VMware Virtual SAN™ 6.2 Performance with Online Transaction Processing Workloads

Performance Study

TECHNICAL WHITE PAPER

vmware®

Table of Contents

Executive Summary

This white paper examines the performance of Online Transaction Processing (OLTP) applications with Virtual SAN 6.2. The OLTP applications, which model an online retail store and a brokerage house workload, involve a large number of client transactions. The performance of such workloads are crucial to businesses such as banks, airlines, and retailers. Overall, Virtual SAN 6.2 performs well and provides stable and consistent I/O capability to the two OLTP workloads. Furthermore, tests show that Virtual SAN 6.2's space efficiency features provide substantial disk space savings and can significantly reduce customers' storage costs per gigabyte (GB). Lastly, Virtual SAN 6.2 performs well during source failure with only a small drop during recovery.

Introduction

Virtual SAN is a distributed layer of software that runs natively as part of the VMware vSphere® hypervisor. Virtual SAN aggregates local or direct-attached storage disks in a host cluster and creates a single storage pool that is shared across all hosts of the cluster. This eliminates the need for external shared storage and simplifies storage configuration and virtual machine provisioning operations. In addition, Virtual SAN supports vSphere features that require shared storage such as VMware vSphere® High Availability (HA), VMware vSphere® vMotion®, and VMware vSphere® Distributed Resource Scheduler™ (DRS) for failover. More information on Virtual SAN design can be obtained in the Virtual SAN design and sizing guide [1].

Note: Hosts in a Virtual SAN cluster are also called nodes. The terms "host" and "node" are used interchangeably in this paper.

Virtual SAN 6.2 New Features

Virtual SAN 6.2 introduces new features to improve space efficiency and data integrity. These features provide benefits to users but may consume more resources. For a full review of Virtual SAN 6.2 new features, please refer to the datasheet [2], white paper [3], and blog post [4]. The data in this white paper demonstrates performance numbers with the following new features and illustrates the trade-off between performance and resource cost.

- **Data integrity feature: software checksum**
  Software checksum is introduced to enhance data integrity. Checksum works on a 4KB block. Upon a write, a 5-byte checksum is calculated for every 4KB block and stored separately. Upon a read operation, a 4KB data element is checked against its checksum. If the checksum does not match the calculation from data, it indicates there is an error in the data. In this case, the date is fetched from a remote copy instead, and the data with the error is updated with a remote copy.
- **Space efficiency feature: erasure coding (RAID-5/RAID-6)**
  Previous Virtual SAN releases support only RAID-1 configuration for data availability. To tolerate 1 failure, 1 extra data copy is required and there is a 100% capacity overhead. Similarly, 200% capacity overhead is needed to tolerate 2 failures. However, in Virtual SAN 6.2, a RAID-5 configuration tolerates 1 failure by storing 1 parity from 3 different data objects. Therefore, only a 33% capacity overhead is needed. Furthermore, a RAID-6 configuration tolerates 2 failures by storing 2 parities for every 4 different data objects in a 6-node cluster. Hence, only 50% capacity overhead is needed to tolerate 2 failures.
- **Space efficiency feature: deduplication and compression**
  The data stored in Virtual SAN may be de-duplicable or compressible in nature. Virtual SAN 6.2 introduces deduplication and compression features which reduce the space required while the data is being persisted to disks. Deduplication and compression are always enabled or disabled together. The scope of the features is per disk group. Deduplication works when the data is de-staged from the caching tier to the capacity tier, and its granularity is 4KB. Upon writing a 4KB block, it is hashed to find whether an identical block already exists in the capacity tier of the disk group. If there is one, only a small meta-data is updated. If no such identical block is available, compression is then applied to the 4KB block. If the compressed size of the 4KB

block is less than 2KB, Virtual SAN writes the compressed data to the capacity tier. Otherwise, the 4KB block is persisted to the capacity tier uncompressed.

Erasure coding, and deduplication and compression features are available only on an all-flash configuration. The software checksum feature is available on both hybrid and all-flash configurations. Erasure coding and software checksum features are policy driven and can be applied to an individual object on Virtual SAN. Deduplication and compression can be enabled or disabled across clusters.

Virtual SAN Cluster Setup

Because the all-flash Virtual SAN configuration supports all the new features discussed in this paper, two all-flash clusters are chosen as the hardware platform. A 4-node cluster is used for the RAID-5 configuration, and an 8-node cluster is used for the RAID-6 configuration. Individual nodes in these 2 clusters have identical hardware configurations.

Virtual SAN Hardware Configuration

Briefly, each node is a dual-socket Intel Xeon CPU E5-2670 v3 @ 2.30 GHz system with 48 Hyper-Threaded (HT) cores, 256GB memory, 2x 400GB Intel P3700 PCIe SSDs, and 1 LSI MegaRAID SAS controller hosting 6x 800GB Intel S3500 SATA SSDs. Please note that in the actual experiment, the PCIe SSDs and SATA SSDs were shuffled between hosts to form disk group configurations as specified in the Virtual SAN configuration section. Each node is configured to use a 10GbE port dedicated to Virtual SAN traffic. The 10GbE ports of all the nodes are connected to a 10GbE switch. Jumbo frames (MTU=9000 bytes) is enabled on the Virtual SAN network interfaces. A 1GbE port is used for all management, access, and inter-host traffic. Details on the hardware are available in Appendix A.

Workloads and Virtual Machine Configurations

Online Transaction Processing (OLTP) is extensively used to support businesses such as online retail stores and financial institutes. Two OLTP workloads are used in this white paper: the DVD Store workload and the Brokerage workload. These workloads strain the database software's ability to handle transactions and its underlying infrastructure (most importantly, the storage system) in order to test each workload's performance with Virtual SAN.

The open-source DVD Store version 2.1 [5] is used as the first workload. DVD Store simulates an online ecommerce DVD store, where customers log in, browse, and order products. The benchmark tool is designed to utilize a number of advanced database features, including transactions, stored procedures, triggers, and referential integrity. The primary performance metric of DVD Store is orders per minute (OPM).

The entire DVD Store benchmark tools, including the query generator and the database backend, are placed in a single virtual machine, which runs the Microsoft Windows Server 2008 R2 operating system and Microsoft SQL Server 2008. The virtual machine is configured with 4 virtual CPUs (vCPUs) and 4GB of memory. The virtual machine is configured with 3 virtual disks: a 50GB boot disk containing Windows Server 2008 R2 and Microsoft SQL Server 2008, a 200GB database disk, and a 10GB log disk. The DVD Store workload used a database size of 100GB with 200 million customers and 10 million products. A more detailed workload configuration can be found in Appendix C. Unless otherwise mentioned in the text, 4 DVD Store virtual machines per node are used in the experiments.

The second workload is the Brokerage workload, which models a brokerage house responding to customer requests and stock market data based on the TPC-E[1] benchmark. It uses TPC-E fairly but is implemented in a way

[1] Disclaimer: The Brokerage workload in this paper is a fair-use implementation of the TPC-E business model; the Brokerage workload results are not TPC-E compliant and are not comparable to official TPC-E results. TPC Benchmark, TPC-E, and tpsE are trademarks of the Transaction Processing Performance Council [6].

that focuses on the storage performance on a virtualized platform. The term "Brokerage" workload is used to refer to this workload in the rest of this paper. Similar to the DVD Store workload, one instance of the Brokerage workload is encapsulated into a single virtual machine and the workload scales with numbers of virtual machines in the cluster.

A Brokerage virtual machine uses a Microsoft Windows Server 2012 R2 as the operating system and Microsoft SQL Server 2014 edition as the database. The virtual machine was configured with 4 virtual CPUs (vCPUs) and 64GB of memory. In terms of storage, it has 1 operating system VMDK and 12 VMDKs to accommodate the database for brokerage firm, market data, customer, and associated database log files. The total size of the disks are 1.9TB and all of them are deployed to Virtual SAN. The details of the Brokerage virtual machine are available in Appendix B. Unless otherwise mentioned in the text, 3 virtual machines per node are used in the experiments.

Metrics

In all the experiments, two important metrics are measured: I/Os per second (IOPs) and the average latency of each I/O operation. The IOPs metric is important because for the workloads discussed in this paper, the guest I/Os are all invoked by the transaction requests from the application. Therefore, the transaction rate is proportional to the guest IOPs. In the Brokerage workload, instead of using "tpsE" as the benchmark metric, guest IOPs and latency are measured to reflect the transaction performance.

The primary performance metric of the DVD Store benchmark is orders per minute (OPM). The DVD Store benchmark driver outputs a moving average of orders per minute and a cumulative number of transactions every 10 seconds. An absolute value of orders per minute is computed from the cumulative number of transactions.

CPU utilization and storage space usage are recorded as the resource consumption metric. The overall system utilization implies how busy the system is under the workload. The Virtual SAN CPU utilization reflects the resources consumed by software to support the workload. The Virtual SAN CPU utilization is important because it implies the overhead of Virtual SAN's software-based storage solution. Intuitively, the Brokerage workload consumes more Virtual SAN CPU than the DVD Store workload because Brokerage has a higher IOPs requirement. This is illustrated with experiment results in a later section.

Storage space usage is measured by adding up the space consumed by the capacity tier disks in the whole Virtual SAN cluster. The measure is taken when all the data from workload is de-staged to the capacity tier disks and no data is buffered in the caching tier for accuracy. The space usage number in the cluster is in gibibytes (GiB), and a space saving ratio in percentage (%) is also presented when comparing with the baseline of Virtual SAN 6.2 default configuration. This percentage directly reflects the benefit of the space efficiency features.

Virtual SAN Configurations

In the experiments, a baseline is first established on the Virtual SAN 6.1 release. Then several Virtual SAN 6.2 feature combinations are used. For the rest of paper, the abbreviations in Table 1 are used to represent the configurations of features.

Name	Failure to Tolerate	Checksum	RAID level	Deduplication and compression
6.1	1	No	1	No
6.2	1	Yes	1	No
6.2 R5	1	Yes	5	No
6.2 D	1	Yes	1	Yes
6.2 R5+D	1	Yes	5	Yes
6.1 FTT2	2	No	1	No
6.2 FTT2	2	Yes	1	No
6.2 R6	2	Yes	6	No
6.2 R6+D	2	Yes	6	Yes

Table 1. Test name abbreviations and configurations

Unless otherwise specified in the experiment, the Virtual SAN cluster is designed with the following common configuration parameters:

- Stripe width of 1
- Default cache policies are used and no cache reservation is set
- 3 disk groups for the Brokerage workload, each disk group has 6 capacity SSDs
- 2 disk groups for the DVD Store workload, each disk group has 3 capacity SSDs

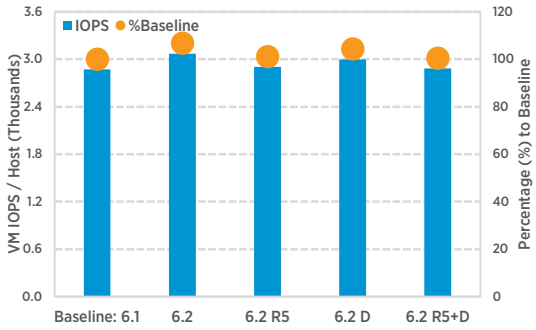
Performance of Virtual SAN Cluster

DVD Store Workload

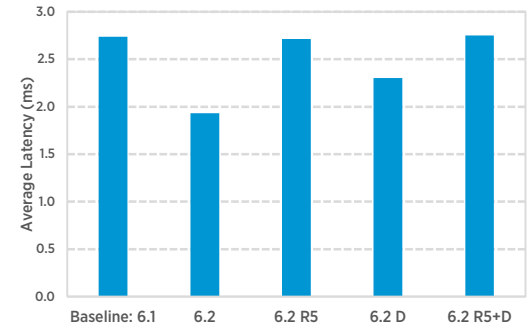
The DVD Store workload has a read/write ratio of 1:2 and 1 virtual machine generates 700-800 IOPs. This represents a low/moderate IOPs requirement from the underlying storage. The results show Virtual SAN 6.2 provides good, steady storage performance to the workload.

In the first DVD Store experiment, a Failures to Tolerate setting of 1 is used on the 4-node cluster. One run of test lasts for a duration of 4 hours. The aforementioned Virtual SAN 6.2 feature combinations are used and the performance metrics are collected and compared with a Virtual SAN 6.1 baseline. Please note that all 6.2 test cases have software checksum enabled, while the feature is not available in the 6.1 baseline. Figure 1 (a) and (b) show the IOPs and latency, respectively. In 6.1, the guest IOPs per host is around 2.9K and average latency is 2.7ms. For all the Virtual SAN 6.2 cases, the IOPs numbers are higher than or equal to the baseline, and the average latency numbers are lower than or equal to the baseline despite every I/O goes through the verification or extra write due to software checksum. As a result, the application's operations per minute sustained steadily at

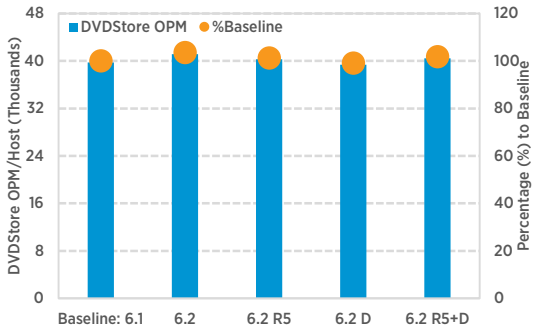
around 40K per host for all the Virtual SAN 6.2 cases as shown in Figure 1 (c). The results show Virtual SAN 6.2 delivers constant, well performing storage for the DVD Store workload.



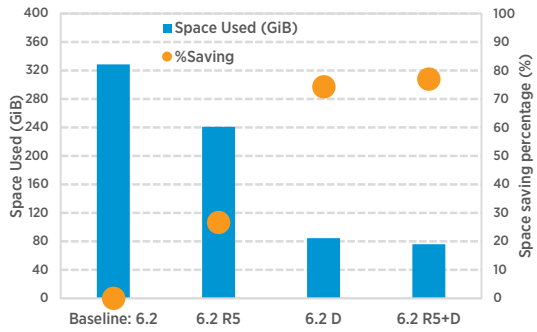
(a) Guest IOPs per host



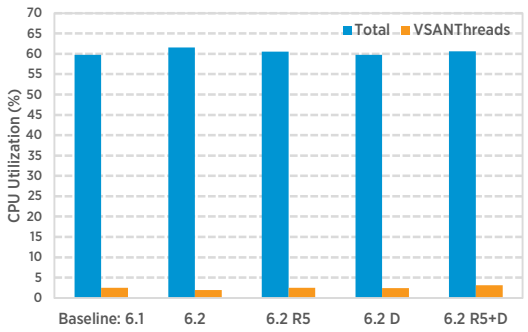
(b) Guest average latency



(c) DVD Store OPM per host



(d) Space usage in the cluster and savings in %



(e) System and Virutal SAN CPU utilization in the cluster

Figure 1. DVD Store RAID-5 results of (a) IOPs per host (b) average latency (c) OPMs per host, (d) space usage and savings percentage, and (e) CPU utilization

Meanwhile, the space efficiency features reduce the space taken on disk. The test is modified to run with a single DVD Store instance in the whole Virtual SAN cluster to eliminate the artificially duplicated data between instances. Virtual SAN 6.2, rather than 6.1, is used as baseline because its space usage is very similar to 6.1 except for a very small meta-data overhead[2]. Figure 1 (d) shows the space usage in the cluster and the saving

[2] Virtual SAN 6.2 default configuration has the software checksum feature, in which there is a 5 Bytes checksum for a 4KB block. This is the extra meta-data overhead compared with Virtual SAN 6.1

percentage. The 6.2 baseline is a RAID-1 mirroring configuration and takes about 330GiB space in the cluster. 6.2 R5 saves by 27% and only takes 241GiB on disks. 6.2 D gives a 74% space saving (85GiB on disk) and 6.2 R5+D gives a 77% saving (76GiB on disk). The substantial space saving ratio is observed because the DVD Store workload VM contains duplicable and compressible data that can be reduced by the deduplication and compression feature in Virtual SAN 6.2 (The actual space saving ratio in production environment will depend on the workload.).

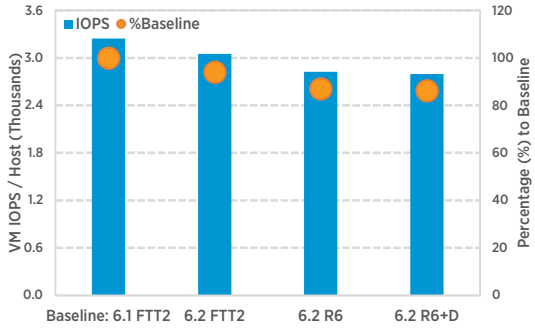
The benefit of Virtual SAN 6.2 features comes at a minimal CPU cost for DVD Store workload. Figure 1 (e) shows the system and Virtual SAN CPU utilization in the cluster corresponding to the tests in Figures 1 (a) to (c). The overall system CPU utilization for the 6.1 baseline is 59.7%, while for all the Virtual SAN 6.2 cases, the system CPU utilization is very similar and in the range of 59.8% to 61.6%. The Virtual SAN CPU utilization in the cluster for the 6.1 baseline is 2.5%, while for all the Virtual SAN 6.2 cases, it is in the range of 2.0% to 3.1%.

The second experiment for the DVD Store workload studies the performance when Failures to Tolerate is set to 2 on the 8-node cluster with a RAID-6 configuration. Test results show that Virtual SAN has only a small performance impact while delivering the benefits of all features. In the case where all features are enabled, DVD Store OPM drops by only 13% compared with 6.1 but the space consumed in the Virtual SAN cluster is reduced by 74% compared with the 6.2 baseline.

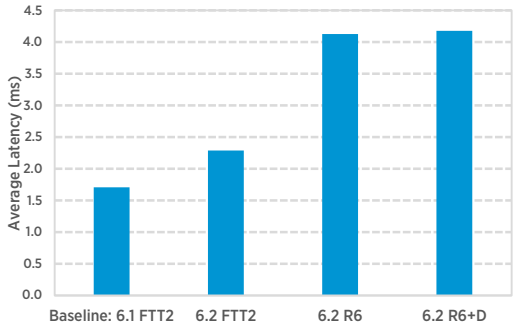
Figure 2 plots the results against Virtual SAN 6.1 baseline. Figures 2 (a) and (b) plot the guest IOPs per host and average latency. With Failures to Tolerate set to 2, Virtual SAN provides 3 replicated copies of data in the default configuration which uses RAID-1. As a result, compared to the case of tolerating 1 failure, software checksum has to be performed on 1 more extra replica upon a write operation. The baseline 6.1 FTT2 provides 3.25K IOPs per host. 6.2 FTT 2 has software checksum enabled, and the IOPs drop by 5.9% to 3.05K IOPs per host. In terms of latency, 6.1 has 1.7ms average latency without checksum protection, and 6.2 FTT2 provides protection with a latency of 2.2ms. This is tolerable considering that every read is verified against the checksum and every write introduces an extra checksum write. 6.2 R6 and 6.2 R6+D performed very similarly: for both cases, the IOPs dropped by 13% to 2.80K IOPs per host and the latency increased to 4.2ms. Again, this is tolerable considering the benefit of checksum and space saving features. Figure 2 (c) plots the orders per minute, which is proportional to IOPs per host.

The space efficiency features also help reduce the space consumed on disks in the RAID 6 experiments. The test is modified to run with 1 DVD Store instance in the whole Virtual SAN cluster to eliminate the duplicable data between instances. Virtual SAN 6.2 is used as the baseline. Figure 2 (d) shows the space used and saving ratio in percentage for the 6.2 baseline, 6.2 R6, and 6.2 R6+D cases[3]. The 6.2 baseline takes 521GiB in the cluster. 6.2 R6 gives 37% saving (327GiB on disks) while providing the same level of failure tolerance by RAID-6. Further, 6.2 R6+D gives an impressive 74% saving (134GiB on disks) by enabling deduplication and compression on top of RAID-6.

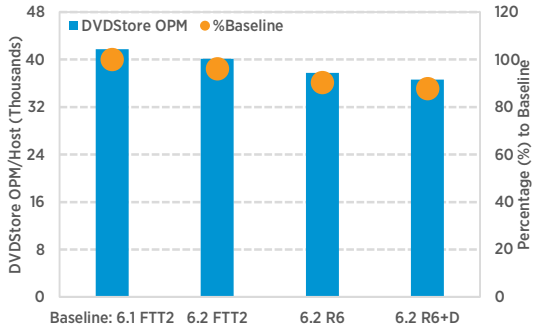
[3] The space saving data for 6.2 D case is not shown because the impact of deduplication feature is already studied in the first experiment of DVD Store workload



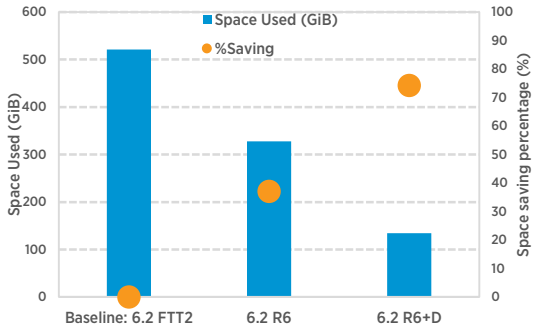
(a) Guest IOPs per host



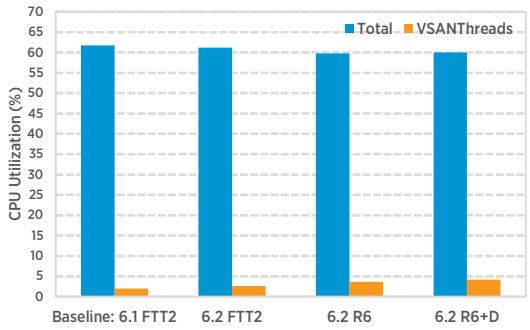
(b) Guest average latency



(c) DVD Store OPM per host



(d) Space usage in the cluster and savings in %



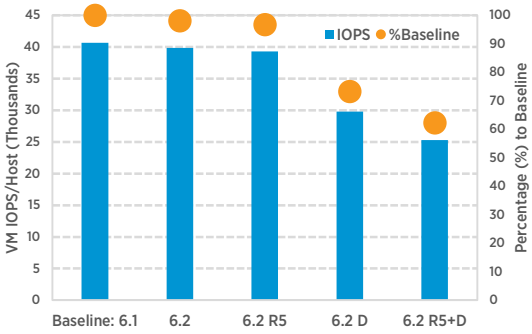
(e) System and Virtual SAN CPU utilization in the cluster

Figure 2. DVD Store RAID-6 results of (a) IOPs per host (b) average latency (c) OPMs per host, (d) space usage and savings percentage, and (e) CPU utilization

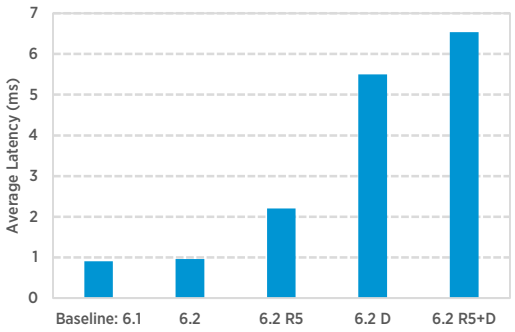
The system and Virtual SAN CPU utilization in the cluster are plotted in Figure 2 (e). The overall system CPU utilization is very similar for all cases. The Virtual SAN CPU utilization for the 6.1 FTT2 baseline was 2%, and it increased to 2.6% for 6.2 FTT2. This reflects the calculation cost of software checksum on every I/O. For 6.2 R6 and 6.2 R6+D, this metric further increased to 3.6% and 4.1% respectively. This is because RAID-6 involves more IOPs on the parity and needs to do the additional calculation. Overall, Virtual SAN 6.2 performs well in a RAID-6 configuration with the DVD Store workload, although with a small IOPs degradation and slightly higher Virtual SAN CPU consumption.

Brokerage Workload

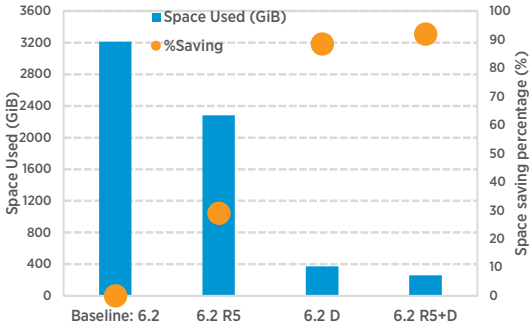
The Brokerage workload uses reads and writes at the ratio of 9:1 and the average I/O size is 8KB. One instance of the Brokerage workload generates approximately 13K IOPs, which represents an OLTP workload that requires relatively high IOPs from the underlying storage. The duration for each test run is 8 hours. The experiment results show Virtual SAN 6.2 performs well and sustains the IOPs required with the default and RAID-5 configurations given the software checksum and RAID-5 space saving benefit. With deduplication and compression, the IOPs drops by 25% percentage, but a space saving of nearly 90% is achieved because the workload data are de-duplicated and compressed.



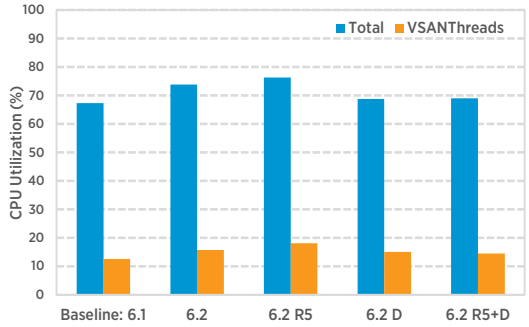
(a) Guest IOPs per host



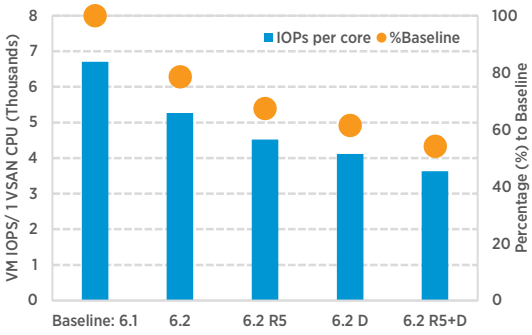
(b) Average guest latency



(c) Space usage in the cluster and savings in %



(d) System and Virtual SAN CPU utilization in the cluster



(e) Virtual SAN CPU efficiency: IOPs per CPU

Figure 3. Brokerage workload results of (a) IOPs per host, (b) average latency, (c) space usage, (d) CPU utilization, and (e) IOPs per Virtual SAN CPU

Figure 3 shows the resulting plots for the Brokerage workload experiment. Figures 3 (a) and (b) plot the guest IOPs and average latency respectively against the baseline of Virtual SAN 6.1. Clearly 6.2 and 6.2 R5 performed

very well: IOPs dropped by only 2% and 4%. In the 6.2 case, with software checksum, the average latency is almost the same as the 6.1 baseline case where no software checksum is available. In the 6.2 R5 case, where RAID-5 is used on top of software checksum, the average latency increases to 2.2ms, which is still low. 6.2 D, which has deduplication and compression enabled, brings IOPs down by 25% to 30K and pushes average latency to 5.5ms. In the 6.2 R5+D case, where RAID-5 is used together with deduplication and compression, IOPs is down by a further 12% to 25K and guest latency increases to 6.5ms.

The deduplication and compression feature has lower IOPs and latency performance in the Brokerage workload, but it provides a good space saving benefit. Figure 3 (c) plots the space consumed in the cluster and saving ratio against Virtual SAN 6.2 baseline. Please note in this space saving experiment, the test configuration is modified to run only 1 instance of the Brokerage workload in the cluster to eliminate the artificially duplicable data across multiple instances. The total disk space taken on disks in the cluster is measured. The 6.2 baseline, which has only software checksum but no storage efficiency features, takes 3211GiB in the cluster. 6.2 R5 gives a 30% space saving and takes 2279GiB on disks by using RAID-5 to provide the same failure to tolerant. 6.2 D saves 88.4% and takes only 372GiB on disks. In the extreme case where RAID-5 is used together with deduplication and compression, 6.2 R5+D saves 92% of space and only uses 260GiB space in the cluster. The substantial space saving ratio is observed because a single Brokerage workload VM contains duplicable and compressible data that can be reduced significantly by the deduplication and compression feature in Virtual SAN 6.2 (The actual space saving ratio in production environment will depend on the workload.).

The CPU resource impact of the features is explored in Figure 3 (d). Please note the CPU number being discussed is corresponding to the test runs in Figures (a) and (b). The overall system and the Virtual SAN CPU utilization in the cluster is plotted. 6.2 with software checksum imposes 7% more system usage and 3% more Virtual SAN usage.(6.2 R5) imposes 9% more system usage and 5% more Virtual SAN usage compared to 6.1. 6.2 D and 6.2 R5+D both use 2% more CPU for the system and Virtual SAN usage compared to the 6.1 baseline.

The Virtual SAN CPU efficiency number is calculated by dividing the total guest IOPs in the cluster by the CPU used by Virtual SAN. This normalized “IOPs per CPU core” is plotted in Figure 3 (e). The impact of each feature is clearly shown. For the 6.1 baseline, 1 CPU used by Virtual SAN can drive 6.7K IOPs. For 6.2, which has software checksum, drives 5.3K IOPs/CPU and this is 21% drop from baseline. The IOPs/CPU is 4.5K or 33% drop from baseline for 6.2 R5 because of the extra operations and parity calculation. Deduplication and compression feature costs slightly more: 6.2 D gives 4.1K IOPs per CPU or 39% drops, however nearly 90% space saving is achieved. Lastly, in the extreme case where all features are enabled, the IOPs per Virtual SAN CPU usage is 3.6K or 46% drop from 6.1 baseline.

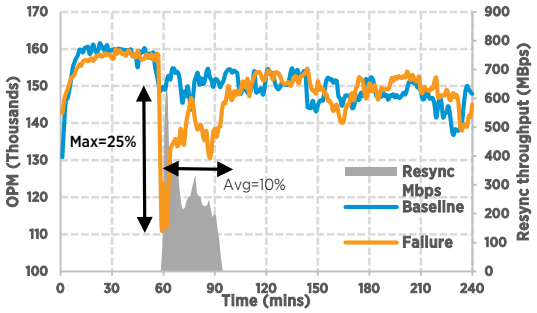
Performance During Failure of Resource

In this experiment, the failure performance for DVD Store workload is discussed. A Virtual SAN RAID-5 configuration in a 5-node cluster and a RAID-6 configuration in 8-nodes cluster are used respectively. In both cases, software checksum, and the deduplication and compression features are enabled as well. In both cases, a baseline run of 4 hours is first established without any failure, and then the same test is repeated but with a failure introduced. The recovery traffic performance is observed and the impact to running workload is measured. Overall Virtual SAN 6.2 performs very well: the recovery finishes with reasonably high throughput and the DVD Store orders per minute drops only by an average of 10% during recovery. After recovery, DVD Store performance recovers to the same level as the baseline.

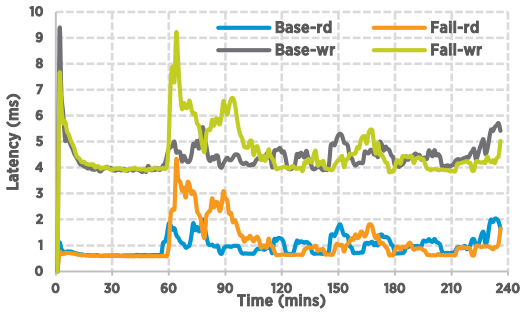
At 60 minutes into the test, a failure of disk group decommission is introduced to the Virtual SAN cluster by removing the SSD from the disk group on one host. This disk group contains multiple objects that are now marked as degraded. This triggers Virtual SAN to re-create replicas for those objects. At the same time, the read and write from workload needs to continue, and will be served from replicas that are not in the host on which the failure happens. In order to let the DVD Store VMs all have a uniformed performance during the recovery, the

workload only running on the hosts that do not have the failure, that is, 4 good hosts in the RAID-5 case and 7 good hosts in the RAID-8 case. The baseline is established in a similar fashion.

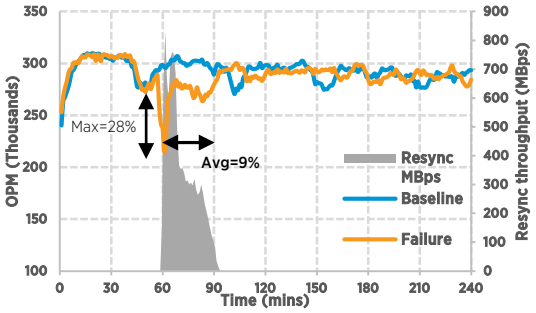
Figure 4 shows the results of both 5-nodes and 8-nodes cases. Figure 4 (a) shows the DVD Store orders per minute over time. The disk group on 1 node is decommissioned at 60 minutes and the Virtual SAN 6.2 starts recovery traffic immediately. The recovery lasts for 35 minutes with an average throughput of 270 megabytes per second. The DVD Store workload continues while the recovery traffic is going. The orders per minute drops by 10% in average compared with the baseline without failure. A max 25% drop of orders per minute is observed at the beginning of the recovery, at which point the recovery traffic also sees a peak point. The IOPs is proportional to orders per minute in the DVD Store workload, hence the IOPs should have similar impact. Figure 4 (b) shows the latency change over time. Taking the read latency as an example, it is flat at 0.8ms during the first 60 minutes of tests. During the recovery phase, the read latency increases, and fluctuates between 2ms and 4ms. After the recovery, the read latency is back to the level of 0.8ms to 2ms. Write latency shows an identical trend except that the minimal write latency is around 4ms, and the peak write latency is up to 9ms.



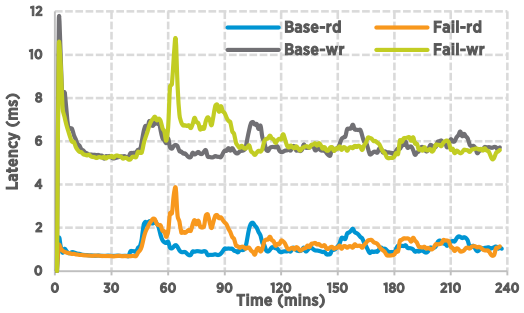
(a) RAID-5: OPM in the cluster and recovery traffic



(b) RAID-5: average latency



(c) RAID-6: OPM in the cluster and recovery traffic



(d) RAID-6: Average latency

Figure 4. DVD Store result during failure (a) RAID-5 OPM and recovery traffic, (b) RAID-5 average latency, (c) RAID-6 OPM and recovery traffic, and (d) RAID-6 average latency

Figures 4 (c) and (d) show the failure result on the 8-node cluster with a RAID-6 configuration. The impact is very similar to the 5-node cluster: the orders per minute sees an average of a 9% drop while Virtual SAN recovers the missing components from the failed disk group. The OPM sees a max of 28% OPM drop at the beginning of recovery. The average latency increases during recovery but drops back to the same level as baseline after recovery finishes.

Clearly, Virtual SAN 6.2 performs well during source failure: for RAID-5 and RAID-6 configuration, DVD Store OPM only sees an average of 10% performance drop and it recovers to a similar level without failure.

Conclusion

This paper shows the performance results of Virtual SAN 6.2 when used as the storage for two OLTP workloads that model an online retail store (DVD Store 2.1) and a brokerage house, respectively. In the DVD Store workload, Virtual SAN performs exceptionally by providing constant IOPs and latency, and delivers good space savings with the new features at a minimal CPU overhead. In the Brokerage workload, where high IOPs is needed, Virtual SAN performs very well with software checksum and RAID 5 features as the high IOPs is sustained; while with the deduplication and compression feature and RAID 5 configuration, a 37% IOPs drop is observed but a substantial space saving of up to 90% is achieved. The performance experiment results of the two OLTP workloads demonstrate Virtual SAN 6.2's performance capability of handling such workloads and the benefits of space savings that can reduce customers' storage costs per GB.

Appendix A. Hardware Configuration for All-Flash Virtual SAN Cluster

The servers were configured as follows:

- Dual-socket Intel® Xeon® CPU E5-2670 v3 @ 2.30GHz system with 48 Hyper-Threaded (HT) cores
- 256GB DDR3 RAM @1866MHz
- 1x LSI / Symbios Logic MegaRAID SAS Fusion Controller with driver version: 6.603.55.00.1vmw, build: 4852043
- 2x 400GB Intel P3700 PCIe SSDs
- 6x 800 GB Intel S3500 SSDs
- 1x dual-port Intel 10GbE NIC (82599EB, Fibre Optic connector)
- 1x quad-port Broadcom 1GbE NIC (BCM5720)

Appendix B. Brokerage Virtual Machine and Workload Configuration Detail

The Brokerage workload virtual machine was configured as follows:

- 64-bit Microsoft Windows Server 2012 R2
- 4 vCPUs, 64GB memory
- VMXNET3 driver version 1.6.6.0, PVSCSI driver version 1.3.4.0
- 160GB disk for the operating system with the LSI Logic controller
- 8 x 155GB broker database disks on the PVSCSI controller
- 70GB database log disk on the PVSCSI controller
- 100GB market data disk and 280GB customer data disk on the PVSCSI controller
- 80GB disk for miscellaneous data on the PVSCSI controller
- Microsoft SQL Server 2014

Appendix C. DVD Store Virtual Machine and Workload Configuration Detail

The DVD Store workload virtual machine was configured as follows:

- 64-bit Microsoft Windows Server 2008 R2
- VMXNET3 driver version 1.2.20.0, PVSCSI driver version 1.1.1.0
- 50GB disk for the operating system with the LSI Logic controller
- 200GB database disk and 10GB log disk on the PVSCSI controller
- Microsoft SQL Server 2008

The DVD Store workload was configured with:

```
n_threads=8
ramp_rate=100
run_time=180
db_size=100GB
think_time=0.002
pct_newcustomers=20
n_searches=15
search_batch_size=15
n_line_items=15
virt_dir=ds2
page_type=php
windows_perf_host=
linux_perf_host=
detailed_view=n
```

References

[1] John Nicholson. VMware Virtual SAN 6.2 Design and Sizing Guide.
https://www.vmware.com/files/pdf/products/vsan/virtual-san-6.2-design-and-sizing-guide.pdf

[2] VMware, Inc. VMware Virtual SAN 6.2.
https://www.vmware.com/files/pdf/products/vsan/VMware_Virtual_SAN_Datasheet.pdf

[3] VMware, Inc. What's New with VMware Virtual SAN 6.2.
https://www.vmware.com/files/pdf/products/vsan/vmware-virtual-san-6-2-technical-white-paper.pdf

[4] VMware, Inc. Whats New Vmware Virtual SAN 6.2.
https://blogs.vmware.com/virtualblocks/2016/02/10/whats-new-vmware-virtual-san-6-2/

[5] Todd Muirhead and Dave Jaffe. DVD Store benchmark.
http://en.community.dell.com/techcenter/extras/w/wiki/dvd-store

[6] TPC. TPC-E Specification.
http://www.tpc.org/tpce/

About the Author

Zach Shen is a Member of Technical Staff in the I/O Performance Engineering group at VMware. His work strives to improve the performance of networking and storage products of VMware.

Acknowledgements

The authors would like to acknowledge Amitabha Banerjee, Julie Brodeur, Duncan Epping, Christos Karamanolis, Eric Knauft, Tuyet Pham, Srinath Premachandran, Sandeep Rangasawamy, Lenin Singaravelu, Rajesh Somasundaran and Ruijin Zhou for their valuable contributions to the work presented in this white paper.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com