Optimizing TCO for Data Protection Storage

A long-term TCO analysis of Cohesity's DataPlatform in a data protection use case

John Webster

July, 2016

Evaluator Group

Enabling you to make the best technology decisions

COHESITY

Evaluator Group   COHESITY

Executive Summary

Many IT enterprise administrators are transforming their data centers to a services delivery engine that can simultaneously support existing applications while delivering new mobile, social, and analytics applications. They are also finding that they may have to first increase spending on servers, networking and both high performance and long-term storage. As this occurs, many have decided to invest in "future forward" technologies rather than simply spend more money refreshing existing ones.

While evaluating forward-looking storage technologies such as all-flash arrays, object storage, and distributed storage platforms for primary storage, IT administrators should not overlook the secondary storage environment, and in particular, data protection storage as part of the evaluation process. Key considerations include:

- New technologies are now being blended into data protection storage. How can they be leveraged?
- What are the most cost effective data protection platforms for today and for the future as technology advances?
- Is IT bound to repeat the same costly three- to four-year hardware refresh cycles of the past, or is there a better way to extend the useful life of the data protection infrastructure?

This paper reviews Cohesity's DataPlatform in the context of two use cases. First we review the use of Cohesity as an efficient distributed storage platform to replace traditional target storage devices and media for backup and recovery operations. Second, we look at the use of Cohesity as an integrated, end-to-end data protection solution for backup, recovery, replication and cloud tiering use cases. Our research presented here in the form of total cost of ownership (TCO) models shows that enterprise storage administrators can reduce their overall cost for data protection target storage by 50% when using Cohesity's DataPlatform to replace purpose-built backup appliances (PBBAs). But more importantly, it shows that using Cohesity's end-to-end solution that integrates backup and recovery software with the Cohesity DataPlatform can also reduce the initial first year TCO of a data protection environment—software included—by as much as 80%.

Our TCO modeling is presented using a 3-4-year time horizon as is typically done by IT administrators given the life of a storage platform before a technology refresh cycle is required. However, we note that Cohesity offers a long term data protection solution that can be scaled and upgraded in place without replacing the entire system. As a result, the Cohesity architecture can extend the time horizon used in typical TCO modelling exercises from 3-4 years to 7-10 years. Our long-term modelling includes a comparison with competing solutions that require platform replacement and data migration activities as part of the total cost of ownership vs. the Cohesity solution that does not. Doing so gives a more accurate picture of cost for long-term and sustainable data protection services delivery.

Cohesity DataPlatform

Cohesity DataPlatform is a distributed, highly scalable platform designed to function as a consolidation point for a variety of secondary data storage types and workflows in both virtual and physical server environments. Use cases include secondary storage for data protection, files, test/dev copies, cloud

tiering and cluster analytics. Native integration with public cloud resources is supported for long-term archival and disaster recovery applications.

The DataPlatform consists of consists of a distributed, shared nothing cluster of nodes that that can be scaled upward without limitation. For data protection applications, it can be used as a target data protection storage system in place of conventional deduplication appliances, or as a completely-integrated data protection system with the addition of Cohesity DataProtect that includes built-in services for backup and recovery without requiring the popular backup applications from vendors such as Commvault, Veritas and Veeam (see Figure 1 below).
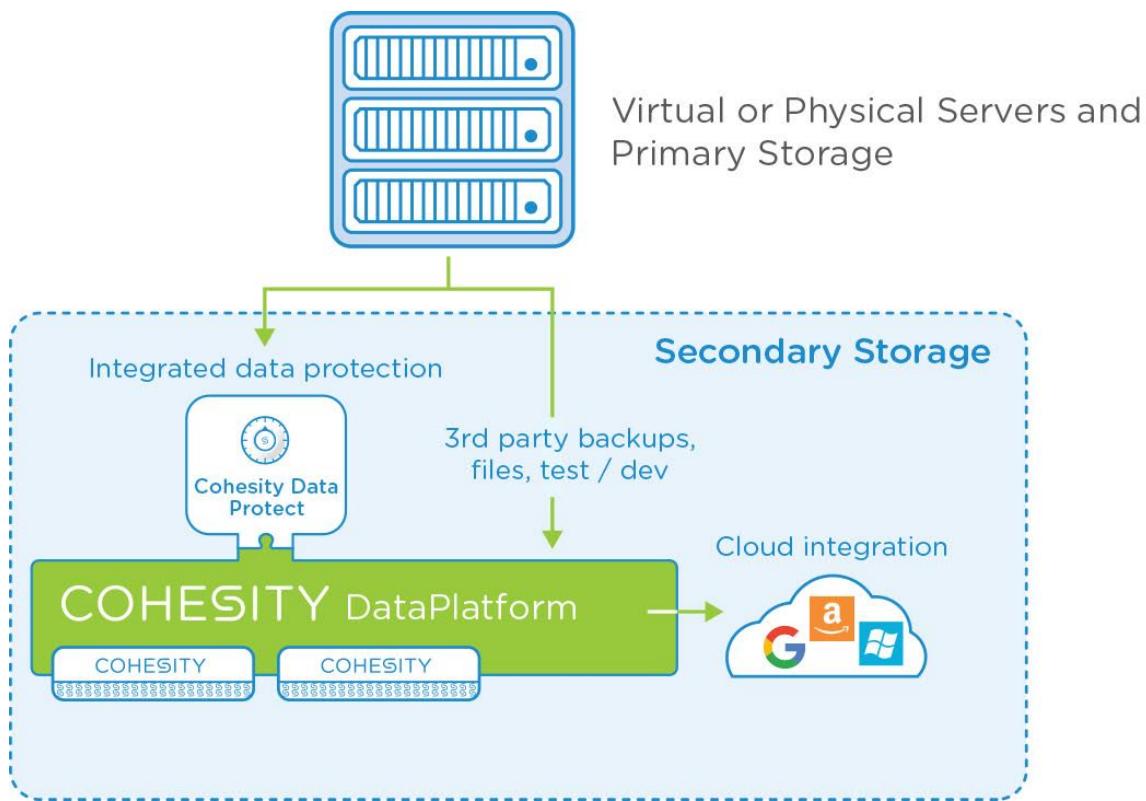

Figure 1. Basic architecture of the Cohesity scale-out DataPlatform as a secondary storage consolidation point consisting of clustered nodes. Cohesity's optional DataProtect software supports integrated data protection. Native cloud integration is also supported.

The Cohesity Data Platform can be deployed initially as a 2U system which includes 4 nodes, for a total of either 48 TB raw HDD capacity and 3.2 TB of flash (model C2300) or 96 TB raw HDD capacity and 6.4 TB of flash (model C2500). Each node also has two 10Gb SFP+ and two 10 GbE interfaces as well as an out-of-band management interface for remote configuration. (See Figure 2 below C2000 Series configuration details.)

Technical Specifications

Technical Specifications	C2300		C2500	
	Per Node	Per Block (4 Nodes)	Per Node	Per Block (4 Nodes)
Hard Disk Capacity*	12 TB	48 TB	24 TB	96 TB
PCI-e-Based Flash Storage*	800 GB	3.2 TB	1.6 TB	6.4 TB
Memory	64 GB	256 GB	64 GB	256 GB
Compute	2x Intel Xeon E5-2600 2.4 GHz 8-Core CPUs	8x Intel Xeon E5-2600 2.4 GHz 8-Core CPUs	2x Intel Xeon E5-2600 2.4 GHz 8-Core CPUs	8x Intel Xeon E5-2600 2.4 GHz 8-Core CPUs
On-Board Network Connectivity	2x 10 GbE 2x 1 GbE 1x IPMI	8x 10 GbE 8x 1 GbE 4x IPMI	2x 10 GbE 2x 1 GbE 1x IPMI	8x 10 GbE 8x 1 GbE 4x IPMI
Operating Environment	Cohesity DataPlatform			
Storage Protocol Support	NFSv3, SMB 2.1, 3.0 REST			
Data Protection Support	VMware VADP, Oracle RMAN, Microsoft SQL, Physical Windows and Physical Linux Servers			
High Availability	Native Site-to-Site Replication			
Cloud Archive	Google Nearline, Microsoft Azure, Amazon S3 & Glacier, S3 compatible			
Encryption	AES256			

Figure 2. Cohesity C2000 Series Configuration Details (Source: Cohesity)

Nodes can be added individually and without limitation if and when required for linear scaling of performance and capacity. Individual nodes can also be upgraded with new software and hardware. Both operations can be performed non-disruptively. This aspect of the DataPlatform leads to increased system longevity without the traditional technology refresh that require the addition of system images and/or disruptive data migrations. The resulting impact to total cost of ownership (TCO) is discussed and quantified later in this report.

Software Architecture

The DataPlatform is a commodity hardware-based, software-defined, cluster architecture with no single points of failure. The cluster provides compute capacity to run global deduplication across the cluster as well as data operations directly on the platform. The overall software architecture is diagrammed in Figure 3. Note once again that this graphic includes a plug-in module called "Cohesity DataProtect" which is an additional-charge add-on that allows administrators to converge backup software, infrastructure and target storage onto one unified platform (described in more detail below).
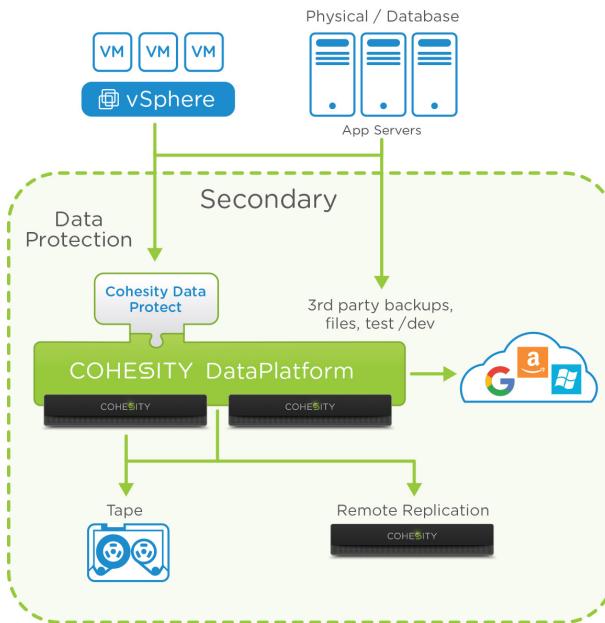
Figure 3. The Cohesity DataPlatform Architecture

Cohesity implements a shared nothing file system that supports full consistency as opposed to eventual consistency commonly found in other large scale distributed files systems that can lead to inconsistencies over time. On ingest, data is optimally placed on the tiered storage layers—flash and HDD—that are distributed across the cluster. All data is written to two nodes simultaneously for high availability and fault tolerance. In addition, all upgrades/replacements can be done without disruption to normal operation including hardware replacements and software updates. This is a core capability that leads to significantly lower total cost of ownership (TCO) over time as we will demonstrate below.

Significant features and work flow add-ons include:

> **Global deduplication and compression**—variable length, global deduplication spanning that features both in-line and post-process deduplication. Administrators can choose between in-line or post process deduplication or no deduplication as needed.

> **SnapTree**—snapshots are created using a tree of pointers to reduce latency rather than more conventional methods that traverse an entire snapshot chain when assembling and retrieving snapshot copies of data. This allows administrators to automate snapshot creation by scheduling snapshots on a minute-to-minute basis if required for near continuous data protection. Every node in a cluster can access any snapshot through a bounded pointer depth independent of the number of accumulated snapshot images.

> **Real-time analytics**—monitors capacity utilization and trends in real-time at the cluster, VM and file level.

Automated global indexing —tracks high-level indices such as files names and metadata as well as higher-level objects such VMDKs and VMs. It is used to enable rapid search and recover capabilities even at large scale.

Software-based encryption—for data at rest and in transit over a network. In transit encryption can be applied to data replicated to a remote Cohesity cluster or when data is tiered/archived to the cloud.

Replication and DR orchestration—backups generated by a source DataPlatform using SnapTree can be replicated to one or multiple Data Platform targets at remote sites for business continuance and disaster recovery operations. In addition, multiple sources can replicate to a single or multiple targets. Global deduplication is also supported for replication. DR Orchestration software automates the connection and discovery of virtualized infrastructure at target sites.

CloudTier—native cloud integration for off-site data tiering. Cohesity CloudTier currently supports Google Cloud Storage, Microsoft Azure, and Amazon S3.

CloudArchive—long-term off-site archival storage on leading cloud providers including Google Cloud Storage Nearline, Microsoft Azure Cool, Amazon S3 and Glacier.

Protocols—supported include NFSv3, SMB 3.0 and REST.

Deployment Options

As mentioned earlier, IT administrators are given the ability to deploy the Cohesity DataPlatform as a backup and recovery target system (Target Mode), as an integrated data protection system (Integrated Mode) with Cohesity DataProtect software, or deploy both modes on the same cluster. These two deployment options are not mutually exclusive and customers have the choice to use the Cohesity DataPlatform as a backup target for some applications, and as an integrated data protection solution for other workloads. Either mode provides all of the functionality of the DataPlatform as described above with the following exceptions:

1. At this point in time, Remote Replication, Analytics, and CloudTier can be bundled into the price of the DataPlatform at no additional charge.
2. CloudArchive is also available on request and can be included with DataProtect at no additional charge. However, it could become a separately chargeable add-on in the future.

Backup target (Target Mode):

In Target Mode, Cohesity DataPlatform provides all of the functionality and add-ons described above. In this mode, DataPlatform can be integrated with third party data protection software. Currently

supported applications include those from Commvault, Veritas (NetBackup and Backup Exec), Veeam, HP, EMC (Networker and Avamar), and IBM.

Cohesity DataProtect for Integrated Data Protection (Integrated Mode)

The acquisition and integration of Cohesity DataProtect with the DataPlatform allows IT administrators to remove the need for third party data protection software and hardware, leveraging instead the native data protection capabilities of Cohesity DataProtect which converges backup software, backup infrastructure and target storage onto one unified platform. With DataProtect, the Cohesity platform becomes an agentless data protection system for backup and recovery of virtual server environments that gain performance vs. traditional dedupe appliances because of the Data Platform's ability to parallelize backups. With Global Indexing, administrators can recover individual VMs, restore files to source VMs, and individual application objects for Microsoft Exchange, SQL and SharePoint. Recovery points are created quickly using SnapTree snapshots and clones. Recovery time can be minimized to near-instantaneous by creating a clone of a backup VM and running that clone directly from the DataPlatform or (in VMware environments) moving the clone back to primary storage via vMotion.

VMware, Windows servers, Linux servers, and MS SQL Server are supported natively. For VMware environments, full integration with vSphere and vCenter provides a view of VMs protected by Cohesity. Administrators can apply policies on a per VM basis. Support for vSphere APIs for Data Protection leverages SnapTree-based snapshot copy protection. For other applications, Cohesity provides a remote adapter facility for managing direct backups to a Cohesity cluster.

The Cohesity Data Protection TCO Advantage

A critical aspect of Cohesity's TCO advantage over many of the purpose-built backup appliances available today is the ability to non-disruptively update the DataPlatform cluster in a modular fashion over time. This means that IT administrators can retire cluster nodes from service and replace them with faster, denser nodes while data is transparently migrated to other nodes to maintain normal operation, and then rebalance data across the cluster once the new node is on line. In addition, administrators can upgrade individual nodes in place for additional performance and/or capacity.

A significant result of these capabilities will be that Cohesity DataPlatform customers can expect TCO savings over the short-term (3-4 years) resulting from greater capacity efficiency and the convergence of backup target, software and other infrastructure components with DataProtect as well as longer system life cycles (5-10 years) than those typically encountered in enterprise data center environments. With Cohesity, IT administrators can extend or replace the typical cadence of three- to four-year technology refresh cycles with ones that are longer in duration. These critical differences result in major economic benefits:

1. <u>Consolidation of Backup Software, Media Servers and Backup Targets</u>: Traditional data protection systems are typically a conglomeration of point products—often from different vendors—that are bought, managed, supported, and tracked separately because each has its own technology lifecycle to be concerned with. Depending on the capacity and scope of the data

protection environment, the hard cost of point products can include data protection software and associated servers, dedupe appliances and cloud gateways when a link to off-site cloud storage is required. Additional costs that are harder to quantify depending on the environment include those associated with managing a complex environment as well as increased risk of downtime due to multiple points of failure.

2. <u>Fewer system replacements</u>: Technology and capacity refresh cycles involving the replacement of an entire storage system become less numerous over time. In fact, because greater performance and capacity can be added on a module-by-module basis, the useful life of the Cohesity scale-out system can be 2-3 times that of a traditional purpose built backup appliance. This cuts the costs associated with refresh cycles in half while more closely aligning them with infrastructure amortization schedules. The ability to non-disruptively add/replace modules as new technology emerges and enters mainstream usage (SSDs, higher performance I/O ports, etc.) also helps to ensure the long-term quality of service delivery without disruption. This enhances IT administrative productivity.

3. <u>Possible elimination of data migration costs</u>: The significant costs associated with migrating data from an obsolete system to a new system are greatly reduced over time and possibly eliminated when cluster nodes can be swapped-out or older storage media needs to be replaced while data is redistributed across the cluster non-disruptively. These costs include those related to system downtime and a corresponding loss in productivity, as well as an extension of lease terms at premium rates when the replaced system is leased and not returned on schedule.

Short and Long-term TCO Comparison of Cohesity vs. Competing Solutions

As mentioned earlier, the DataPlatform can used as a large scale, secondary storage consolidation point. We've also highlighted the fact that its scale-out cluster architecture can be upgraded in place and without disruption. Upgrades include the installation of new software revisions, node additions and hardware replacements. The resulting total cost of ownership (TCO) benefits of this scalable, non-disruptive, upgrade-in-place architecture are demonstrated in Figures 4 through 9 below that show the results from Evaluator Group's TCO model.

We first compared the commonly used 3-year TCO of Cohesity's DataPlatform to competitive solutions. Figure 4 shows the DataPlatform in backup target mode vs. a popular competitive purpose built backup appliance PBBA—also often called a dedupe appliance. Here we found that significant TCO savings are achievable. We then ran the same TCO model calculations comparing the DataPlatform in integrated mode with the optional DataProtect software (Figure 5) vs. competing solutions that include backup software and a PBBA. However, even greater savings were seen in integrated mode—up to 80% in the first year. Finally, we ran the same Cohesity vs. competing solutions comparisons using a ten-year time horizon (Figures 7 through 9) instead of three years to demonstrate the impact of a non-disruptive upgrades to the platform in place on long-term TCO. Again we saw significant savings of over 60%.

TCO comparisons of 3-years and 10-years were done using pricing that included the additional charges for Cohesity Protection in integrated mode. Note that three of these model calculations are set for the commonly-used 3-year time horizon. However, we also ran 3 additional models set for a ten-year time horizon in order to emphasize savings when the costs of infrastructure refresh cycles and data migrations requited for the competing solutions are included in the TCO model.

Storage workloads modeled include a mix of VMs, structured data and unstructured data. Assumptions include:

- Year 1 VM storage of 80 TB/10% growth rate per year with 10 full backups retained and data reduction ratio of 6:1
- Year 1 structured data storage of 20 TB/10% with growth rate per year with 5 full backups retained and data reduction ratio of 3:1
- Year 1 unstructured data storage of 100 TB/25% growth rate per year with 10 full backups retained and data reduction ratio of 5:1
- Software cost is perpetual license charged on the basis of stored capacity
- Data reduction is deduplication and compression
- Annual Support cost of 20% per year
- Life span of 3 years for the competing Purpose-Built Backup Appliance (PBBA)
- Cost of Cohesity DataPlatform model C2500 used for all calculations

As noted above, there is an additional charge for storage capacity when cluster nodes, partial nodes, or entire clusters are used under Cohesity DataProtect. Even so, we believe that the underlying scale-out architecture as well as the consolidation of the entire data protection environment onto a single platform forms the basis for a significant TCO advantage when compared to that of a more traditional data protection system composed of individual elements from multiple vendors that have to be integrated and supported individually by administrators. Calculations reflect mixed workload conditions (VM support, structured and unstructured data). The yearly and cumulative models depicted below show comparisons for Cohesity in the Target Mode as well as the Integrated Mode of operation. All models show a significant reduction in TCO that results from Cohesity's scale-out, software defined architecture and Global Deduplication and Compression implementation.

Three-year TCO Comparisons Cohesity Data Platform in Target and Integrated Modes—Yearly and Cumulative

Evaluator Group first compared the three-year TCO of a popular PBBA target to the Cohesity DataPlatform in target mode. We then ran the same models for Cohesity in integrated data protection mode (DataProtect software costs included) vs. a combination of a popular third-party backup software suite running on a dedicated server and the same PBBA used in the target mode calculation. Yearly and cumulative costs are modeled.

Target Mode Yearly

The yearly cost comparison in Figure43 below shows that the Cohesity DataPlatform in target mode operation (red plot) costs less than half of that of the PBBA over the first year of operation while the 3-year TCO differential is even greater. Costs for the popular competing PBBA accelerate in year three to account for an increase in capacity that requires an additional PBBA. This is so because the PBBA chosen for this model does not scale linearly as does the Cohesity DataPlatform.

Cohesity Target Mode vs. Competing PBBA, Yearly TCO
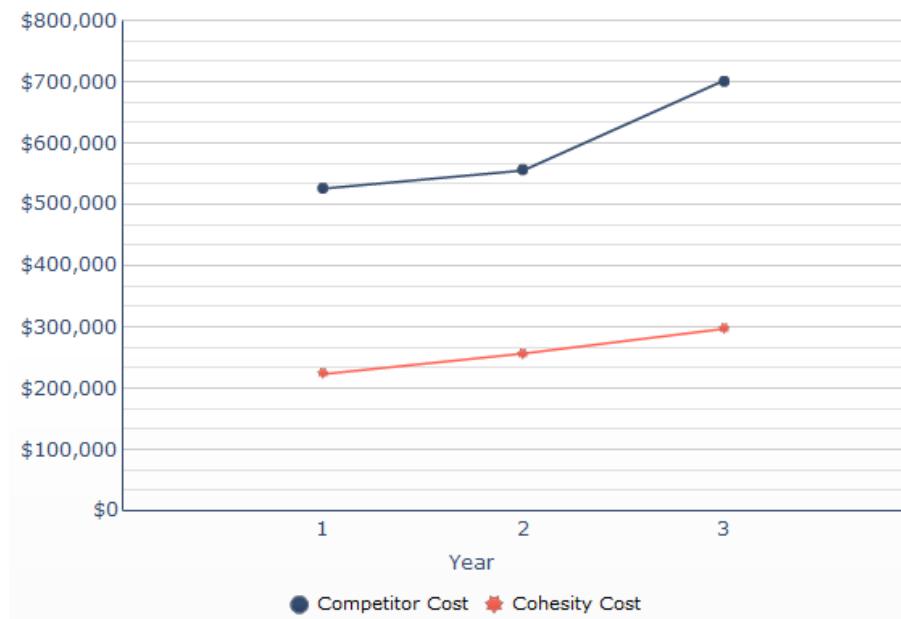
Figure 4 – Comparison of 3-year TCO of Cohesity DataPlatform (red plot) as a backup target vs. a competing PBBA as calculated on a yearly basis.

Integrated Mode Yearly

The yearly comparison in Figure 5 below shows that the Cohesity DataPlatform (red plot) in Integrated Mode operation costs substantially less than that of a competing solution consisting of a popular data protection application plus the same PBBA over the first year of operation. This is due mostly to the fact that the competing third party software licensing costs are eliminated. We note here that costs of up to 80% can be realized in the first year of operation depending mostly on the cost of the data protection software suite.

Cohesity Integrated Mode vs. Competing Data Protection Set, Yearly TCO
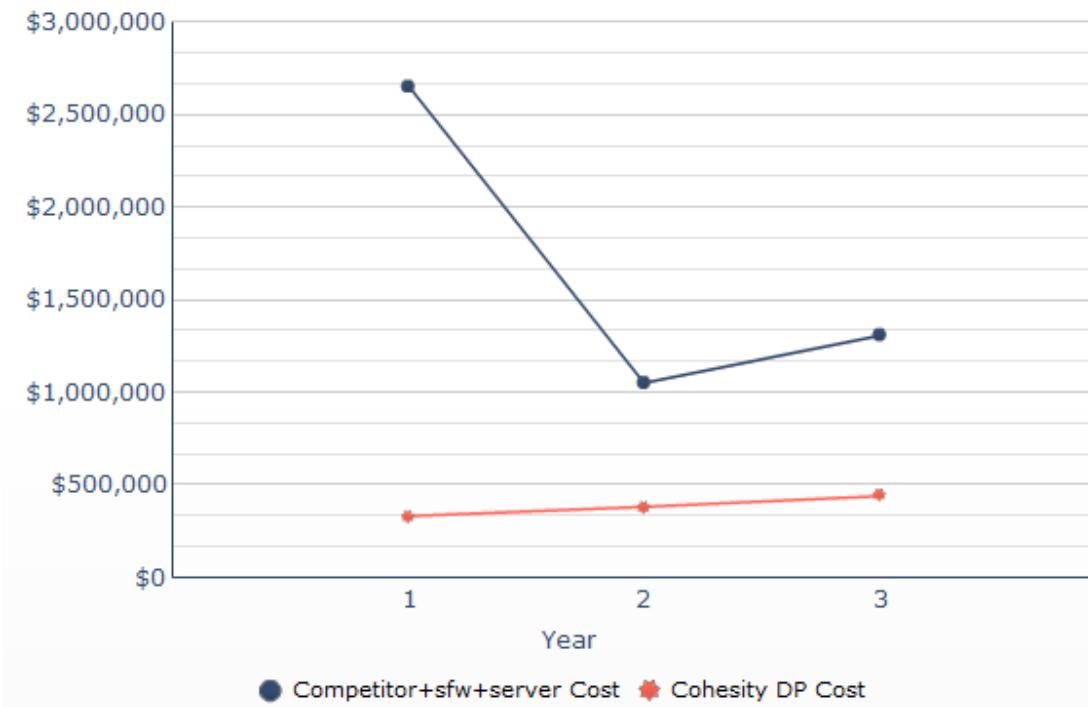
Figure 5 - Comparison of 3-year TCO of Cohesity DataPlatform as an integrated data protection system vs. a competing hardware and software elements as calculated on a yearly basis

Target and Integrated Modes Cumulative

A comparison of cumulative costs with Cohesity in both target and integrated modes (Figure 6 below) also shows that the most significant savings achievable with the Cohesity DataProtection solution of up to 80% over a three-year period (blue plot) is possible vs. the competing hardware/software point product solution (red plot).

Cohesity Target and Integrated Modes vs. Competing PBBA and Data Protection Set, Cumulative TCO
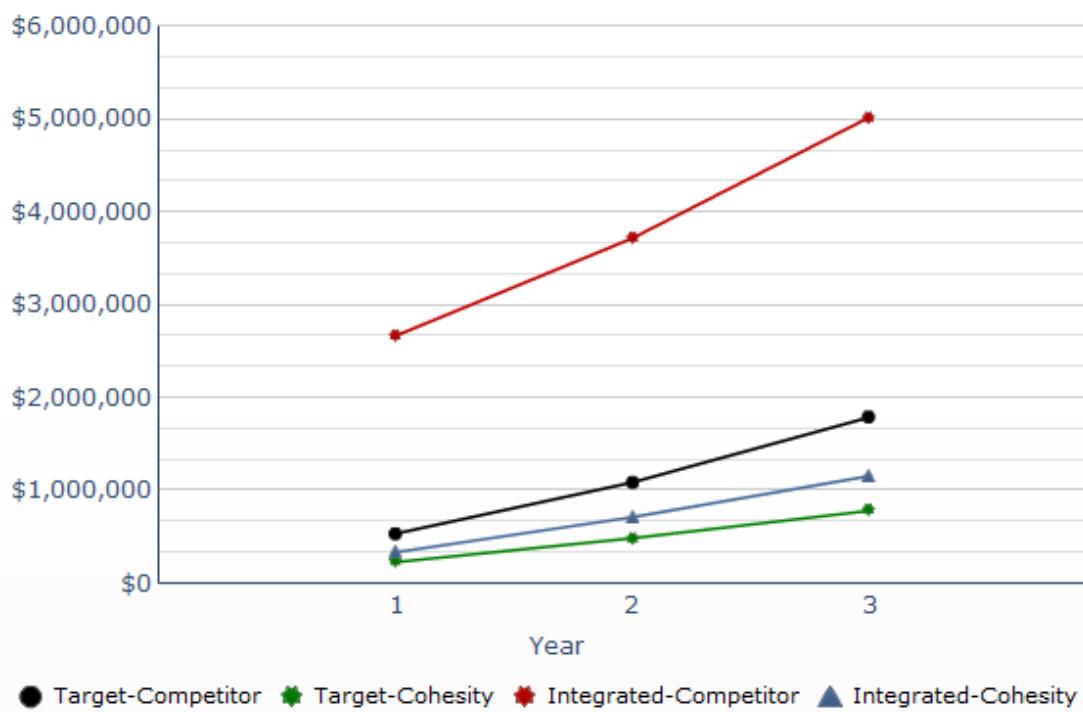
Figure 6 - Comparison of 3-year TCO of Cohesity DataPlatform in both Target and Integrated Modes as vs. competing solutions as calculated on a cumulative basis

Ten-year TCO Comparisons of Target and Integrated Modes—Yearly and Cumulative

As stated earlier, Cohesity offers a long term data protection solution that can be scaled and upgraded in place without replacing the entire system. As a result, the Cohesity architecture can extend the time horizon used in typical TCO modelling exercises from 3-4 years to 7-10 years. Our long-term modelling as depicted in figures 7through 9 below includes a comparison with competing solutions that require platform replacement and data migration activities as part of the total cost of ownership vs. the Cohesity solution that does not. We believe that doing so gives a more accurate picture of cost for long-term and sustainable data protection services delivery and allows the enterprise IT administrator to see that overall cost for long-term secondary storage can be reduced by more than 60% over a ten-year period as well.

Target Mode Yearly

### Cohesity Target Mode vs. Competing PBBA, Yearly TCO
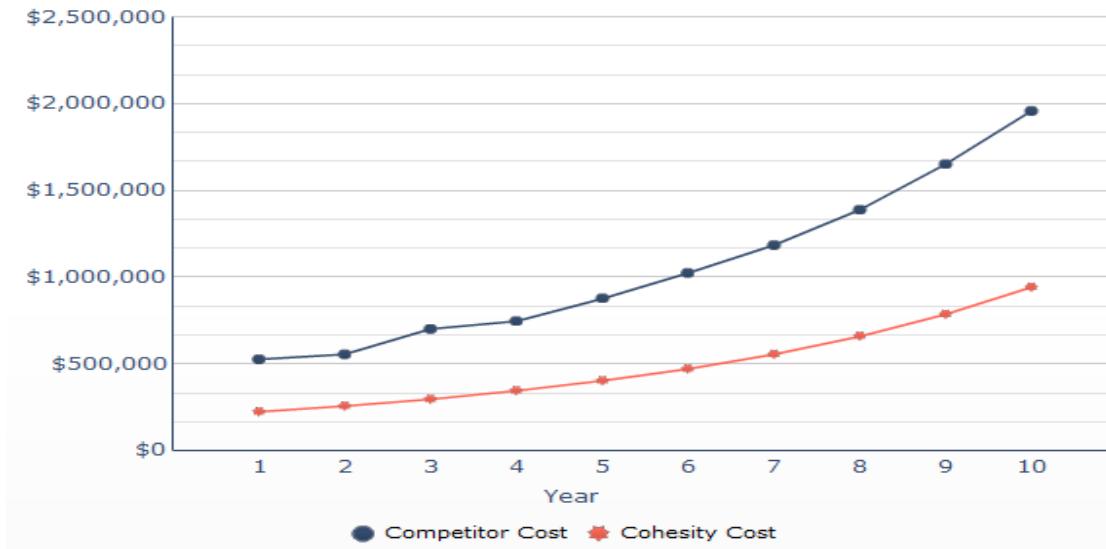
Figure 7: Ten-year TCO comparison of Cohesity's Data Platform in Target Mode vs a competing PBBA computed on a yearly basis. Note that the yearly TCO over a ten-year period is roughly half that of the competing PBBA throughout the operational lifecycle.

Integrated Mode Yearly

Cohesity Integrated Mode vs. Competing Data Protection Set, Yearly TCO
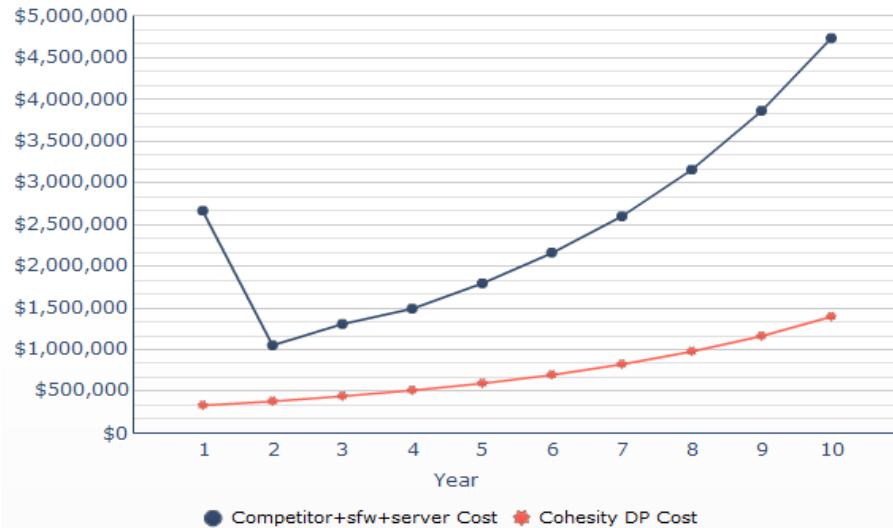
Figure 8: 10-year TCO comparison of Cohesity's Data Platform in Integrated Mode vs a competing solution. Note that that the yearly TCO of the point-product set rises at a greater rate than does Cohesity's integrated DataProtection solution over the ten-year period with savings accelerating beyond five years. This is due Cohesity's ability to upgrade and add capacity linearly without requiring the costs of hardware refresh cycles as well as the fact that the competing third party software licensing costs are eliminated.

Target and Integrated Modes Cumulative

Cohesity Target and Integrated Modes vs. Competing PBBA and Data Protection Set, Cumulative TCO
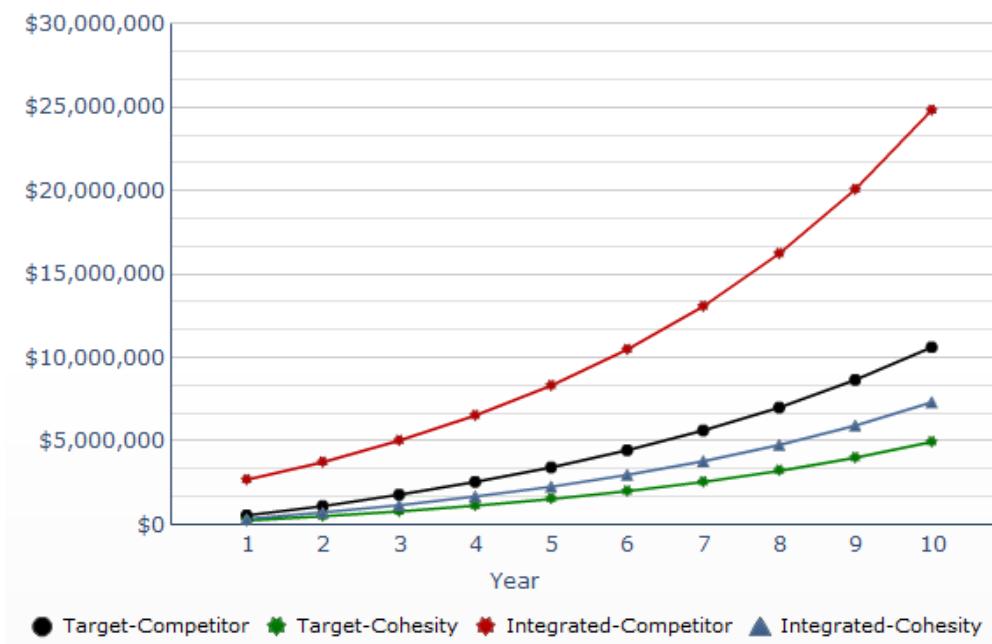

Figure 9: Ten-year cumulative TCO comparison of Cohesity's Data Platform in data protection Target and Integrated Modes vs. a competing solution. Again we note a similar cost acceleration for the point product set vs. the Cohesity DataProtection solution and for the same reasons as noted in Figure 7.

Conclusion

We believe that Cohesity offers a unique secondary storage platform because of its scale-out, distributed cluster architecture and capacity-optimized operational environment. This combination allows Cohesity customers to define either parts or an entire cluster as a self-contained data protection platform under Cohesity Protection as well as other applications including test/development and analytics.

Here we have highlighted the notable viability of Cohesity's DataPlatform as an alternative the current "business as usual" approach to enterprise data protection systems and in particular, the popular PBBA. Based on our review of the Cohesity architecture and software suite, we have shown that significant TCO savings can be realized by deploying the DataPlatform as a data protection solution. And while the DataPlatform can be deployed in Target Mode, we believe that most customers will gravitate to Integrated Mode for additional cost savings resulting from the replacement of third party suites with Cohesity's native data protection software.

About Evaluator Group

*Evaluator Group Inc. is dedicated to helping **IT professionals** and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver **in-depth, unbiased analysis** on storage architectures, infrastructures and management for IT professionals. Since 1997, Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and **education**. **www.evaluatorgroup.com** Follow us on Twitter @evaluator_group*