|Broomfield, CO 80038-0492|
720-221-7270|www.silvertonconsulting.com

Using the projected workloads, the enterprise model determines the deduplication product configurations and the resulting costs of the HP and competitor systems as shown in Figure 1. The year 1 capital expense includes only the acquisition cost of the configurations. For example, the required HP configuration consists of three system nodes for required performance throughput with three 40TB storage arrays, a 20-port FC switch with deduplication licensed. To meet end of year 1 storage requirements and using a much industry touted 20:1 deduplication factor, it is projected that 32TB of physical storage space will be needed. The superior performance throughput gives the HP product an advantage as seen in year 1 costs of $1,006,857 versus $1,525,100. After the first year both HP's and the competitor's systems require significant upgrades. Because of the new incremental configurations, HP's projected cost is higher than its competitor in years two, three and five but is expected to be significantly less over the cumulative five year projection period. Again much of this advantage results from the HP product's superior per node throughput performance.



	Year 1	Year 2	Year 3	Year 4	Year 5
VLS9000	$205,588	$327,132	$471,948	$689,171	$1,033,757
Competition	$326,751	$463,876	$643,314	$991,221	$1,434,331

Figure 2 Enterprise Solution - Administration Expense

Once more, using the projected workloads, the enterprise model determines a projected annual administrator cost over the next five years as shown in Figure 2. The actual computation of this value is complex and is based on three variables including administrator time/system, administrator time/backup and administrator time/TB stored. No definitive published numbers for these factors were available but modeled administration time is conservative. In all years HP's VLS9000 shows a distinct cost advantage due to its more integrated system architecture and management capabilities. For example, the HP system does not require separate management for its integrated frontend and backend system.

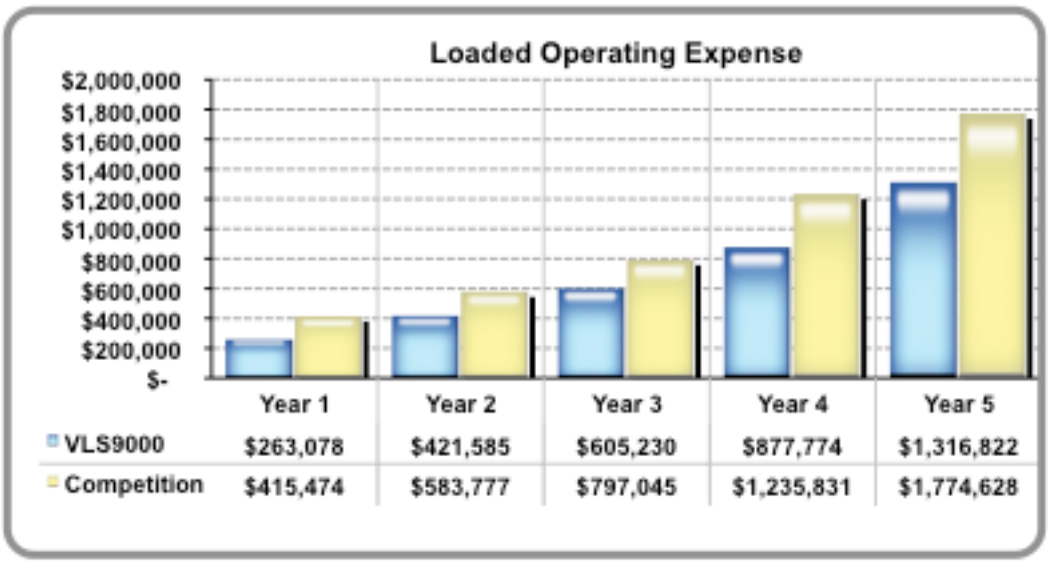


Figure 3 Enterprise Solution - Loaded Operating Expense

The loaded operating expense as shown in Figure 3 is a consolidated chart of FTE administrator, power and cooling, data center footprint, and system maintenance costs. As expected the FTE administrator costs from Figure 2 represent most of the total operating expense. The other factors including power and cooling and data center footprint costs are individually calculated and charted but have not been shown.

Because of the significance of FTE administrator costs, it is not surprising that HP once again shows a distinct advantage every year over its competition. However, in every category (e.g., power and cooling) HP enjoys a small, yearly advantage.

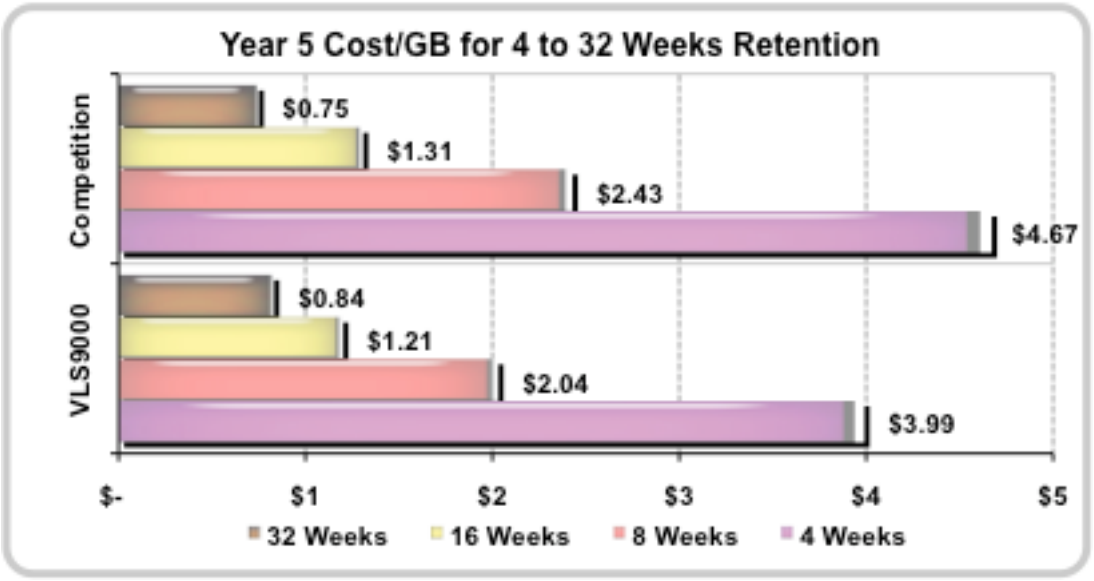


Figure 4 Enterprise Solution - Year 5 Cost/GB for 4 to 32 Weeks Retention

Figure 4, the final illustrated chart for the enterprise model, shows cumulative acquisition year one through five costs/GB of storage assuming retention periods of 4 to 32 weeks. Specifically, this output was calculated as the cumulative acquisition costs from years one through five plus the year five operating cost divided by the required storage needed by end of year 5 in GBs. With its overall capital expense advantage, the HP VLS9000 as shown in Figure 4 is the definitive cost leader over three of the four projected retention periods.

To reiterate, the model and its resulting charts were designed to provide easily understandable pictorial illustrations of various capital and operating costs given a certain potential customer specified workload. When modeling other workloads, HP's VLS9000 may not always be advantageous, particularly for each cost component for each and every year. However, it is expected that the VLS9000 will result in significant cost savings in a vast majority of realistic enterprise storage configurations – the overall architecture and performance capabilities are just too compelling.

A SMB Deduplication Example

Small to medium business (SMB) data centers typically store less data and so, support less intensive backup environments. Similar to the enterprise class model above, the SMB model starts with a prescribed customer workload. As such, the backup environment for the SMB example includes:

- 5TB of full backup data backed up once per week
- 1% of incremental backup data per day for five days during the week
- 20% annual backup data growth
- 8 weeks of backup data retained online and
- 20 hour full backup window

Given these requirements and projected storage growth, the storage requirements at the end of the next five years are as follows in Table 3. The workload calculated below more closely matches the HP targets for the D2D4112. However, the model only supports FC configurations; iSCSI configurations have not been considered.

	Year 1	Year 2	Year 3	Year 4	Year 5
YE full backup (TB)	6	7.2	8.6	10.4	12.4
Retained capacity (TB)	43	51	61	74	88
Required perf. (MB/s)	83	100	120	144	173

Table 3 SMB backup workload requirements

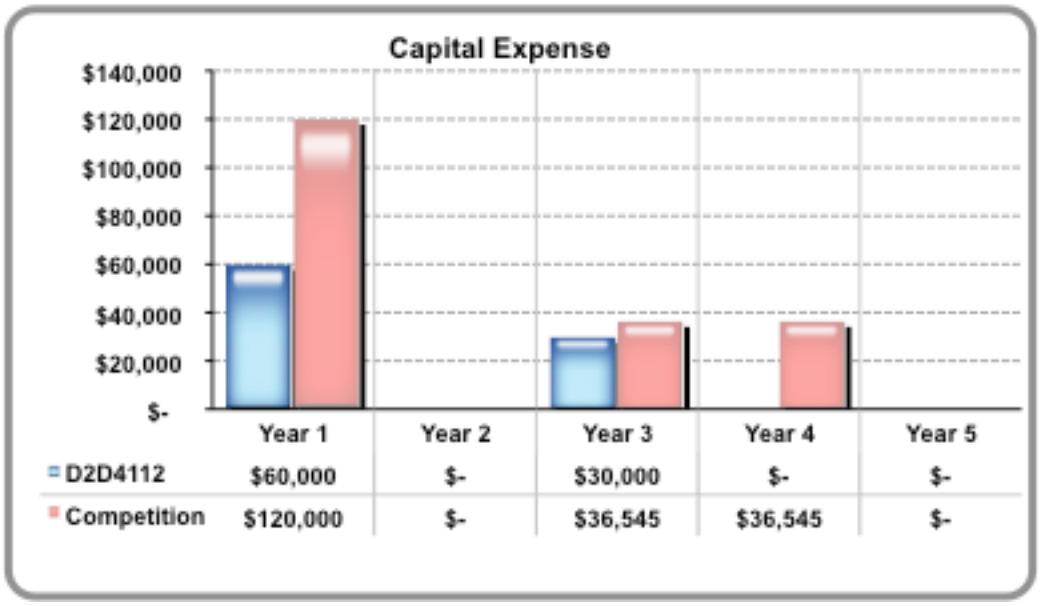


Figure 5 SMB Solution - Capital Expense

Again like the enterprise model, the year 1 costs represent the acquisition cost of the equipment configuration as determined by the model. The resulting HP configuration is a 12TB array system with 100MB/s deduplication performance. As seen, the initial acquisition cost of the D2D4112 is substantially less than the comparable competitor product. In year 3 additional capital outlays need to be made for performance rather than additional storage capacity. HP's D2D4112 only bests the competition by a small margin in that year. Over the five years of the model, however, the HP product retains a significant cost advantage over its competitor.

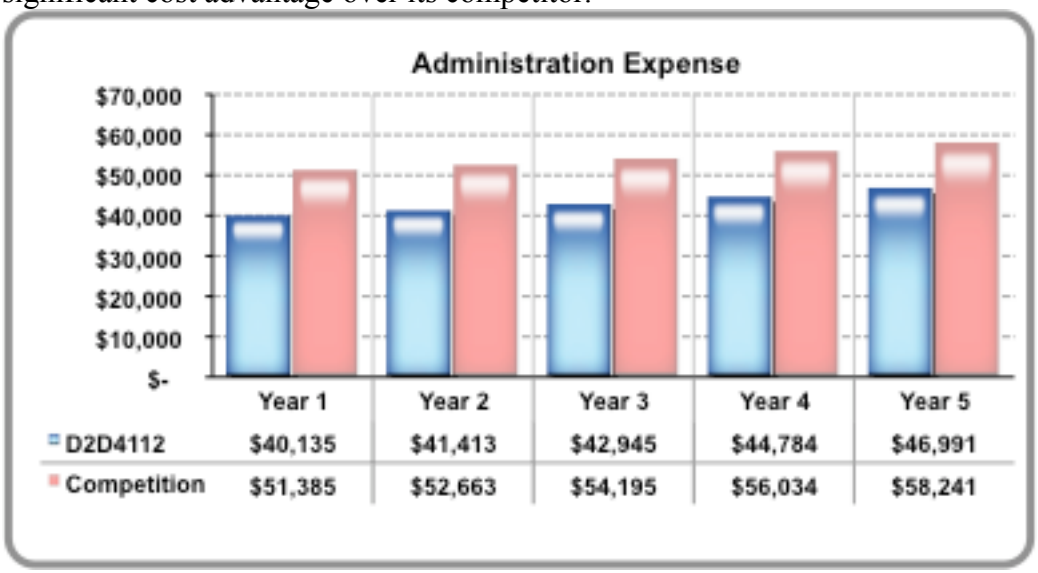


Figure 6 SMB Solution - Administration Expense

Based on the workload parameters of the SMB model, Figure 6 illustrates the annual administrator cost advantage of the HP product. Like the enterprise model, the actual estimated administrator cost is calculated from three variables, including administrator time/system, administrator time/backup and administrator time/TB stored. Again, no published numbers for these factors are available but modeled time is conservative. The HP D2D4112 shows a distinct cost advantage, albeit not as dramatic as the enterprise class, because of higher integration and less complexity.

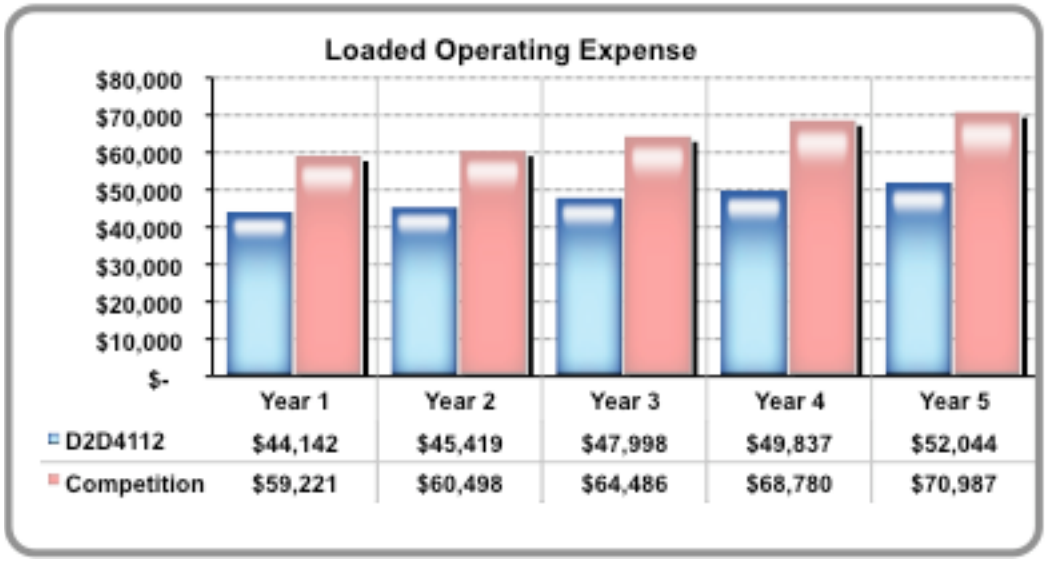


Figure 7 SMB Solution - Loaded Operating Expense

Figure 7 illustrates the combined annual FTE administrator, power and cooling, data center system footprint and system maintenance costs. Annual FTE administrator costs again represent a large percentage of these costs. Power and cooling and data center footprint costs are individually charted but have not been shown. As noted before, with the annual FTE administrator cost advantage, the HP product maintains its loaded operating cost advantage throughout the five year period and in fact, even improves its advantage slightly.

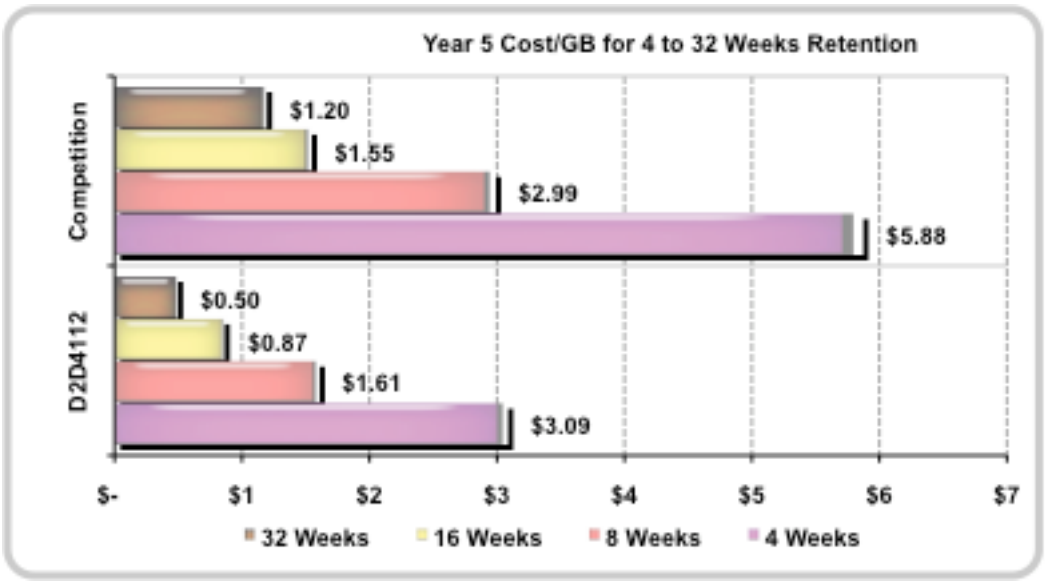


Figure 8 SMB Solution - Year 5 Cost/GB for 4 to 32 Week Retention

The final illustrated chart for the SMB model, Figure 8, was calculated as the cumulative acquisition cost years one through five plus the year five operating cost divided by the required storage needed by the end of year five. This calculation is then illustrated for four retention periods ranging from 4 to 32 weeks. Not surprisingly in each case the HP D2D4112 shows a significant overall cost advantage.

The foregoing charts illustrate that the D2D4112 is a cost effective solution to support SMB backup workloads. The SMB model realistically simulates the HP and the competitor offerings to meet customer specified workloads. The model may sometimes result in transient competitor cost advantages but the lower entry cost of the D2D4112 together with its greater capacity will likely provide the lowest cost solution when examined across a wide variety of realistic SMB workloads.

Low bandwidth replication

The models above conclusively showed that both the HP VLS9000 and the HP D2D4112 are outstanding choices for backup storage systems for their respective market targets. However, one factor often overlooked is the attractiveness of these systems for disaster recovery through the use of low bandwidth replication. With the deduplication and compression capabilities of both products, asynchronous replication becomes a realistic, viable disaster recovery alternative.

For example, in the enterprise model above, without deduplication the 75TB backup could be replicated over a 164.52 Mbs OC3 link in just under 1300 hours or ~7.5 weeks. However, with the VLS9000 deduplication capability using a 20:1 reduction, the 75TB backup workload could be compressed and deduplicated down to ~3.75TB of unique data, and replicated in just over 63 hours over the same OC3 link. Similarly, although not as dramatic, in the SMB model's 6TB backup, the data could be replicated over an

OC3 link in ~101 hours or ~4.2 days. However, on HP's D2D4112 with a 20:1 reduction, replicating this backup data would take just over 5 hours. As such, both HP products could perform asynchronous replication of large amounts of deduplicated data within a respectable timeframe at a reasonable cost.

Summary

The two models and resulting storage configurations clearly reflect that HP's VLS9000 and D2D4112 are tough, if not superior products in their markets. Not only do both products enjoy competitive acquisition price but also can be expanded competitively. Because of their less complex, more integrated architectures both products can be operated at an ongoing significant cost savings. Perhaps even more significantly though, the deduplication capability of both systems makes them an ideal, viable alternative to use low bandwidth asynchronous replication to support enhanced disaster recovery.

Silverton Consulting, Inc. is a Storage, Strategy & Systems consulting services company, based in the USA offering products and services to the data storage community.

Silverton Consulting
Strategy, Storage & Systems
P.O. Box 492|Broomfield, CO 80038-0492|
720-221-7270|www.silvertonconsulting.com