SSD Cache

White Paper

July 2022

ANNOUNCEMENT

Copyright

QSAN believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

Trademarks

- QSAN, the QSAN logo, QSAN.com, XCubeFAS, XCubeSAN, XCubeNXT, XCubeNAS, XCubeDAS, XEVO, SANOS, and QSM are trademarks or registered trademarks of QSAN Technology, Inc.
- Microsoft, Windows, Windows Server, and Hyper-V are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark of Linus Torvalds in the United States and/or other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Mac and OS X are trademarks of Apple Inc., registered in the U.S. and other countries.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- VMware, ESXi, and vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other countries.
- Citrix and Xen are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries.
- Other trademarks and trade names used in this document to refer to either the entities claiming the marks and names or their products are the property of their respective owners.

TABLE OF CONTENTS

FIGURES

TABLES

NOTICES

Information contained in this manual has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

PREFACE

Executive Summary

The SSD cache is a large second-level cache that uses enterprise SSDs that sit between the main memory cache and the HDDs. The SSD read and write cache improves the random IOPS of the system's read and write I/O by copying frequently accessed random data to the SSDs which are faster than the HDDs. With this technology, SSD cache can improve random read performance by up to 92 times and random write performance by up to 171 times. SSDs also provide a larger, scalable cache than memory. Therefore, it can improve overall performance by using only a few SSDs.

SSD cache is a feature available on XEVO and SANOS. It requires license to activate. This document discusses the SSD cache technology and describes its features, functions, management, and best practice.

Audience

This document is applicable for QSAN customers and partners who are familiar with QSAN products and considering using SSD cache function. Any settings which are configured with basic operations will not be detailed in this document. If there is any question, please refer to the user manuals of products, or contact QSAN support for further assistance.

Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: https://www.qsan.com/technical_support
- Via Telephone: +886-2-77206355
- (Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Skype ID: qsan.support

- (Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com

Information, Tip, and Caution

This document uses the following symbols to draw attention to important safety and operational information.

INFORMATION

INFORMATION provides useful knowledge, definition, or terminology for reference.

TIP

TIP provides helpful suggestions for performing tasks more effectively.

CAUTION

CAUTION indicates that failure to take a specified action could result in damage to the system.

---

1. Overview

Traditionally, data has been stored on traditional rotating memory, or HDDs (Hard Disk Drives) and SSDs (Solid-State Drives) are mainly used for mission-critical applications that demand high-speed storage systems, however tend to be costly. In recent years, the capacity of HDDs has increased, but their random I/O (Input / Output) has not kept pace. For some applications such as web commerce, clouds, and virtualization that require both high capacity and performance, HDDs, though capacious, simply are not fast enough and SSDs have increased capacity and have declined in cost, making them more attractive for caching in SAN storage networks.

Smart Response Technology (also known as SSD cache technology) leverages the strengths of both HDDs and SSDs, to cost-effectively meet the capacity and performance requirements of enterprise applications. Data is primarily stored on HDDs while SSDs serve as an extended HDD memory cache for many I/O operations. One of the major benefits of using SSD cache is the improved application performance, especially for workloads with frequent I/O activity. The read data of an application that is frequently accessed is copied to the SSD cache; the write data is stored to the SSD cache temporary and then flush to HDDs in bulk. So the application receives an immediate performance boost. QSAN SSD cache enables applications to deliver consistent performance by absorbing bursts of read/write loads at SSD speeds.

Another important benefit is improved TCO (Total Cost of Ownership) of the system. SSD cache copies the hot or frequency of data to SSDs in chunks. By offloading many if not most of the remaining IOPS after SSD cache, the user can fill the remainder of their storage needs with low cost, high capacity HDDs. This ratio of a small amount of SSD paired with a lot of HDD offers the best performance at the lowest cost with optimal power efficiency.

Generally, SSD read cache is particularly effective when:

- Reads are far more common than writes in the production environment, common in live database or web service applications.
- The inferior speeds of HDD reads cause performance bottlenecks.
- The amount of repeatedly accessed data is smaller than the capacity of the SSD cache.

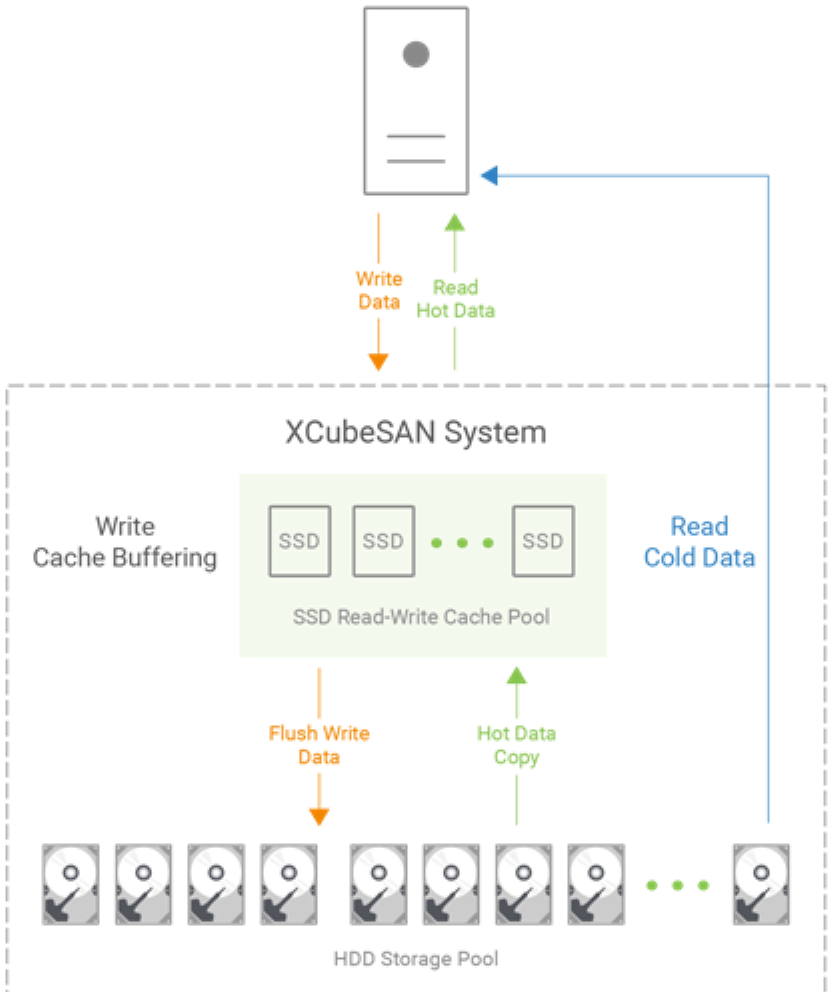

Figure 1-1 SSD Read and Write Cache

SSD read-write cache is particularly effective when:

- Reads and writes mix in the production environment, common in file service applications.
- The inferior speeds of HDD reads and writes cause performance bottlenecks.
- Same as SSD read cache case; the size of repeatedly accessed data is smaller than the capacity of the SSD cache.
- Willing to take a little risk to increase write performance because it's write cache buffering at SSD cache pool. Of course, these write data in SSD cache can be used at the next read.

2. Theory of Operation

SSD cache allows an SSD to function as read cache or write buffer for a HDD volume. In SSD read cache, it is a secondary cache that improves performance by keeping frequently accessed data on SSDs where they are read far more quickly than from the HDD volume. When reads or writes are performed, the data from the HDDs are copied into the SSD cache. Although the data is duplicated to SSD cache pool, it does not matter if the read cache pool is corrupted.

In SSD write cache, SSDs are a write buffering storage that improves performance by storing the write data in SSDs temporary where they are write far more quickly than to the HDD volume. And then the write data will be flushed to the HDD volume at the appropriate time. It may take risk of losing data during the period that the write data is stored in SSD cache if the SSD cache pool is corrupted. The write data has not yet written back to the HDD volume. So the read-write cache pool needs data protection to protect the write data.

INFORMATION

SSD cache is only available for hybrid storage models, optional and not included in the default package.

CAUTION

Using SSD read-write cache may take risk of losing data if the SSD cache pool is corrupted. User has to monitor the health of the SSD cache pool carefully.

2.1. SSD Cache Pool Architecture

An SSD cache pool is grouped to provide capacity for SSD cache usage of a dedicated storage pool. The maximum SSD cache pool quantity per system (either dual controller or single controller) is 4. The following is the storage architecture of SSD cache pool.

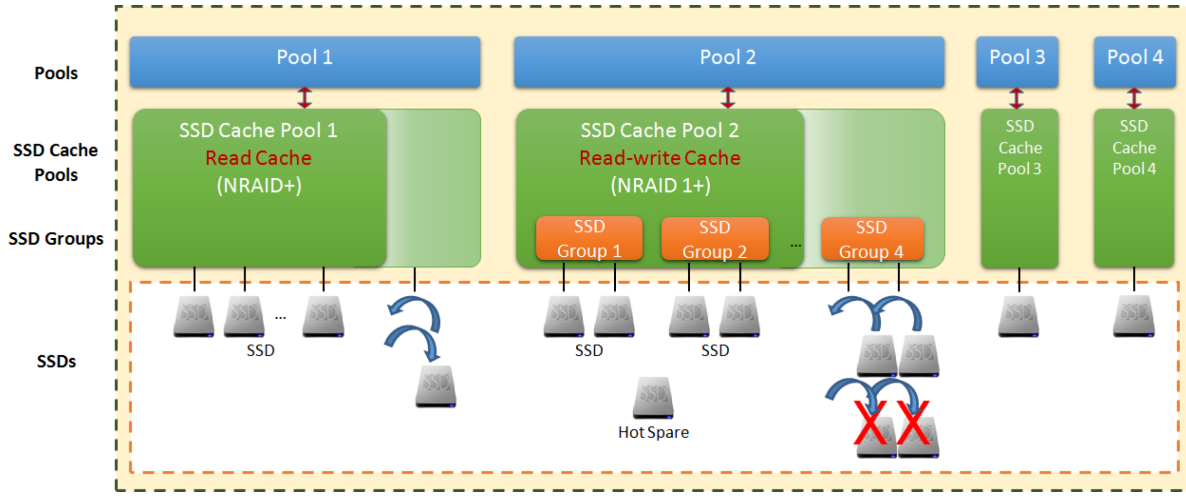

Figure 2-1 Storage Architecture of SSD Cache Pool

Take an example of SSD read cache; one or more SSDs with NRAID+ (which is described in the next section) are grouped into the "SSD Cache Pool 1" and assigned to the "Pool 1" for use. The maximum SSD quantity in an SSD cache pool is 8. When the SSD read cache is performing, the capacity of the SSD cache can be increased by adding an SSD or decreased by removing an SSD.

Another example of SSD read-write cache; two SSDs with NRAID 1+ (which is also described in the next section) are grouped into an SSD group. One or more SSD groups combine to the "SSD Cache Pool 2" and assigned to the "Pool 2" for use. When the SSD read-write cache is performing, the capacity of the SSD cache can be increased only by adding an SSD group with two SSDs at a time. The SSD read-write cache pool can set SSDs as dedicated hot spares. The maximum dedicated spare SSD quantity in an SSD cache pool is 4. The following table is the summary of the SSD cache parameters.

TIP

Note that the capacity allocated to the SSD cache pool is not counted in the regular data storage.

Table 2-1 SSD Cache Parameters

ITEM	VALUE
Maximum SSD cache pool quantity per system (either dual controller or single controller)	4
Maximum SSD quantity in an SSD cache pool	8
Maximum addressable capacity of an SSD cache pool	32 TB
Maximum quantity of volume shared in an SSD cache pool	32
Maximum dedicated spare SSD quantity in an SSD cache pool	4

The volumes in pool can be selected to enable SSD cache function. The following is the relationship between SSD cache pool and storage pool.

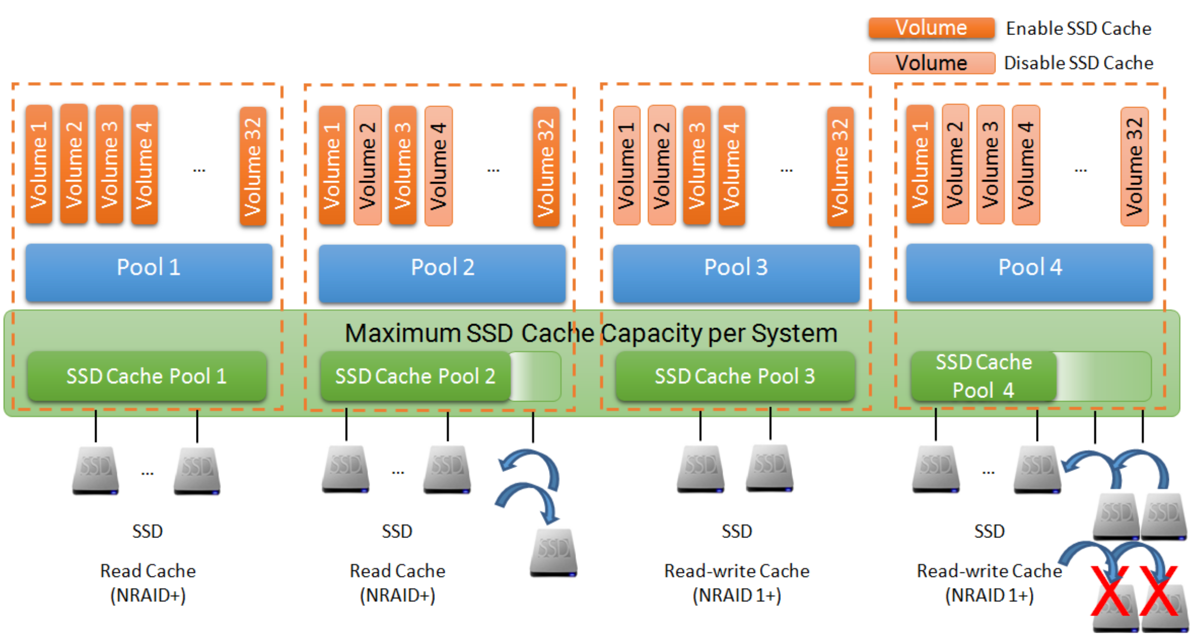

Figure 2-2 Relationship between SSD Cache Pool and Storage Pool

The volumes enabled SSD cache with hot data will consume the capacity of the SSD cache pool. Users have to consider which volumes are enabled according to the SSD cache resources. When the SSD cache is performing, you can enable or disable the volume. The maximum quantity of volume shared in an SSD cache pool is 32.

2.2. RAID Level of SSD Cache Pool

SSD read cache with NRAID+

Generally, SSD read cache uses NRAID (Non-RAID) or RAID 0 without data protection to create the SSD cache storage space. Our SSD read cache technology uses NRAID+ which is parallel NRAID without striping.

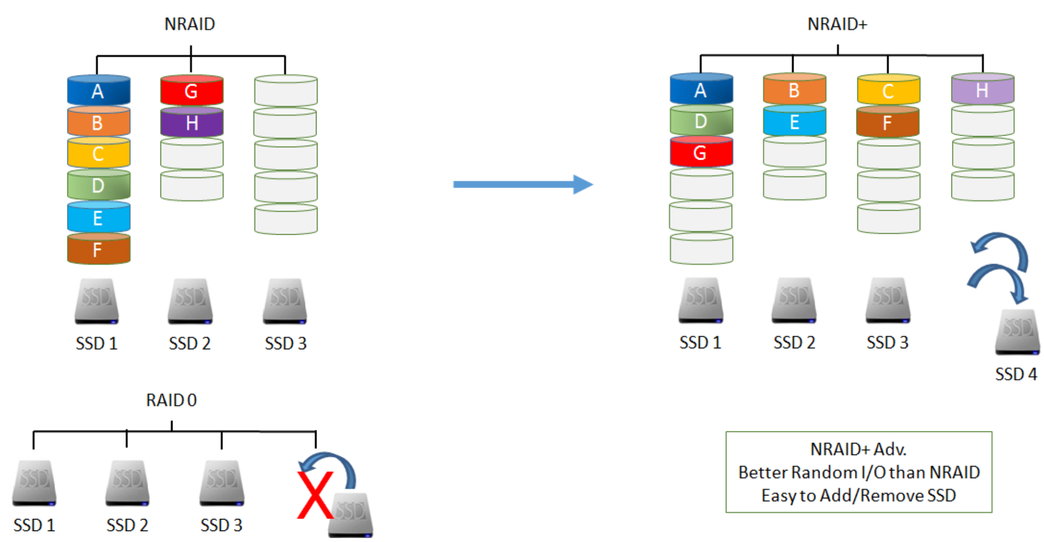

Figure 2-3 SSD Read Cache with NRAID+

NRAID is a method of combining the free space on multiple disk drives to create a spanned capacity. A NRAID is generally a spanned volume only, as it can contain disk drives of various types and various sizes. RAID 0 consists of stripes, without mirroring or parity. Adding a disk drive into RAID 0 needs to re-stripe the data. For SSD cache, it will have terrible performance when performing SSD cache and migrating RAID 0 at the same time. Let alone remove a disk drive.

Compare to the NRAID or RAID 0, NRAID+ distributes cache data over all SSDs. This NRAID+ technology combines with the advantages of NRAID and has better random I/O than NRAID. It also has the advantage of easy to add or remove SSDs from the SSD cache pool to increase/decrease the capacity.

Although the SSD read cache is a technique of duplicating data. If the SSD cache pool corrupts, the original data will be safe but stop SSD read cache.

SSD read-write cache with NRAID 1+

SSD read-write cache needs data protection, so it usually uses RAID 1 or RAID 10 level to create the SSD cache storage space. Our SSD read-write cache technology uses NRAID 1+ which is parallel NRAID with mirror.

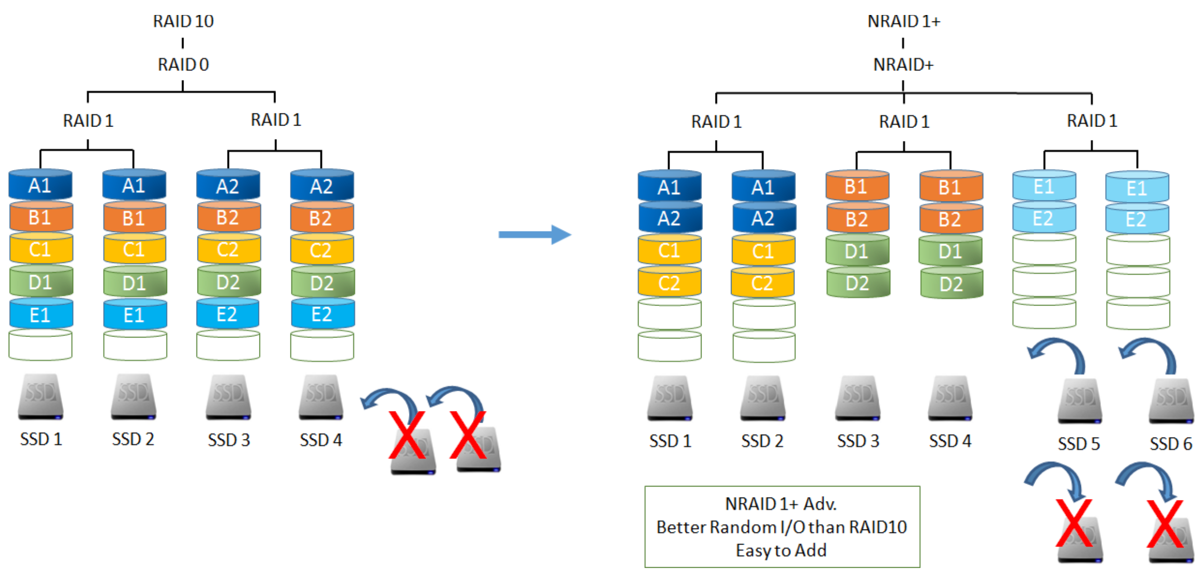

Figure 2-4 SSD Read-write Cache with NRAID 1+

RAID 10 creates a striped set from a series of mirrored drives. The same, adding pair disk drives into RAID 10 needs to re-stripe the data. So it is hard to add SSDs into SSD cache pool.

Compare to the RAID 10, NRAID 1+ distributes cache data over all mirrored SSDs. This NRAID 1+ technology has the data protection and better random I/O than RAID 10. It also has the advantage of easy to add SSDs into the SSD cache pool to increase the capacity. In this case, we disallow the SSDs to be removed from the SSD cache pool because it has the risk of losing write cache data.

If one SSD failure occurs in the SSD cache pool, the status of the SSD cache pool will be changed to degrade. It will stop write cache and start flushing write cache data into HDDs. At this time, read cache data is still working but no more new read cache data will be added into SSD cache pool. The SSD read-write cache pool can set SSDs as dedicated hot spares to prevent SSD cache pool failure. After inserting an SSD as a hot spare, the SSD read-write cache pool will be rebuilt until it is completed. And then it reverts to SSD read-write cache service.

2.3. Read / Write Cache Cases

The following describes the cases of read / write cache. In SSD read cache, the processes of read/write data will be disassembled to the following cases.

- Read Data with Cache Miss
- Read Data with Cache Hit
- Write Data in SSD Read Cache

Besides, in SSD read-write cache, the processes of read data are the same as above. But the write data is different. The processes of write data will be described on following case.

- Write Data in SSD Read-write Cache

Each case will be described below.

Read Data with Cache Miss

The following figure shows how the controller handles a host read request when some or all of the data are not in the SSD cache.

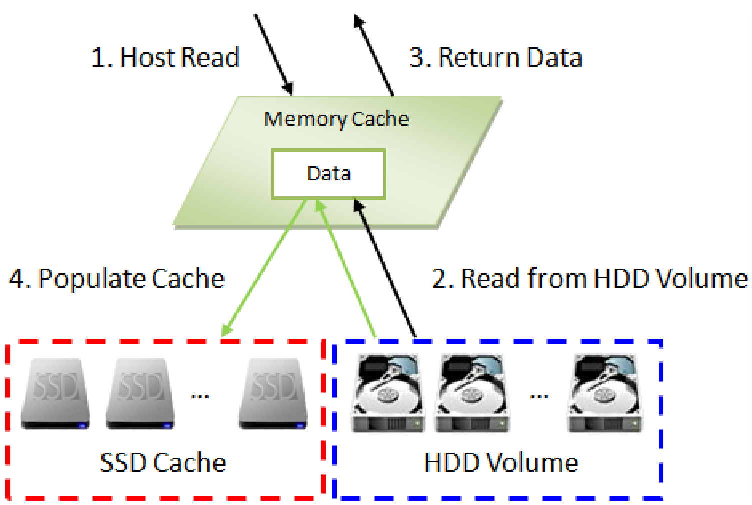

Figure 2-5 Read Data with Cache Miss

These steps are:

1. A host requests to read data. The system will check if the requested data is in memory cache or SSD cache. If not, it is called cache miss.
2. Data is read from the HDD volume because of cache miss.

3. The requested data is returned to the host. And the system will check whether the requested data is hot data.
4. If it is, the SSD cache is populated.

INFORMATION

The actions that read data from the HDD and then write to the SSD are called populating the cache.

Read Data with Cache Hit

The following figure shows how the controller handles a host read request when the data is in the SSD cache.

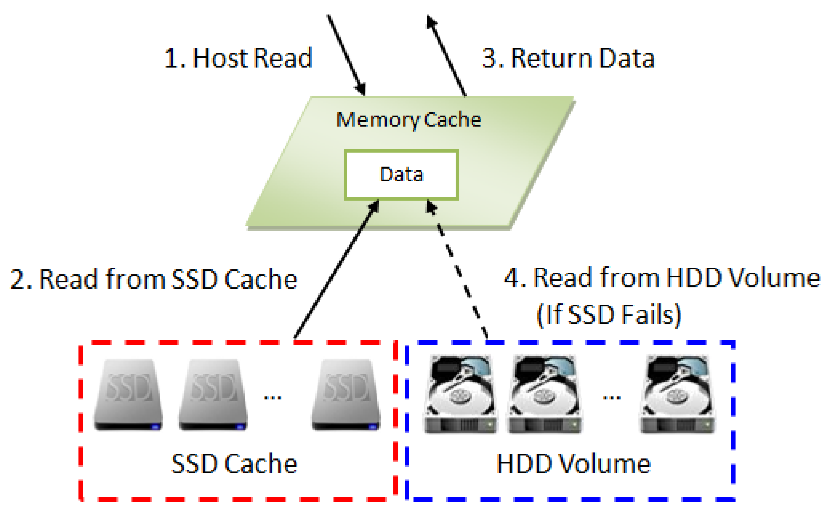

Figure 2-6 Read Data with Cache Hit

These steps are:

1. A host requests a read data. The system finds that the data is in SSD cache, so it is called cache hit.
2. Data is read from the SSD cache.
3. The requested data is returned to the host.
4. If there is an SSD cache error, data is read from the HDD volume.

Write Data in SSD Read Cache

The following figure shows how the controller handles a host write request in SSD read cache. The write data can also be frequently accessed data and populated into SSD cache for next read.

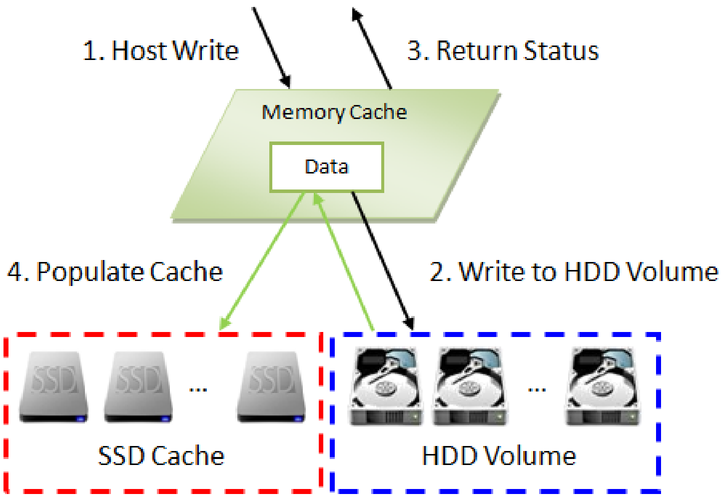

Figure 2-7 Write Data in SSD Read Cache

These steps are:

1. A host requests to write data.
2. Data is written to the HDD volume.
3. The status is returned to the host.
4. The SSD cache is populated if the write threshold is reached.

Write Data in SSD Read-write Cache

The following figure shows how the controller handles a host write request in SSD read-write cache. The write data stays in the SSD cache for a while, and will be flushed to the HDD volume at the appropriate time.

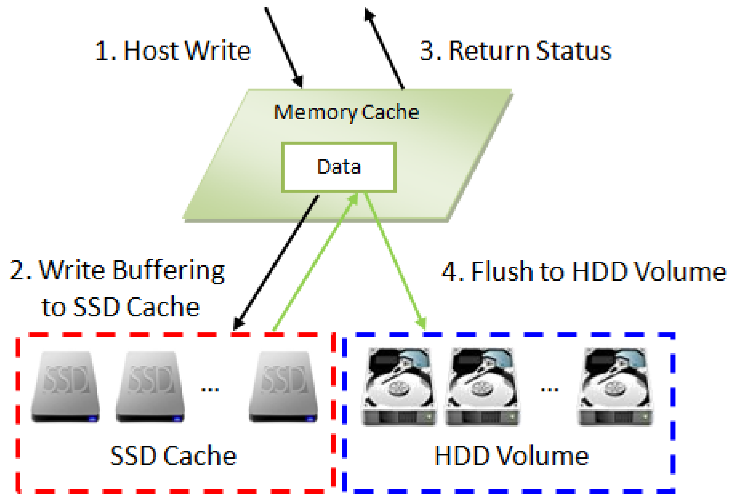

Figure 2-8 Write Data in SSD Read-write Cache

These steps are:

1. A host requests to write data.
2. Data is written to the SSD cache.
3. The status is returned to the host.
4. Data will be flushed to the HDD volume at the appropriate time.

Flush Write Data to HDD Volume

In SSD read-write cache, the write data will be flushed to the HDD volume in the following situations.

- Flush data to the HDD volumes before the SSD cache pool is deleted.
- Flush data to the HDD volumes before the related volume is disabled.
- Flush data when the system is idle.

CAUTION

It will take time to flush data from SSD cache to HDD volumes. Please do not remove SSDs before complete flushing; otherwise it may take risk of losing data. In addition, to avoid taking a long time to complete flushing, please make sure the front I/O traffic is not heavy when deleting SSD cache pool.

TIP

It won't flush data to the HDD volumes when system goes shutdown or reboots.

2.4. Populating the Cache

The actions that read data from the HDD and then write to the SSD are called populating the cache. Typically, this is a background operation that immediately follows a host read or write operation. As the goal of the cache is to store frequently accessed data, not every I/O operation should trigger a cache population, but only ones that pass a certain threshold, implemented as a counter. There are both a populate-on-read threshold, and a populate-on-write threshold.

Populate-on-read Threshold

When the same data block to be read over the threshold, it is called hot data and is populated to the SSD cache. The threshold must be greater than or equal to 1 in SSD read cache or SSD read-write cache. The value is forbidden to set 0 because of no action in read cache. The maximum of the threshold is 4. If larger than 4, the frequently accessed data is hard into SSD cache, so there is no obvious effect.

Populate-on-write Threshold

When the same data block to be written over the threshold, it is called hot data and populated to the SSD cache. The threshold must be greater than or equal to 0. If it is set to 0, no action is performed for a write cache. The value must be greater than or equal to 1 in SSD read-write cache. The value is forbidden to set 0 because of no action in write cache. The same as above, the maximum of the threshold is 4. If larger than 4, there is no obvious effect.

Operation Process

Each cache block on a HDD volume has a read and write counter associated. When a host requests to read data located in that cache block, the read count is increased. If the data is not found in the cache and the read count is greater than or equal to the populate-on-read threshold, then a cache-populate operation is performed concurrently with the host read operation. If a cache hit occurs, the data is immediately returned from the SSD cache and a populate operation is not performed. If the read count is smaller than the threshold, a populate operation is not performed.

Write cases are the same scenario as read.

2.5. SSD Cache Tuning

The SSD cache can be tuned to maximize its efficiency base on application usage. Cache block size, populate-on-read threshold and populate-on-write-threshold are the main parameters.

Cache Block Size

A large cache block suits applications where frequently accessed data is close to each other, known as a high locality of reference. A large cache block will also fill up the SSD cache quickly - this is known as the warm-up time. After the cache is warmed up, the performance would be quite good for applications with high locality of reference. Such as the file system or web service usage, the frequently accessed data are based on some concentrated files which are usually in large block size. However large cache blocks will also generate larger I/O overhead, increasing response time, especially for cache misses.

A smaller cache block size suits applications with data that is less localized, meaning the data is accessed more randomly, such as database usage. The SSD cache will fill up slower, but with more cache blocks, there is greater chance of a cache hit, especially for data with less locality of reference. With a smaller cache block size, cache usage is usually less than with a larger cache block size, but overhead is less, so the penalty for cache misses is less severe.

Population Threshold

The population threshold is the quantity of accesses at which point that cache block is copied to the SSD Cache. A higher number ensures that the cache only stores frequently accessed data so there will not be much cache turnover however it also means the cache will take longer to warm up and be fully effective. A lower number means the cache is warmed up quickly, but may cause excessive cache populations. A populate on read threshold of 2 is sufficient for many applications. Populate-on-write is useful when data that is written to is often read soon after. This is often the case in file systems. Other applications, such as database software, does not have this tendency so populate on write may sometimes even be disabled.

Table 2-2 I/O Type Table for SSD Read Cache

I/O TYPE	BLOCK SIZE (SECTORS)	POPULATE-ON-READ THRESHOLD	POPULATE-ON-WRITE THRESHOLD
Database	1MB (2,048)	2	0
File System	2MB (4,096)	2	2
Web Service	4MB (8,192)	2	0
Customization	1MB/2MB/4MB	≥ 1 and ≤ 4	≥ 0 and ≤ 4

Table 2-3 I/O Type Table for SSD Read-write Cache

I/O TYPE	BLOCK SIZE (SECTORS)	POPULATE-ON-READ THRESHOLD	POPULATE-ON-WRITE THRESHOLD
Database	1MB (2,048)	2	1
File System	2MB (4,096)	2	1
Web Service	4MB (8,192)	2	1
Customization	1MB/2MB/4MB	≥ 1 and ≤ 4	≥ 1 and ≤ 4

As you can see, there are tradeoffs for increasing or decreasing each parameter. Understanding the data locality of the application is essential and it can be useful to do some field testing to see what works best.

3. Configure SSD Cache

The following are tutorials for configuring SSD cache. Follow the steps and you can learn them quickly.

For XEVO

- SSD Cache Tutorial in XEVO

For SANOS

- SSD Cache Tutorial in SANOS
- SSD Cache and Auto Tiering Tutorial in SANOS
- Enable Trial License

4. TEST RESULTS

4.1. Test Case 1: SSD Read Cache with 1 / 2 / 4 / 8 SSDs

This test verifies the dramatic performance gains offered by SSD read cache. We test the SSD read cache with 1 / 2 / 4 / 8 SSDs. According to the design structure of RAID level in SSD read cache pool, in theory, the more SSDs used, the better SSD read cache is. We also set the populate-on-read threshold to 1 which means the data hits once and is populated to the SSD.

Test Equipments and Configurations

- Server
  - Model: HP Z840 (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
  - Model: Dell E25S (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
- Storage
  - Model: XCubeSAN XS5224D
    Memory: 16 GB (2 x 8 GB in bank 1 & 3) per controller
    Firmware 1.1.2
    HDD: 16 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
    SSD: 1 / 2 / 4 / 8 x HGST Ultrastar SSD800MH.B, HUSMH8010BSS200, 100 GB, SAS 12 Gb/s
  - HDD Pool: 1 x RAID 5 Pool with 16 x NL-SAS HDDs in Controller 1
  - HDD Volume: 2 x 45 GB in Pool
  - FC Session: 2 per Volume
- SSD Cache
  - **SSD Cache Pool: Read Cache (NRAID+) with 1 / 2 / 4 / 8 x SAS SSDs**
  - I/O Type: Customization
    Cache Block Size: 4MB
    Populate-on-read Threshold: 1

Populate-on-write Threshold: 0

- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 1
  - Outstanding (Queue Depth): 128
  - **Access Specifications: 4 KB, 100% Read, 100% Random**
- Test Scenario
  - HP server runs 1 x Volume (45 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO
  - Dell server runs 1 x Volume (45 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO

Test Diagram

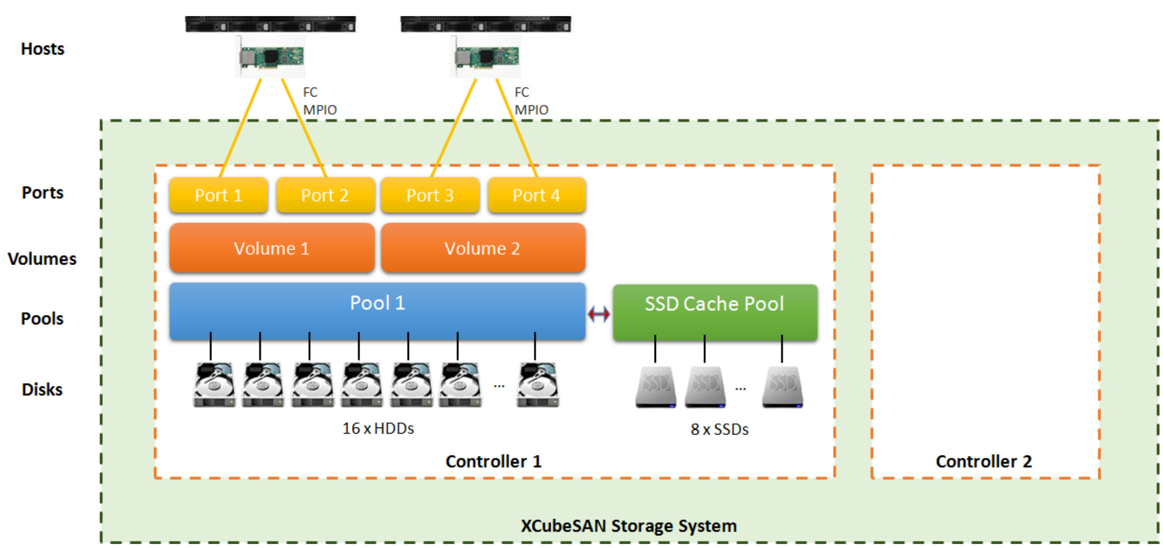

Figure 4-1 SSD Cache Test Diagram

Test Result

Table 4-1 The Test Result of SSD Read Cache with 1 / 2 / 4 / 8 SSDs

Volume Capacity	Disable SSD Cache (IOPS)	Enable SSD Cache				Improved
		No. of SSDs	SSD Capacity	Warm up Time	After Populating (IOPS)	
90GB	4,512	1	100GB	7.5 min.	121,326	2,589%
90GB	4,512	2	200GB	7 min.	192,374	4,164%
90GB	4,512	4	400GB	7 min.	214,986	4,665%
90GB	4,512	8	800GB	7 min.	216,434	4,697%

Without SSD cache, the average IOPS is 4,512. Enable SSD cache with 8 SSDs, IOPS increases to 216,434. It improves (216,434 – 4,512) / 4,512 = 46.968 == 4,697%. The warm-up time is about 7 minutes.

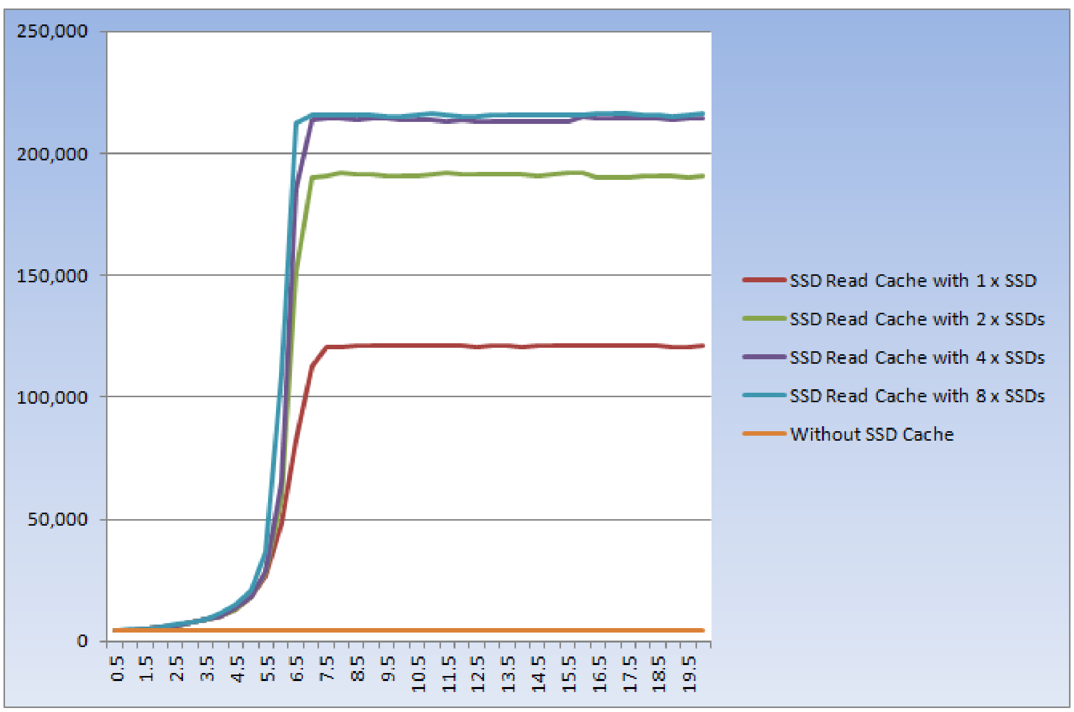

Figure 4-2 The Chart of SSD Read Cache with 1 / 2 / 4 / 8 SSDs

According to the chart for the relationship between IOPS and warm-up time, the orange line is the IOPS without SSD cache. And the others are enabled SSD cache. At the period of 1 to 7 minutes, the data is hit and read from the SSDs. If it is not hit, the data is read from HDD and populated to the SSDs. After 7 minutes, the IOPS climbs to the top and stay stable. This means that the data are all in the SSDs.

Summary

- SSD Cache 2.0 can improve random read performance by up to 47 times.
- The more SSDs used, the better SSD read cache is.

4.2. Test Case 2: SSD Write Cache with 2 / 4 / 8 SSDs

In this test, we tested the SSD read-write cache with 2 / 4 / 8 SSDs. As before, we set the populate-on-write threshold to 1 which means the data hits once and is populated to SSD.

Test Equipments and Configurations

- Server
  - Model: HP Z840 (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
  - Model: Dell E25S (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
- Storage
  - Model: XCubeSAN XS5224D
    Memory: 16 GB (2 x 8 GB in bank 1 & 3) per controller
    Firmware 1.1.2
    HDD: 16 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
    SSD: 2 / 4 / 8 x HGST Ultrastar SSD800MH.B, HUSMH8010BSS200, 100 GB, SAS 12 Gb/s
  - HDD Pool: 1 x RAID 5 Pool with 16 x NL-SAS HDDs in Controller 1
  - HDD Volume: 2 x 30 GB in Pool
  - FC Session: 2 per Volume
- SSD Cache
  - **SSD Cache Pool: Read-write Cache (NRAID 1+) with 2 / 4 / 8 x SAS SSDs**
  - I/O Type: Customization
    Cache Block Size: 4MB

Populate-on-read Threshold: 1

Populate-on-write Threshold: 1

- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 1
  - Outstanding (Queue Depth): 128
  - **Access Specifications: 4 KB, 100% Write, 100% Random**
- Test Scenario
  - HP server runs 1 x Volume (30 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO
  - Dell server runs 1 x Volume (30 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO

Test Result

Table 4-2 The Test Result of SSD Read-write Cache with 2 / 4 / 8 SSDs

Volume Capacity	Disable SSD Cache (IOPS)	Enable SSD Cache				Improved
		No. of SSDs	SSD Capacity	Warm up Time	After Populating (IOPS)	
60GB	1,660	2	100GB	13 min.	41,358	2,391%
60GB	1,660	4	200GB	9 min.	84,405	4,985%
60GB	1,660	8	400GB	6.5 min.	143,898	8,569%

Without SSD cache, the average IOPS is 1,660. Enable SSD cache with 8 SSDs, IOPS increases to 143,898. It improves (143,898 – 1,660) / 1,660 = 85.685 == 8,569%. The warm-up time is about 6.5 minutes.

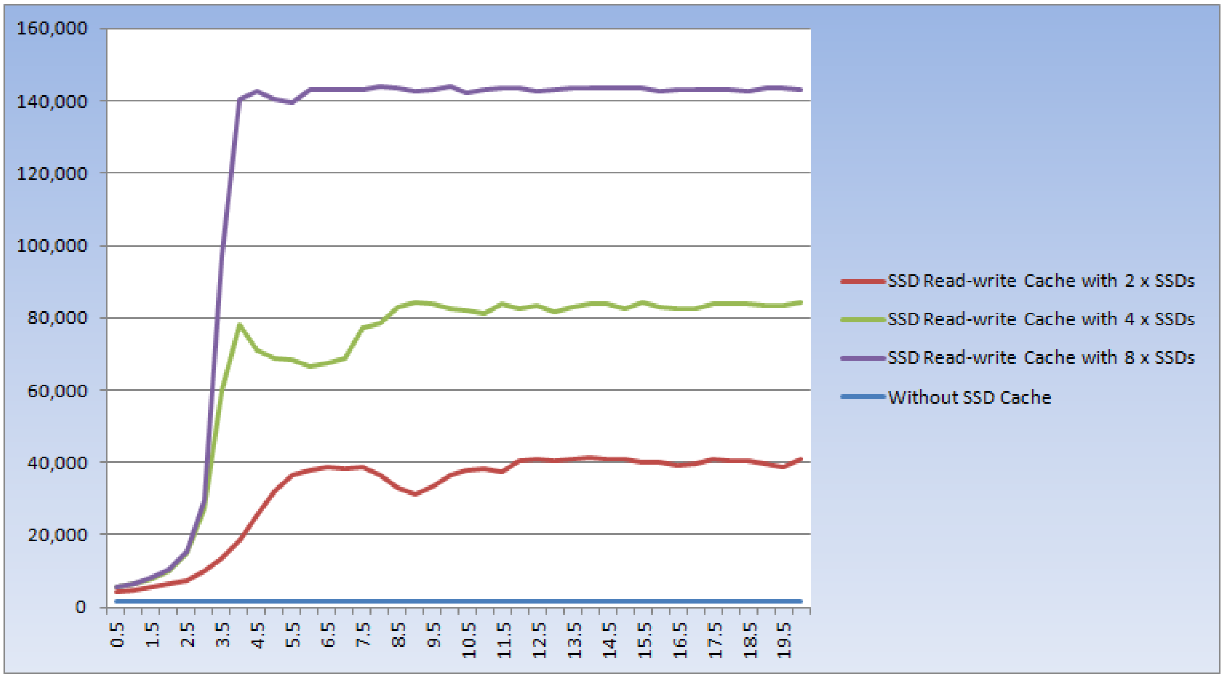

Figure 4-3 The Chart of SSD Read-write Cache with 2 / 4 / 8SSDs

Summary

- SSD Cache 2.0 can improve random write performance by up to 86 times.
- The more SSDs used, the better SSD write cache is.

4.3. Test Case 3: Simulate Database Application

This test simulates the database application. We use the Database I/O type in the configuration of SSD cache pool, and use the database access pattern (8 KB, 67% read, 100% random) to test.

Test Equipments and Configurations

- Server
  - Model: HP Z840 (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
  - Model: Dell E25S (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2

- Storage
  - Model: XCubeSAN XS5224D
    Memory: 16 GB (2 x 8 GB in bank 1 & 3) per controller
    Firmware 1.1.2
    HDD: 16 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
    SSD: 2 x HGST Ultrastar SSD800MH.B, HUSMH8010BSS200, 100 GB, SAS 12 Gb/s
  - HDD Pool: 1 x RAID 5 Pool with 16 x NL-SAS HDDs in Controller 1
  - **HDD Volume: 2 x 30 GB / 2 x 45 GB / 2 x 60 GB / 2 x 90 GB in Pool**
  - FC Session: 2 per Volume
- SSD Cache
  - **SSD Cache Pool: Read-write Cache (NRAID 1+) with 2 x SAS SSDs**
  - **I/O Type: Database**
    Cache Block Size: 1 MB
    Populate-on-read Threshold: 2
    Populate-on-write Threshold: 1
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 1
  - Outstanding (Queue Depth): 128
  - **Access Specifications: 8 KB, 67% Read, 100% Random** (Database Access Pattern)
- Test Scenario
  - HP server runs 1 x Volume (30 GB / 45 GB / 60 GB / 90 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO
  - Dell server runs 1 x Volume (30 GB / 45 GB / 60 GB / 90 GB) of RAID 5 Pool in Controller 1 via 2 x FC ports with MPIO

Test Result

Table 4-3 Test Result of Database Application

Volume Capacity	Disable SSD Cache (IOPS)	Enable SSD Cache				Improved
		No. of SSDs	SSD Capacity	Warm up Time	After Populating (IOPS)	
60GB	2,799	2	100GB	19 min.	64,809	2,215%
90GB	2,799	2	100GB	23 min.	63,660	2,174%
120GB	2,799	2	100GB	N/A min.	3,999	43%

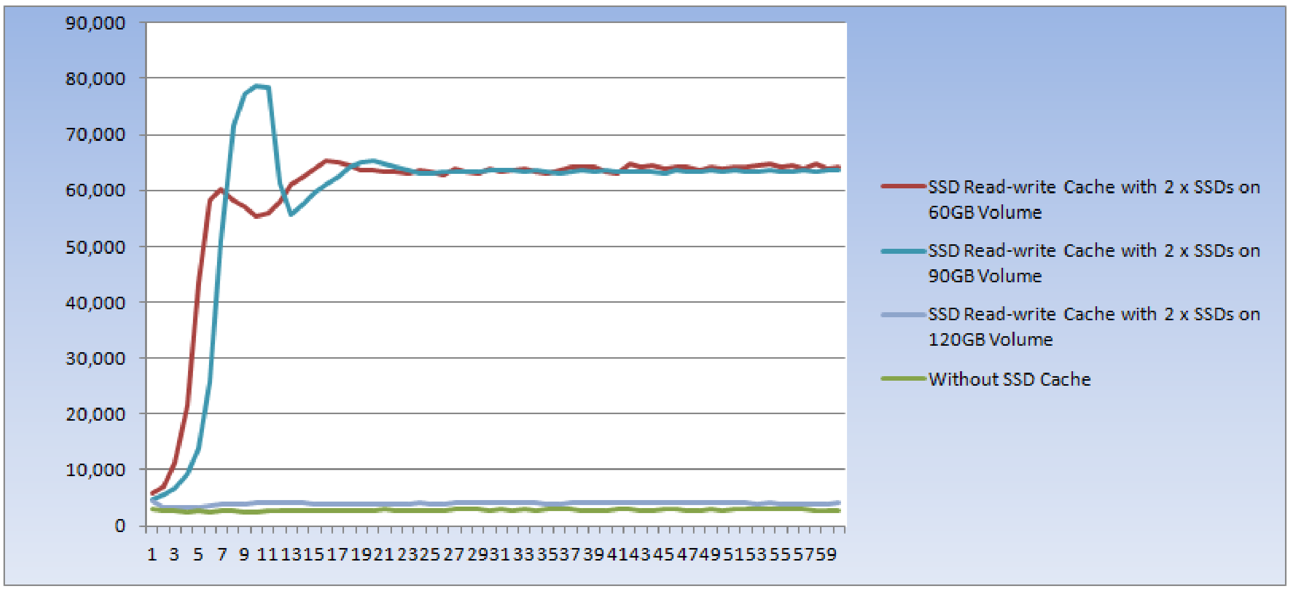

Figure 4-4 The Chart of Database Application on 60 GB / 90 GB / 120 GB / 180 GB Volumes

The result is very good when the amount of the hot data is less than the capacity of SSD cache pool.

Summary

- User has to accurately estimate the amount of hot data used to achieve the best result.

4.4. Test Case 4: Best Practice of SSD Read Cache

The cases above are tested on single controller. The following test provides the best practice of SSD read cache on dual controller. We assume that the performance can be twice than that of the test on single controller.

Test Equipments and Configurations

- Server
  - Model: HP Z840 (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
  - Model: Dell E25S (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
    FC HBA: QLogic QLE2694-SR
    OS: Windows Server 2012 R2
- Storage
  - Model: XCubeSAN XS5224D
    Memory: 16 GB (2 x 8 GB in bank 1 & 3) per controller
    Firmware 1.1.2
    HDD: 16 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
    SSD: 8 x HGST Ultrastar SSD800MH.B, HUSMH8010BSS200, 100 GB, SAS 12 Gb/s
  - HDD Pool: 1 x RAID 5 Pool 1 with 8 x NL-SAS HDDs in Controller 1
    1 x RAID 5 Pool 2 with 8 x NL-SAS HDDs in Controller 2
  - HDD Volume: 2 x 45 GB in Pool 1
    2 x 45 GB in Pool 2
  - FC Session: 2 per Volume
- SSD Cache
  - **SSD Cache Pool: Read Cache (NRAID+) with 8 x SAS SSDs**
  - I/O Type: Customization
    Cache Block Size: 4 MB
    Populate-on-read Threshold: 1
    Populate-on-write Threshold: 0
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 1
  - Outstanding (Queue Depth): 128

- **Access Specifications: 4 KB, 100% Read, 100% Random**

- Test Scenario
  - HP server runs 1 x Volume (45 GB) of RAID 5 Pool1 in Controller 1 via 2 x FC ports with MPIO
  - Dell server runs 1 x Volume (45 GB) of RAID 5 Pool2 in Controller 2 via 2 x FC ports with MPIO

Test Diagram

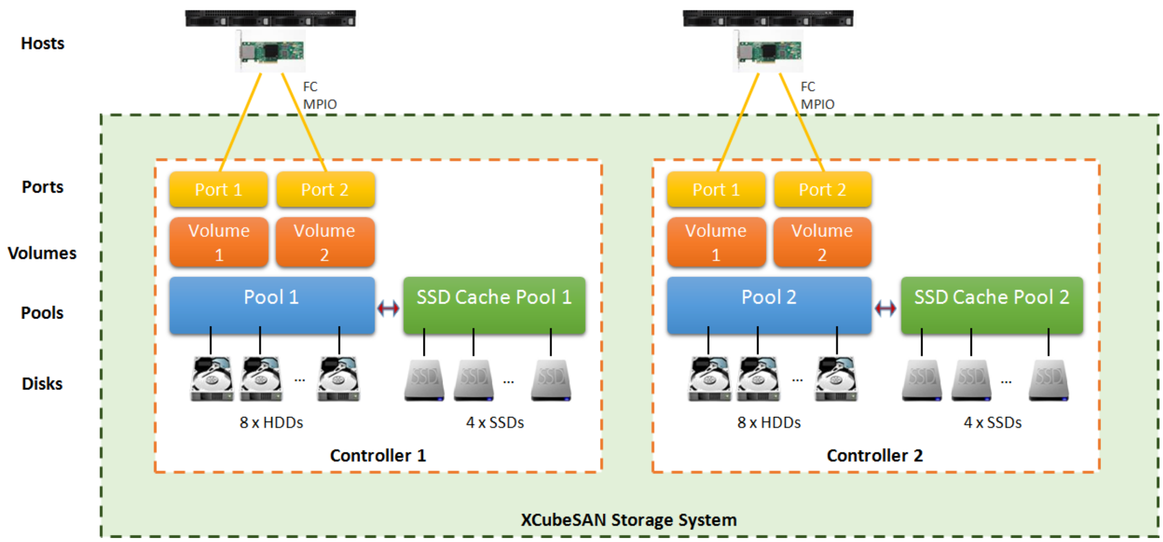

Figure 4-5 SSD Cache Test Diagram

Test Result

Table 4-4 The Test Result of SSD Read Cache with 1 / 2 / 4 / 8 SSDs

Volume Capacity	Disable SSD Cache (IOPS)	Enable SSD Cache				Improved
		No. of SSDs	SSD Capacity	Warm up Time	After Populating (IOPS)	
2 x 90GB = 180GB	4,986	2 x 4 = 8	2 x 400GB = 800GB	13.5 min.	461,037	9,147%

Without SSD cache, the average IOPS is 4,986. After enabling SSD cache with 8 SSDs, IOPS increases to 461,037. It improves (461,037 –4,986) / 4,986 = 91.466 == 9,147%. The warm-up time is about 13.5 minutes.

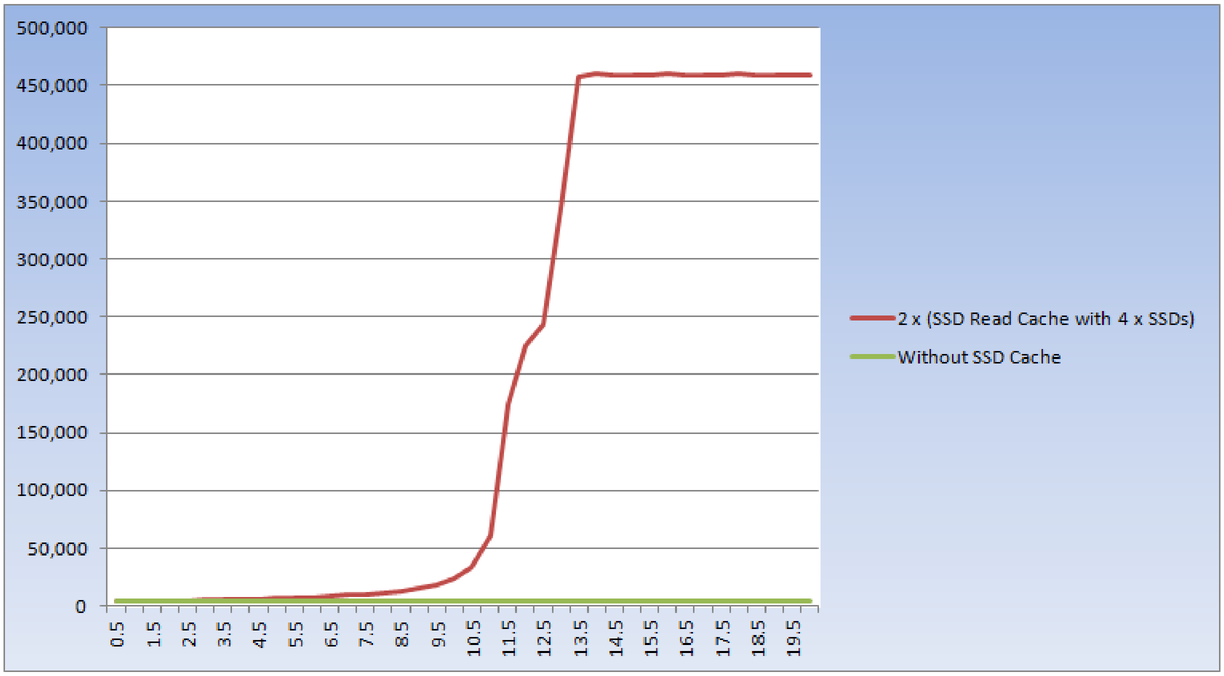

Figure 4-6 The Chart of the Best Practice of SSD Read Cache

Summary

- SSD Cache 2.0 can improve random read performance by up to 92 times.
- The test is the highest SSD cache performance of a system.

4.5. Test Case 5: Best Practice of SSD Write Cache

This test provides the best practice of SSD write cache on dual controller. As before, we assume that the performance can be twice than that of the test on single controller.

Test Equipments and Configurations

- Server
  - Model: HP Z840 (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)

FC HBA: QLogic QLE2694-SR
OS: Windows Server 2012 R2

- Model: Dell E25S (CPU: 2 x Xeon E5-2620v3 2.4Hz / RAM: 32 GB)
  FC HBA: QLogic QLE2694-SR
  OS: Windows Server 2012 R2

- Storage
  - Model: XCubeSAN XS5224D
    Memory: 16 GB (2 x 8 GB in bank 1 & 3) per controller
    Firmware 1.1.2
    HDD: 16 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
    SSD: 8 x HGST Ultrastar SSD800MH.B, HUSMH8010BSS200, 100 GB, SAS 12 Gb/s
  - HDD Pool: 1 x RAID 5 Pool 1 with 8 x NL-SAS HDDs in Controller 1
    1 x RAID 5 Pool 2 with 8 x NL-SAS HDDs in Controller 2
  - HDD Volume: 2 x 45 GB in Pool 1
    2 x 45 GB in Pool 2
  - FC Session: 2 per Volume

- SSD Cache
  - **SSD Cache Pool: Read Cache (NRAID+) with 8 x SAS SSDs**
  - I/O Type: Customization
    Cache Block Size: 4 MB
    Populate-on-read Threshold: 1
    Populate-on-write Threshold: 1

- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 1
  - Outstanding (Queue Depth): 128
  - **Access Specifications: 4 KB, 100% Write, 100% Random**

- Test Scenario
  - HP server runs 1 x Volume (45 GB) of RAID 5 Pool1 in Controller 1 via 2 x FC ports with MPIO
  - Dell server runs 1 x Volume (45 GB) of RAID 5 Pool2 in Controller 2 via 2 x FC ports with MPIO

Test Result

Table 4-5 The Test Result of SSD Read Cache with 1 / 2 / 4 / 8 SSDs

Volume Capacity	Disable SSD Cache (IOPS)	Enable SSD Cache				Improved
		No. of SSDs	SSD Capacity	Warm up Time	After Populating (IOPS)	
2 x 90GB = 180GB	1,268	2 x 4 = 8	2 x 200GB = 400GB	18 min.	217,495	17,053%

Without SSD cache, the average IOPS is 1,267. After enabling SSD cache with 8 SSDs, IOPS increases to 217,495. It improves (217,495 –1,268) / 1,268 = 170.526 == 17,053%. The warm-up time is about 18 minutes.

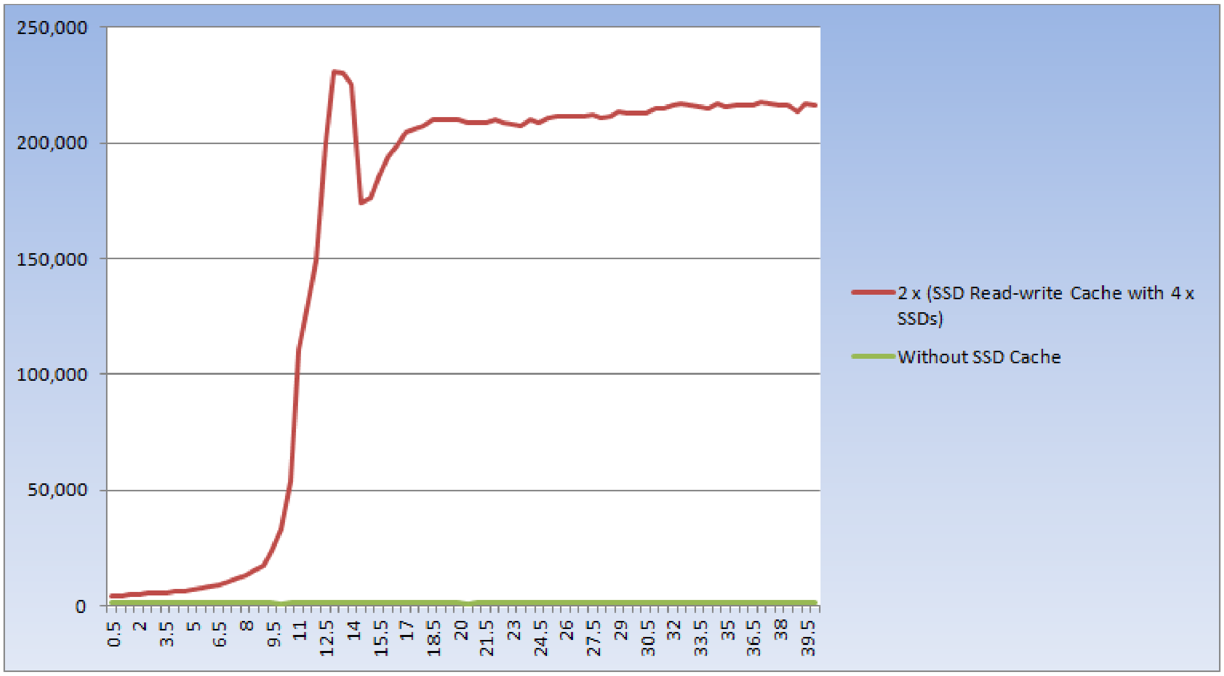

Figure 4-7 The Chart of the Best Practice of SSD Write Cache

Summary

- SSD Cache 2.0 can improve random read performance by up to 171 times.
- The test is the highest increasing rate of SSD cache in a system

5. CONCLUSION

The hybrid storage concept of storage acceleration uses the idea of hot data to accelerate I/O performance of an entire storage system. When hardware and IT administration costs are taken into consideration, it turns out that SSD cache as available in modern SAN systems is generally the best way for most businesses to gain the benefits of the faster performance from flash based storage without sacrificing the reliability of their data.

6. APPENDIX

6.1. Apply To

- XEVO firmware 2.2.0 and later
- SANOS firmware 1.1.2 and later

6.2. Reference

- XEVO Software Manual
- SANOS Software Manual