

# DataStax Enterprise on VMware vSAN for Development

## The Rise of Modern Cloud Applications

Modern cloud applications—applications that create and leverage real-time value and run at epic scale—require a change in data management with an unprecedented transformation to the decade-old way that databases have been designed and operated. Requirements from cloud applications have pushed beyond the boundaries of the Relational Database Management System (RDBMS) and have introduced new data management requirements to handle always-on, globally distributed, and real-time applications.

## Hyper-Converged Infrastructure for Cloud Applications

While local fast storage provides optimal performance for distributed systems, it can be challenging for customers to manage with globally distributed cloud applications because they have outgrown their SAN-backed relational databases.

VMware vSAN™, the market leader in Hyper-Converged Infrastructure (HCI), enables low cost and high performance next-generation HCI solutions, converges traditional IT infrastructure silos onto industry-standard servers and virtualizes physical infrastructure to help customers easily evolve their infrastructure without risk, improve TCO over traditional resource silos, and scale to tomorrow with support for new hardware, applications, and cloud strategies. The natively integrated VMware infrastructure combines radically simple VMware vSAN storage, the market-leading VMware vSphere® Hypervisor, and the VMware vCenter Server® unified management solution all on the broadest and deepest set of HCI deployment options.

DataStax is the always-on data management platform industry leader. DataStax Enterprise (DSE) is the always-on data platform powered by the industry's best distribution of Apache Cassandra™—and includes search, analytics, developer tooling, and operations management, all in a unified security model. DSE makes it easy to distribute your data across datacenters or cloud regions, making your applications always-on, ready to scale, and able to create instant insights and experiences. Your applications are ready for anything—be it enormous growth, handling mixed workloads, or enduring catastrophic failure. With DSE's unique, fully distributed, masterless architecture, your application scales reliably and effortlessly.

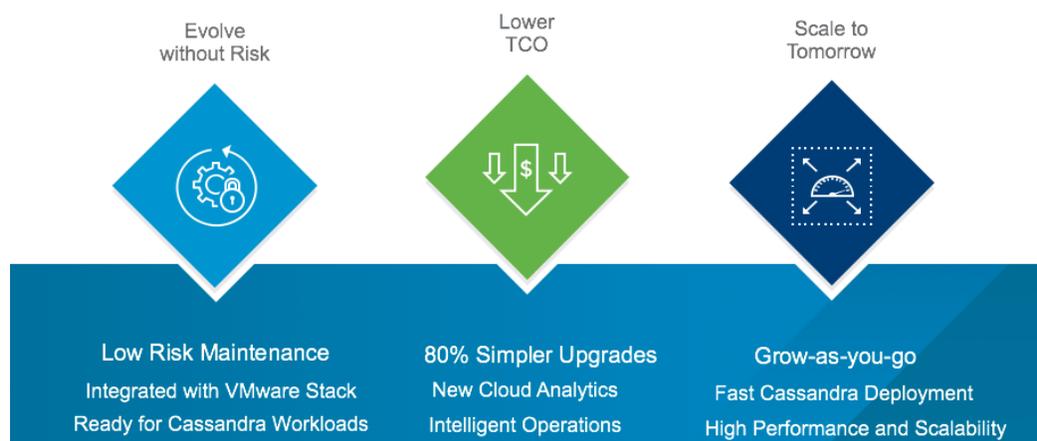


VMware and DataStax have jointly undertaken an extensive technical validation to demonstrate vSAN as a storage platform for globally distributed cloud applications for test and development environments. The companies are working together on shared nothing vSAN enhancements with a design focused on cloud applications that require data to be contextual, always on, real time, distributed, and scalable.

## Why DataStax Enterprise (DSE) on vSAN?

Many customers have asked for reference architecture to modernize their existing DSE applications on vSAN, running on VMware certified x86 servers, to eliminate traditional IT silos of compute, storage, and networking. All storage management moves into a single software stack, thus taking advantage of the security, performance, scalability, operational simplicity, and cost-effectiveness of vSAN in development environments.

Workloads can be easily migrated from aged, bare-metal configurations to a modern, dynamic, and consolidated hyper-converged infrastructure based on vSAN.

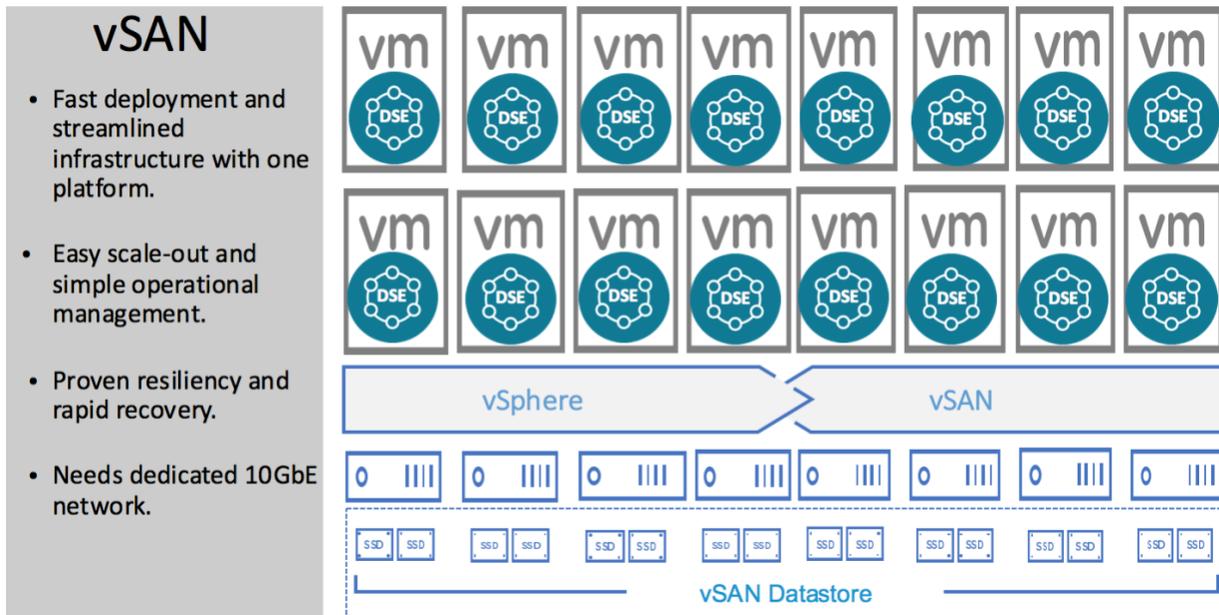


## High Level Architecture

To ensure continued data protection and availability of DSE during planned or unplanned down time, a minimum of four nodes are recommended for the vSAN cluster. In VMware's solution, an 8-node vSAN cluster is used with 16 DSE nodes to validate the cluster functions as expected for typical workloads and scale of test and development environments.



A typical configuration includes two or more disk groups per node. Custom storage policies can be created for different DSE applications to satisfy performance, resource commitment, failure tolerance, checksum protection, and quality of service requirements in an application-centric way. An all-flash configuration is required for performance consistency and predictable latency.



## Test Configuration

The vSAN and DSE testing environment consists of 8 DELL R630 servers for the 16-node DSE cluster. We use an additional 4-node hybrid vSAN cluster for the DSE OpsCenter and 8 testing client nodes that are used to generate load on the cluster.

PROPERTY	SPECIFICATION
Server	DELL R630
CPU cores	2 sockets, 12 cores each of 2.3GHz with hyper-threading enabled
RAM	256GB
Network adapter	2 x 10Gb NIC
Storage adapter	SAS Controller Dell LSI PERC H730 Mini NVMe Controller Intel Corporation DC P3700 SSD
Disks	Cache-layer SSD: 2 x 1.8TB Intel SSD DC P3700 NVMe Capacity-layer SSD: 8 x 800GB 2.5-inch Enterprise Performance SAS SSD S3710
PROPERTY	SPECIFICATION
OS	Ubuntu 14.04
vCPU	12
RAM	64GB (heap size of 24GB)
Cassandra distribution	DataStax Enterprise 5.1
Disks	OS disk: 40GB Data disk: 1,150GB Log disk: 50GB

## Workload

The eight client nodes all run [Cassandra-stress](#), a built-in DSE benchmark tool used for workload testing. To prepare the cluster for testing, Cassandra-stress is used to load the base data set until we get the average data of at least 500GB per DSE node.

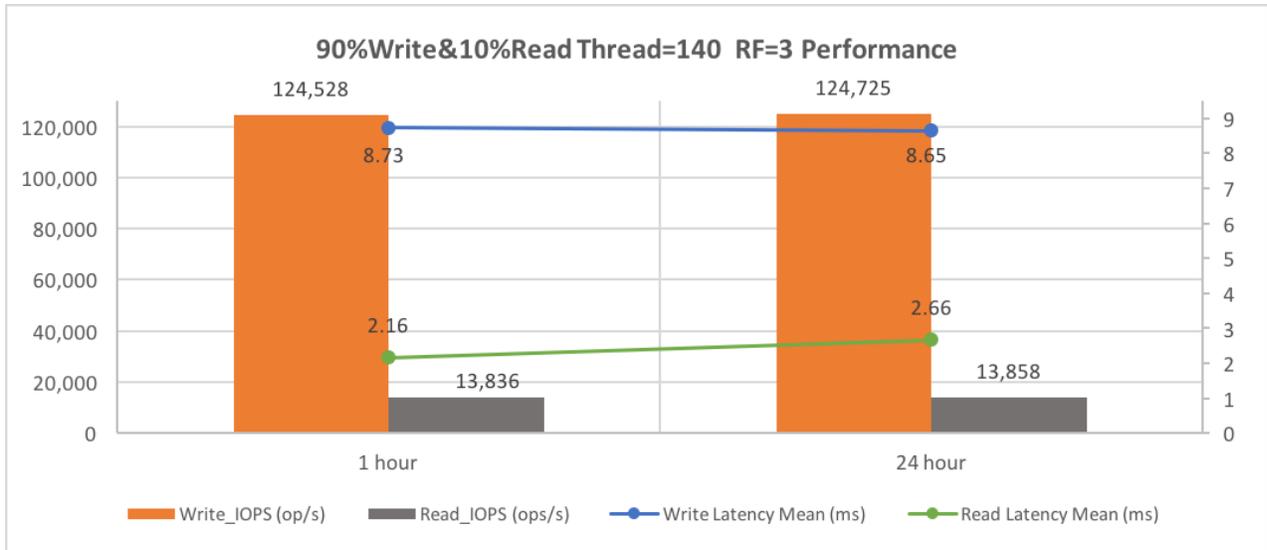
## Performance Result Examples

In the performance testing, Cassandra-stress randomly inserts and reads, running a 90% write and 10% read workload to get a measure of how we are stressing the cluster and to work towards maximizing the peak performance when no maintenance tasks are occurring and then backing off

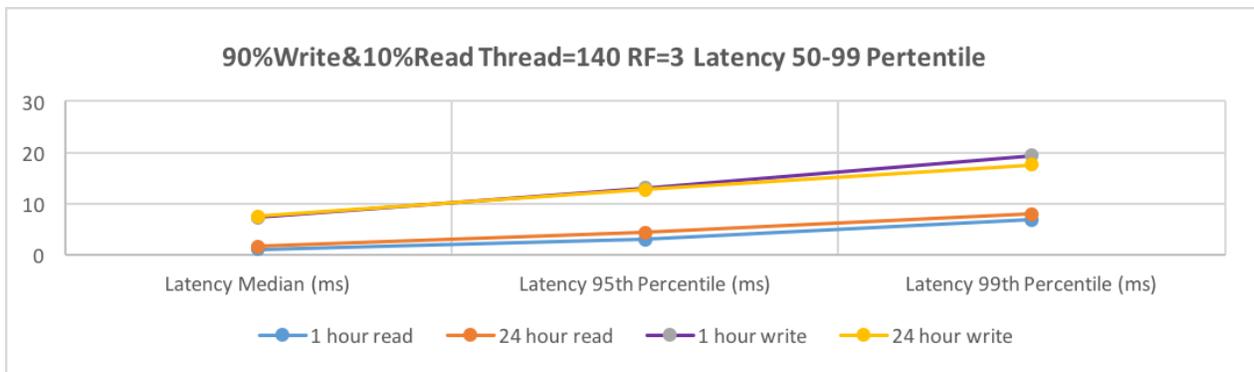


from that to a workload we would advise customers to run in a typical scenario, which reduces threadcount on each client by 30%.

The throughputs and latencies of 1-hour and 24-hour performance tests (90% write and 10% read) are consistent, which validates vSAN provides consistent performance at predictable latency.



In 90% write and 10% read performance tests, median read latency is less than 2ms while median write latency is less than 8ms. 95 percentile and 99 percentile latency numbers are reasonable.



## Resiliency and Availability

vSAN's storage-layer resiliency features combined with DSE's peer-to-peer design enable this solution to meet the data availability requirements of even the most demanding applications. A set of failure scenarios are created to validate data availability. In the failure testing, we again run Cassandra-stress against the preloaded dataset. We validate the disk failure, VM failure, and host failure while running the performance testing workload.

### Best Practices:

- Initial testing was conducted using EXT4 and required journaling to be enabled.
- Users must use a rack-aware switch to ensure that multiple data replicas are not stored on the same ESXi host.

FAILURE TYPE	TEST DESCRIPTION	RESULT
Disk failure	Fail one disk, and bring it back after 20 minutes.	Performance impact is negligible, after bringing the disk back, vSAN resyncs the data in less than 2 minutes.
VM failure	Fail a VM and bring it back after an hour.	Service is not interrupted, performance is degraded as expected due to the loss of a DSE node.
Host failure	Power off a host and bring it back after 20 minutes.	Performance is degraded due to the loss of two DSE nodes as expected, but service is not interrupted since the host is back within one hour. vSAN only resyncs the data without component rebuild.

## Support

Customers can approach VMware or DataStax for adequate advice and support for using vSAN in development at the current stage.

## Summary

VMware vSAN, vSphere, and vCenter Server collectively power the best HCI solutions for deploying, running, and managing DSE applications that require predictable performance and high availability. The integration of vSAN with vSphere simplifies the operational management with a single software stack and automation through Storage Policy-Based Management (SPBM).



The advantages of running DSE applications on vSAN include the acceleration of legacy application migration, lower costs, simple management, high resiliency, and improved availability.

If more capacity or larger workload is needed, it is simple to expand using a scale-up or scale-out approach without incurring any downtime. With the joint efforts of VMware and DataStax, customers can deploy DSE clusters on vSAN for their modern cloud applications with ease and confidence in test and development environments currently.



## Learn More

See more vSAN details and customer stories:

- [vSAN](#)
- [Virtual Blocks Blog](#)
- [Customer Stories](#)

For more information regarding DataStax Enterprise, see [DataStax](#).



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 [www.vmware.com](http://www.vmware.com)  
Copyright © 2017 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

DataStax is a registered trademark of DataStax, Inc. and its subsidiaries in the United States and/or other countries. Apache Cassandra, and Cassandra are trademarks of the Apache Software Foundation or its subsidiaries in Canada, the United States and/or other countries.