Technical Insight

Datrium Open Converged – Architecture for Performance

Author: Russ Fellows

March 2018

Evaluator Group

Executive Summary

IT organizations are continually striving to reduce complexity, while increasing their ability to respond quickly to application needs. One of the trends that has emerged is the move towards using converged and hyperconverged systems which helps reduce the integration and management overhead of traditional IT infrastructures. These trends are part of the drive towards reducing complexity and enabling faster time to revenue for the business.

Converged architectures began as pre-integrated best of breed components, a trend which continued with hyperconverged which includes utilizing servers and software defined storage to further consolidate hardware while increasing simplicity. Hyperconverged is focused on simplicity, via further integration and collapsing storage functions to reside wholly within compute nodes.

However, IT organizations have faced challenges with both approaches for different reasons. While hyperconverged architectures are easily deployed, companies have been reluctant to use these systems for business-critical applications, due to the need for high reliability and performance. In contrast, converged systems offer greater scale and flexibility, but do not offer the simplicity or ease of management promised by hyperconverged architectures.

Evaluator Group is an analyst firm focusing on Enterprise IT needs, conducting hands-on validations in our in-house lab as well as in-depth research into emerging trends and technologies. Working with Evaluator Group, Datrium conducted a large-scale benchmark of their DVX system to validate the performance capabilities of their system. The benchmark, released in October 2017, shows a scale-out cluster of Datrium systems able to support a world record number of VMs, with results available on the iomark.org website.[1]

Additionally, Evaluator Group conducts research utilizing interviews from IT users and has recently published a report on how enterprises are utilizing hyperconverged systems, along with IT users' concerns and decision criteria for choosing these products. As shown on the following page in Figure 1, respondents indicated that performance was their top concern, followed by the system's cost, with scalability the third most important issue.

This paper explores several issues:

- Requirements for converged and hyperconverged infrastructures (HCI)
- Architectural aspects of Datrium DVX that provide large scale performance and resilience
- Considerations for evaluating converged and hyperconverged systems

[1] http://www.iomark.org/content/datrium-announces-new-record-iomark-vm-hc-results

Enterprise Infrastructure Requirements

Evaluator Group conducts in-depth research studies, with one recent study examining hyperconverged usage in the enterprise and Enterprise IT's opinions on hyperconverged based on responses from over 100 IT professionals in companies with more than 5,000 employees.

One of the more interesting results pertained to the primary factors that go into how IT staff decides between hyperconverged solutions. The most important aspect was performance, followed by cost effectiveness, scalability and manageability. Evaluator Group's hyperconverged research study may be found at Evaluator Group website.[2]

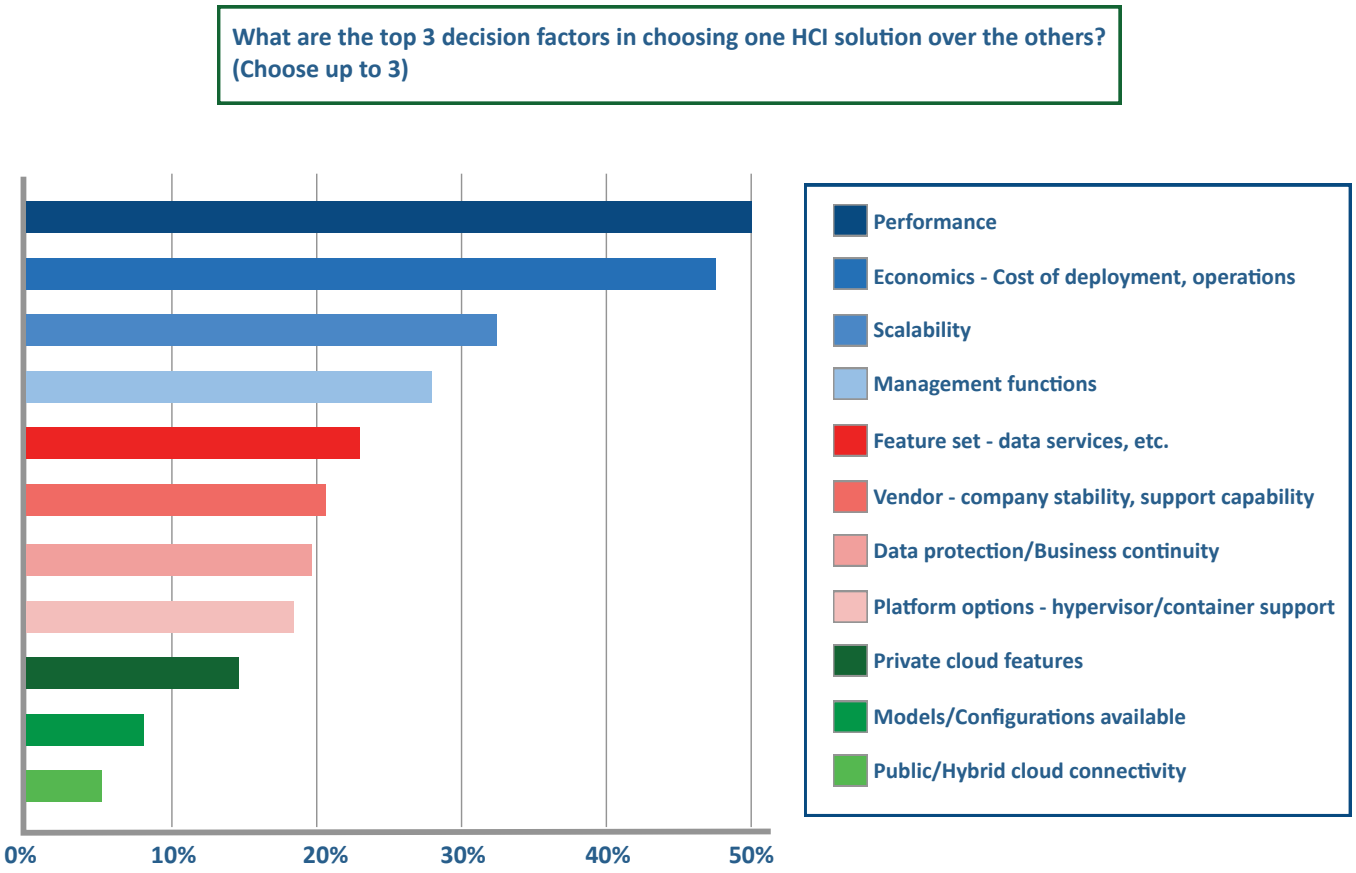

Figure 1: Enterprise IT Criteria for hyperconverged

[2] hyperconverged Infrastructure in the Enterprise: https://www.evaluatorgroup.com/hyperconverged-infrastructure-in-the-enterprise/

Also cited in this study, is the concern for overall resiliency and failure recovery. These concerns are valid, as Evaluator Group's own lab testing has shown that many typical hyperconverged systems have significant performance issues when a single node fails. Failures can arise for a variety of reasons, but the failure of a single SSD cache device can be enough to cause an entire node to be placed into a failed status, requiring data rebuilds on the remaining systems.

Evaluator Group Comments: Feedback from IT clients and survey participants, along with in-house testing of hyperconverged systems shows that failure of a single node can significantly impact the performance of a hyperconverged cluster. Additionally, error recovery can result in an elevated workload on the remaining hyperconverged systems, leading to further degradation and potential outage. As a result, some Enterprise organizations are reluctant to utilize hyperconverged for their mission critical applications.

DVX Architectural Elements

Datrium's DVX architecture utilizes software resident on each hyperconverged-like node that provides I/O control for storage, directing read requests to local solid-state media and directing writes to a scale-out pool of Datrium data nodes. This unique, split-path architecture along with the scale-out design is able to address some of the issues commonly experienced with hyperconverged systems, specifically for resiliency and large scale performance capabilities. The architectural features of Datrium include the following elements:

- DVX Compute Nodes: Scale-out servers with DVX Software and local SSD capacity which provide VM and IO processing, supporting 1 to 128 compute nodes
- DVX Data Nodes: Scale-out durable storage capacity, supporting from 1 to 10 data nodes and up to 1.7 petabytes for a cluster
- DVX Software: Software component on each compute node that provides storage and data services including erasure coding, compression, deduplication, snapshots, replication, encryption, as well as I/O redirection and failure recovery
- DVX management: Software includes provides management for a DVX cluster, includes CLI, APIs, a web-based UI console as well as a VMware vCenter plug-ins

The scale-out architecture increases read performance as compute nodes are added to the cluster, while capacity and write performance scale as additional DVX data nodes are added. With separate read and write data paths, it is possible to scale read and write performance independently. Typically, read performance is the critical factor in many applications, including transactional databases, while database logs and other applications may rely upon write performance. The Datrium architecture utilizes the DVX Software along with compute CPU, RAM and solid-state media to provide scalable read performance and other functions as compute nodes are added.

Datrium Architecture for Performance

Performance claims are made by many companies, often with persuasive arguments about how their design can provide better results than alternatives, but then offer little or no proof. Datrium has validated their performance and scalability claims by publishing benchmark results, which provide validation of their claims.

Evaluator Group worked with Datrium to perform a validated benchmark using a Datrium DVX cluster supporting 8,000 virtual machines. Utilizing standard SATA SSD's together with Datrium's all-flash storage nodes, the Datrium cluster ran the IOmark-VM benchmark. Datrium DVX performance exceeded the results for the previously reported best HCI system by 10X and surpassed the best all-flash storage system by 5X, with all results and test details available on IOmark website.

The IOmark-VM benchmark is an industry standard workload that provides a way to directly compare both performance and price-performance of storage and hyperconverged elements. Shown below in Figure 2 are the published results for IOmark-VM's for Datrium compared to the two closest competitors.

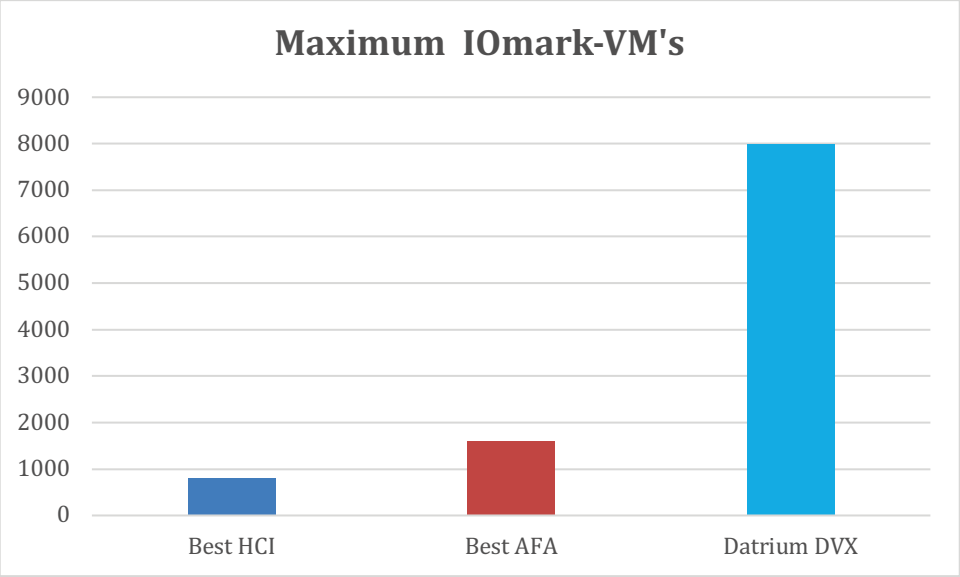

Figure 2: IOmark-VM Benchmark Results

The audited benchmark results show a Datrium DVX system comprised of data nodes and servers (compute nodes) with DVX software achieved passing results for 8,000 IOmark-VM instances at a price of $667.01 / virtual machine[3]. Key features noted during testing include:

[3] www.iomark.org

- A Datrium cluster consisting of 10 DVX data nodes and 60 servers running DVX software was certified to support an 8,000 IOmark-VM virtual server workload
- The tested configuration utilized both data deduplication, compression and erasure coding while also having encryption enabled
- More than 96% of response times were less than 5 milliseconds, with an average read response time of 0.76 milliseconds, one-third lower than the leading HCI and AFA results

IOmark-VM measures storage performance of virtual machines in both traditional and hyperconverged system architectures. This enables comparing performance results of storage only systems to converged systems, although price-performance comparisons are not possible. The 8,000 virtual machine results by Datrium was the largest number of VMs supported by either an HCI or a storage-only system. Moreover, Datrium's performance shows that a Datrium cluster can outperform up to five of the fastest all-flash storage systems tested to date.

Datrium's Architectural Details

Datrium's approach to converged infrastructure and enterprise storage requirements delivers features that are difficult to achieve with traditional architectures, such as scale-out performance and server and SSD fault tolerance, by leveraging standard servers and solid-state devices to accelerate applications. Their "Open Converged" architecture provides the scale and resiliency features of enterprise storage, combined with the simplified management and building block approach inherent with hyperconverged systems.

The ability to provide a scale-out converged system architecture is a challenge that many vendors are still working to meet. Datrium is able to scale capacity and performance of durable storage up to a 10-node cluster, while also have the unique ability to scale data processing and read caching to 128 compute nodes using hundreds of solid-state media all in a single cluster.

One of the key factors for Datrium's performance is their split-path architecture, which separates the read and write I/O paths. With this design, application read requests are provided within each compute node, delivering I/O as close to the application as possible, while writes are cached locally as well as are sent to Datrium DVX data nodes. Additionally, this design eliminates another potential issue with hyperconverged architectures, the inter-host network traffic between compute elements, which impacts latency and scalability during normal operations, and can significantly impact performance during failure and recovery scenarios.

Each DVX data node contains NVRAM for high-speed write acknowledgement to compute nodes, with all data mirrored to the second data node controller to ensure no loss of data. Data nodes provide read access data for any data not resident on a compute node, due to a cache miss or if a compute node fails or goes offline. Thus, performance during errors or recovery only degrades to that of a cache miss.

With support for multiple NVMe or SSD media, individual compute nodes can be configured to support specific application needs that may differ from other nodes in the DVX cluster. Each cache media within a compute node increases the nodes availability and HA capabilities. Write performance is scaled across data nodes, with up to 10 data nodes supported in a cluster. Thus, the ability to support scale-out writes across multiple DVX data nodes, combined with the ability to support up to 128 compute nodes with local cache in a cluster makes Datrium's one of the most scalable systems available.

Another critical aspect of the DVX architecture is the HA design, with no single point of failure. Each data node is similar to a traditional, dual controller system with redundant, hot swappable components and erasure-coded data protection. Additionally, the compute resident DVX Software is designed to fail gracefully by utilizing SSDs for read caching, and then using other compute node resources if all local resources have failed ("Peer Cache Mode"). Perhaps most importantly, loss of multiple compute nodes does not result in loss of data, or reduced availability for remaining compute nodes accessing the DVX data nodes. The unique scale and resiliency characteristics make Datrium's converged architecture well suited for enterprises that demand performance, scale and availability for their applications.

Similar to traditional storage, DVX data nodes are designed to ensure that writes are retained, regardless of any failure occurring after data is acknowledged. This design, along with the failover controller in each data node ensures resiliency and availability. DVX also provides data integrity checks four times daily of all data on the system, ensuring data consistency. While consistency is the most important feature of storage, read operations are often critical to application performance. The contention between data consistency and servicing application reads can result in performance issues within traditional storage systems, even with all-flash designs. By moving read operations to the compute node, Datrium data nodes are optimized for low write latency, while the DVX software combined with solid-state media within compute nodes provides low latency read operations and regular data integrity checks.

This unique approach to converged infrastructure enables Datrium to scale down to small configurations as cost effectively as hyperconverged architectures, while maintaining the ability to scale to multiple rack clusters to meet enterprise requirements. As with other converged architectures, Datrium supports both virtualized VMware and Linux systems as well as bare-metal Linux container compute nodes for operational flexibility. Finally, DVX management is integrated within vCenter for a single point of management simplicity of a hyperconverged solution and offers a built-in GUI on each data node.

Summary

Business executives and IT professionals understand the need to modernize how IT services are delivered, while reducing complexity and the time to value from IT systems. One of the ways to modernize while reducing complexity is by deploying an all-flash converged system, designed to run hybrid cloud applications. While IT professionals desire the simplicity and rapid scaling features inherent with hyperconverged architectures, concerns remain about their performance, scalability and reliability. Many

converged architectures are inflexible, with only a moderate improvement in management simplicity compared to traditional IT.

Datrium's open converged architecture is different than other solutions, providing an architecture that maintains durable data on highly available scale-out data nodes, combined with server resident resources to provide data services and application acceleration on scale-out compute nodes. The Datrium converged architecture provides proven performance and rack-scale clustering while ensuring storage availability meets enterprise applications demand.

By placing application data within server resident resources, the DVX architecture is a highly scalable architecture that also provides the reliability of traditional enterprise storage. The ability to cluster multiple DVX data nodes provides scalable capacity and write durability within the architecture. The Datrium design is highly reliable and just as important, highly available even when SSDs or compute nodes fail. These features, combined with the ability to scale-out writes on 1 - 10 data nodes and reads across 1 - 128 compute nodes provide the scalability and flexibility necessary to optimize DVX for specific application needs.

The Datrium DVX architecture addresses most concerns of Enterprise IT users, while offering reliability and the highest performing system Evaluator Group has tested to date. Additionally, the unique split-path converged architecture is flexible and scalable while providing the simplicity of a hyperconverged solution. The published benchmark results, along with architectural characteristics of Datrium DVX, show that Datrium's approach to converged infrastructure can address the top concerns many IT users have when choosing infrastructure for their business needs.

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*

This document was developed with Datrium funding. Although the document may utilize publicly available material from various vendors, including Intel and others, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.