White Paper

INTRODUCTION TO EMC XTREMSW CACHE

- XtremSW Cache is a server Flash-caching solution
- XtremSW Cache accelerates reads and ensures data protection
- XtremSW Cache extends EMC FAST Suite to the server

Abstract

This white paper provides an introduction to EMC XtremSW Cache. It describes the implementation details of the product and provides performance, usage considerations, and major customer benefits when using XtremSW Cache.

August 2013

EMC²

Part Number: H11946

Table of Contents

Executive summary

Since the first deployment of Flash technology in disk modules (commonly known as solid-state drives or SSDs) by EMC in enterprise arrays, it has been EMC's goal to expand the use of this technology throughout the storage environment.

The combination of the requirement for high performance and the rapidly falling cost-per-gigabyte of Flash technology has led to the concept of a "caching tier". A caching tier is a large-capacity secondary cache using Flash technology that is positioned between the server application and the storage media.

EMC XtremSW Cache™ is a server Flash-caching software that reduces latency and accelerates throughput to dramatically improve application performance, and is most effective when coupled with EMC XtremSF - EMC PCIe Flash technology.

XtremSW Cache accelerates reads and protects data by using a write-through cache to the networked storage to deliver persistent high availability, integrity, and disaster recovery.

XtremSW Cache coupled with array-based EMC® FAST software provides the most efficient and intelligent I/O path from the application to the data store. The result is a networked infrastructure that is dynamically optimized for performance, intelligence, and protection for both physical and virtual environments.

Major XtremSW Cache benefits include:

- Provides performance acceleration for read-intensive workloads
- As a write-through cache, enables accelerated performance with the protection of the back-end, networked storage array
- Provides an intelligent path for the I/O and ensures that the right data is in the right place at the right time
- In split-card mode, enables you to use part of the server Flash for cache and the other part as direct-attached storage (DAS) for temporary data
- By offloading Flash and wear-level management onto the XtremSF PCIe card, uses minimal CPU and memory resources from the server
- Achieves greater economic value when data deduplication is enabled, by providing effective cache size larger than physical size, and longer card life expectancy.
- Works in both physical and virtual environments

As XtremSW Cache is installed in a greater number of servers in the environment, more processing is offloaded from the storage array to the server. This provides a highly scalable performance model in the storage environment.

Introduction

This white paper provides an introduction to XtremSW Cache. Flash technology provides an opportunity to improve application performance by using it in different ways in a customer environment. Topics covered in this white paper include implementation in physical and virtual environments, performance considerations, best practices, usage guidelines and characteristics, and some application-specific uses cases. For more information on EMC XtremSF PCIe cards, please refer to the "Introduction to EMC XtremSF" white paper.

Audience

This white paper is intended for EMC customers, partners, and employees who are considering the use of XtremSW Cache in their storage environment. It assumes a basic understanding of Flash technology and its benefits.

Terminology

Cache page: The smallest unit of allocation inside the cache, typically a few kilobytes in size. The XtremSW Cache cache page size is 8 KB.

Cache warm-up: The process of promoting new pages into the XtremSW Cache after they have been referenced, or a change in the application access profile that begins to reference an entirely new set of data.

Cache promotion: The process of copying data from the SAN storage in the back end to XtremSW Cache.

Hot spot: A busy area in a source volume.

Spatial locality of reference: The concept that different logical blocks located close to each other will be accessed within a certain time interval.

Temporal locality of reference: The concept that different logical blocks will be accessed within a certain time interval.

Working set: A collection of information that is accessed frequently by the application over a period of time.

Data deduplication: Eliminating redundant data by storing only a single copy of identical chunks of data, while keeping this data referenceable.

Use cases of Flash technology

There are different ways in which Flash technology can be used in a customer environment depending on the use case, application, and customer requirements. EMC's architectural approach is to use the right technology in the right place at the right time. This includes using Flash:

- As direct-attached storage
- As a cache in a server
- As a cache in an array
- As a storage tier in an array
- In an all-Flash array

In addition, there are different types of Flash with different cost structures, endurance considerations, and performance characteristics. All types of Flash have a proper place in the vast continuum of use cases. Some of the use cases for Flash are below (some may overlap):

- Applications with performance requirements with or without protection requirements that are read and write heavy. For example, temporary or mission-critical data (protected by application or operating system tools), may be a good fit for PCIe Flash in the server as DAS—for example, EMC XtremSF.
- Applications with high performance and protection requirements that are read heavy are a perfect fit for PCIe Flash in the server as a cache—for example, XtremSF hardware combined with XtremSW Cache software.
- Applications with performance and protection requirements that are read and write heavy may be a good fit for Flash in the array as a cache—for example, EMC FAST Cache on an EMC VNX storage system.
- Applications with mixed workloads and changing data "temperatures" are a perfect fit for Flash as part of a tiering strategy—for example, Fully Automated Storage Tiering for Virtual Pools (FAST VP) on an EMC VMAX storage system.
- Applications requiring highly consistent performance may be a good fit for Flash as the single tier of storage, for example, an all-Flash array.

XtremSW Cache advantages over DAS

One option to use PCIe Flash technology in the server is to use it as a DAS device where the application data is stored on the Flash. Advantages of using XtremSW Cache over DAS solutions include:

- DAS solutions do not provide performance with protection, because they are not storing the data on an array in the back end. If the server or the Flash card is faulted, you run the risk of data unavailability or even data loss. XtremSW Cache, however, provides read acceleration to the application, and

simultaneously mirrors application writes to the back-end storage array, thereby providing protection.

- DAS solutions are limited by the size of the installed Flash capacity and do not adapt to working sets of larger datasets. In contrast, after the working set of the application has been promoted into the Server Flash using XtremSW Cache, application performance is accelerated. Then, when the working set of the application changes, XtremSW Cache adapts to it and promotes the new working set into Flash over a period of time.
- DAS solutions lead to stranded sets of data in your environment that have to be managed manually. This is in contrast to application deployment on a storage array where data is consolidated and can be centrally managed.

Cache in the storage array

Another way in which some solutions use PCIe Flash technology is to use it as a cache in the storage array. However, XtremSW Cache uses PCIe Flash in the server and is closer to the application in the I/O stack. XtremSW Cache does not have the latency associated with the I/O travelling over the network to access the data.

Flash cell architecture

In general, there are two major NAND-based Flash cell technologies used in all Flash drives:

- Single-level cell (SLC)
- Multi-level cell (MLC)

A cell is the smallest unit of storage in any Flash technology and is used to hold a certain amount of electronic charge. The amount of this charge is used to store binary information.

NAND Flash cells have a very compact architecture; their cell size is almost half the size of a comparable NOR Flash cell. This characteristic, when combined with a simpler production process, enables a NAND Flash cell to offer higher densities with more memory on a given semiconductor die size. This results in a lower cost per gigabyte.

Flash storage devices store information in a collection of Flash cells made from floating gate transistors. SLC devices store only one bit of information in each Flash cell (binary), whereas MLC devices store more than one bit per Flash cell by choosing between multiple levels of electrical charge to apply to its floating gates in the transistors (See Figure 1).

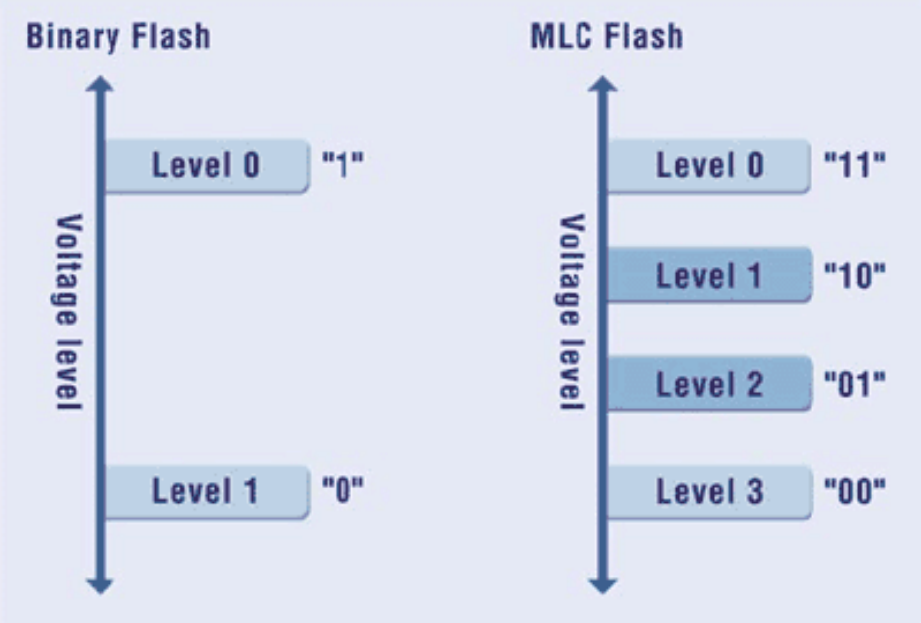

Figure 1: Comparison between SLC and MLC Flash cell data storage[1]

Since each cell in MLC Flash has more information bits, an MLC Flash-based storage device offers increased storage density compared to an SLC Flash-based version. However, MLC Flash has lower performance and endurance because of its inherent architectural tradeoffs. Higher functionality further complicates the use of MLC Flash, which requires advanced Flash management algorithms and controllers. There are two grades of MLC Flash produced today; consumer-grade (cMLC) used in consumer storage products such as thumb drives, and higher quality enterprise-grade (eMLC) used in the MLC versions of EMC XtremSF.

SLC Flash and MLC Flash offer capabilities that serve two very different types of applications—those requiring high performance at an attractive cost per bit (eMLC), and those who are less cost sensitive and seek even higher performance and endurance over time (SLC).

Taking into account the varying types of I/O profiles and requirements of enterprise applications, EMC XtremSW Cache provides customers the flexibility of choosing between eMLC and SLC Flash architectures.

Table 1 compares the SLC and MLC Flash characteristics (typical values).

[1] Kaplan, Francois. "Flash Memory Moves From Niche to Mainstream." Chip Design Magazine. April/May 2006.

Features	eMLC	SLC
Bits Per Cell	2	1
Endurance (Erase/Write Cycles)	~30K	~100K
Read Page (Average)	50 µs	35 µs
Program Page (Average)	1,600 µs	300 µs
Block Erase (Average)	5,500 µs	700 µs

Table 1: SLC and MLC Flash comparison

Although SLC NAND Flash offers a lower density, it also provides an enhanced level of performance in the form of faster reads and writes. Because SLC NAND Flash stores only one bit per cell, the need for error correction is reduced. SLC also allows for higher write/erase cycle endurance, making it a better fit for use in applications that require increased endurance and viability in multi-year product life cycles.

For more details on various Flash cell architectures, refer to the *Considerations for Choosing SLC versus MLC Flash* Technical Note on EMC Online Support (https://support.emc.com).

XtremSW Cache design concepts

Over the past decade, server processing technology has continued to advance along the Moore's Law curve. Every 18 months, memory and processing power have doubled, but disk drive technology has not. Spinning drives continue to spin at the same rate. This has caused a bottleneck in the I/O stack whereby the server and the application have capacity to process more I/O than the disk drives can deliver. This is referred to as the I/O gap, as shown in Figure 2.

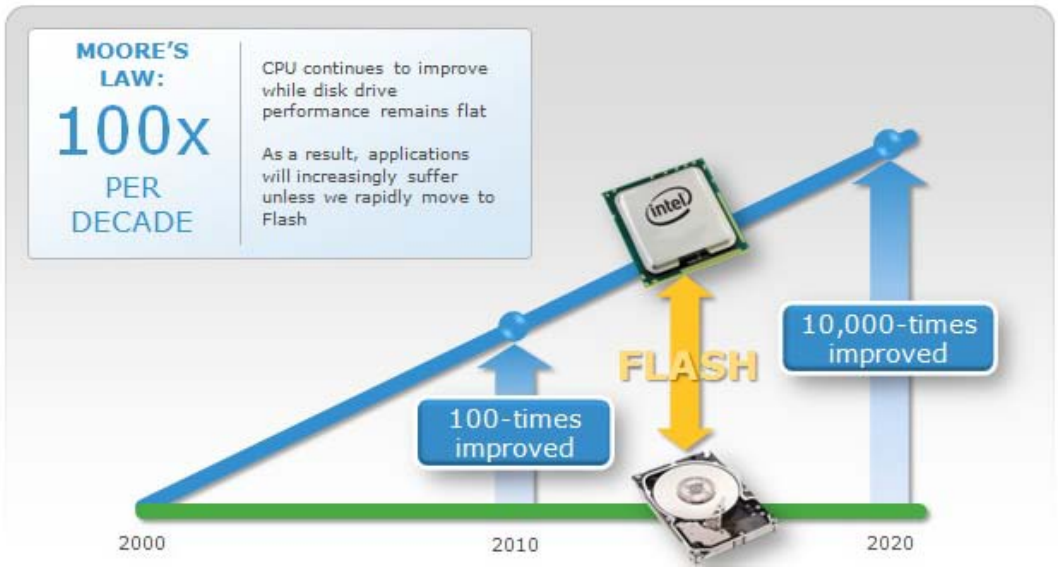

Figure 2: I/O gap between the processor and storage sub-systems

Flash drives in the storage system have helped to close this gap. Flash is a silicon technology, not mechanical, and therefore can enjoy the same Moore's Law curve.

Flash technology itself can be used in different ways in the storage environment. Figure 3 shows a comparison of different storage technologies based on the I/O per second (IOPS) per gigabyte (GB) of storage that they offer.

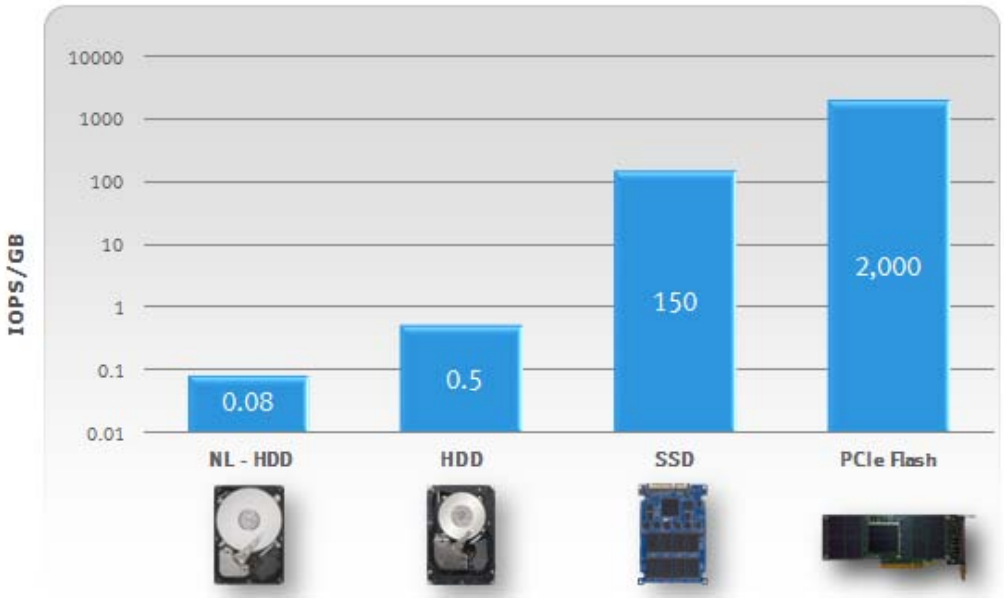

Figure 3: Comparison of storage technologies

Mechanical spinning drives provide a great dollar-per-gigabyte economic value to cold datasets, but they do not provide the best performance. Putting Flash drives in the array provides an order of magnitude better performance. Putting Flash in the server on a PCIe card can accelerate performance by even another order of magnitude over Flash drives.

FAST technology on EMC storage arrays can help place the application data in the right storage tier based on the frequency with which data is being accessed. XtremSW Cache extends FAST technology from the storage array into the server by identifying the most frequently accessed data and promoting it into a tier that is closest to the application.

EMC XtremSW Cache is server caching software that dramatically improves your application response time and delivers more IOPS. It intelligently determines, through a fully automated tiering (FAST) algorithm, which data is the “hottest” data and would benefit by sitting in the server on PCIe Flash and closer to the application. This avoids the latencies associated with I/O accesses over the network through to the storage array. Once enough data from the application working set has been promoted into server flash device, future accesses to the data will be at very low latencies. This results in an increase of performance by up to 300 percent and a decrease in latency by as much as 50 percent in certain applications.

Because the processing power required for an application’s most frequently referenced data is offloaded from the back-end storage to the PCIe card, the storage array can allocate greater processing power to other applications. While one

application is accelerated, the performance of other applications is maintained or even slightly accelerated.

EMC XtremSW Cache is EMC's newest intelligent software technology which extends EMC FAST into the server. When coupled with FAST, XtremSW Cache creates the most efficient and intelligent I/O path from the application to the data store. With both technologies, EMC provides an end-to-end tiering solution to optimize application capacity and performance from the server to the storage. As a result of the XtremSW Cache intelligence, a copy of the hottest data automatically resides on the PCIe card in the server for maximum speed. As the data slowly ages and cools, this copy is discarded and FAST automatically moves the data to the appropriate tier of the storage array – from Flash drives to FC/SAS drives and SATA/NL-SAS drives over time.

XtremSW Cache ensures the protection of data by making sure that all changes to the data continue to persist down at the storage array, and uses the high availability and end-to-end data integrity check that a networked storage array provides. Figure 4 shows a XtremSW Cache deployment in a typical environment.

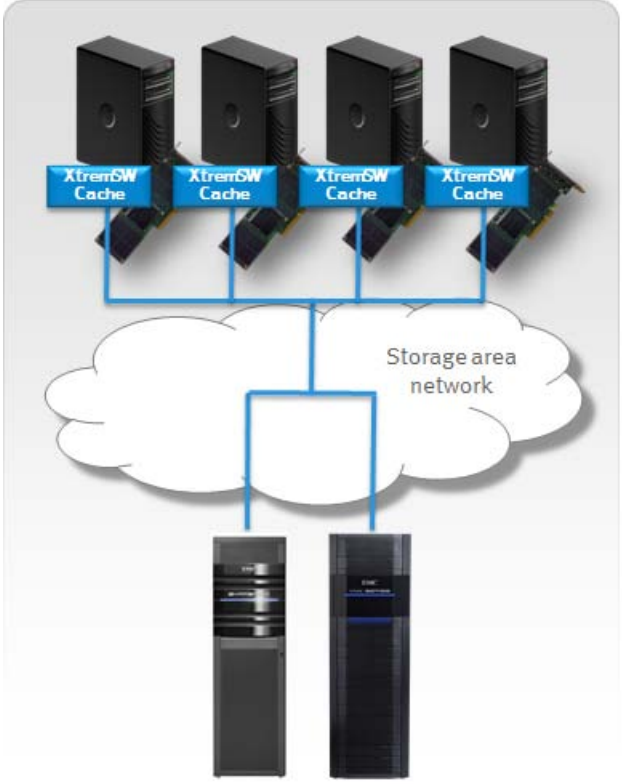

Figure 4: Typical EMC XtremSW Cache deployment

XtremSW Cache is designed to follow these basic principles:

- **Performance:** Reduce latency and increase throughput to dramatically improve application performance.
- **Intelligence:** Add another tier of intelligence by extending FAST into the server.

- **Protection:** Deliver performance with protection by using the high availability and disaster recovery features of EMC networked storage.

Business benefits

XtremSW Cache provides the following business benefits:

- Because of the way XtremSW Cache works, a portion of I/O processing is offloaded from the storage array to the server where XtremSW Cache is installed. As XtremSW Cache is installed on more servers in the environment, more I/O processing is offloaded from the storage array to the servers. The result is a highly scalable I/O processing storage model—a storage environment with higher performance capability.

  As XtremSW Cache helps in offloading workload from the storage array, the disk drives may become less busy and can be reclaimed and used for other applications.

  Note This should be done only after carefully studying the workload patterns and current utilization of disk drives.

- XtremSW Cache increases the performance and reduces the response time of applications. For some businesses, this translates into an ability to do faster transactions or searches, and more of them.

  For example, a financial trading company may be limited in the number of transactions it can process because of the number of IOPS that the storage environment can provide. XtremSW Cache increases throughout to enable more trades, thereby generating more revenue for the company.

  As another example, visitors to an eCommerce website may experience delays because of the speed at which data can be read from the back-end storage. With reduced latencies from XtremSW Cache, searches will be faster and web pages will load in less time, which in turn improves the user experience of the site.

- Typical customer environments can have multiple applications accessing the same storage system in the back end. If some of these applications are more important than others, you want to get the best performance for these applications while making sure that the other noncritical applications continue to get “good enough” performance.

  Because XtremSW Cache is installed in the server instead of the storage, it provides this flexibility. With multiple applications accessing the same storage, XtremSW Cache improves the performance of the application on the server where it is installed, while other applications on other servers continue to get good performance from the storage system. In fact, they can get a small performance boost because part of the back-end storage system’s workload gets offloaded to XtremSW Cache, and the storage system has more processing power available for these applications.

XtremSW Cache also provides you with the capability to configure XtremSW Cache at the server volume level. If there are certain volumes, like application logs, which do not need to be accelerated by XtremSW Cache, those specific devices can be excluded from the list of XtremSW Cache-accelerated volumes.

In a virtual environment, XtremSW Cache provides the flexibility to select the virtual machines and their source volumes that you want to accelerate using XtremSW Cache.

- XtremSW Cache is a server-based cache and therefore completely infrastructure agnostic. It does not require any changes to the application above it, or the storage systems below it. Introducing XtremSW Cache in a storage environment does not require you to make any changes to the application or storage system layouts.
- Because XtremSW Cache is a caching solution and not a storage solution, you do not have to move the data. Therefore data is not at risk of being inaccessible if the server or the PCIe card fails.
- XtremSW Cache does not require any significant memory or CPU footprint, Split-card mode in XtremSW Cache allows you to use part of the server Flash for cache and the other part as DAS for temporary data.

Implementation details

This section of the white paper provides details about how I/O operations are handled when XtremSW Cache is installed on the server. In any implementation of XtremSW Cache, the following components need to be installed in your environment:

- Physical XtremSF card
- XtremSF card driver
- XtremSW Cache software

In a physical environment (nonvirtualized), all the components need to be installed on the server where XtremSW Cache is being used to accelerate application performance. For more information about the installation of these components, see the *EMC XtremSW Cache Installation and Administration Guide for Windows and Linux*.

Figure 5 shows a simplified form of XtremSW Cache architecture. The server consists of two components – the green section on top shows the application layer, and the blue section on the bottom shows the XtremSW Cache components in the server (SAN HBA shown in the figure is not part of XtremSW Cache).

XtremSF hardware is inserted in a PCIe Gen2, x8 slot in the server. XtremSW Cache software is implemented as an I/O filter driver in the I/O path inside the operating system. One or more back-end storage LUNs or logical volume manager volumes are configured to be accelerated by the XtremSW Cache. Every I/O from the application to an accelerated LUN or volume is intercepted by this filter driver. Further course of action for the application I/O depends on the particular scenario when the I/O is intercepted.

In the following examples, if the application I/O is for a source volume on which XtremSW Cache has not been enabled, the XtremSW Cache driver is transparent to the application I/O, and it gets executed in exactly the same manner as if there was no XtremSW Cache driver in the server I/O stack. You can assume that the application I/O is meant for a source volume which is being accelerated by XtremSW Cache in the following examples.

Read Hit example

In this example, you can assume that the XtremSW Cache has been running for some time, and the application working set has already been promoted into XtremSW Cache. The application issues a read request, and the data is present in XtremSW Cache. This process is called “Read Hit”.

The sequence of steps is detailed after Figure 5.

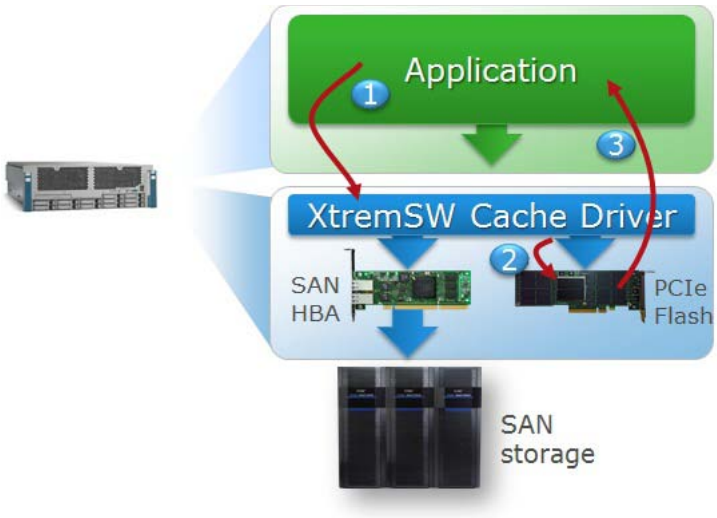

Figure 5: Read Hit example with XtremSW Cache

1. The application issues a read request that is intercepted by the XtremSW Cache driver.
2. Because the application working set has already been promoted into XtremSW Cache, the XtremSW Cache driver determines that the data being requested by the application already exists in the XtremSW Cache. The read request is therefore forwarded to the PCIe XtremSF card, rather than to the back-end storage.
3. Data is read from the XtremSW Cache and returned back to the application.

This use case provides all the throughput and latency benefits to the application, because the read request is satisfied within the server itself rather than incurring all the latencies of going over the network to the back-end storage.

Read Miss example

In this example, you can assume that the application issues a read request, and that data is not present in XtremSW Cache. This process is called “Read Miss”. The data cannot be in XtremSW Cache because the card has just been installed in the server, or the application working set has changed so that this data has not yet been referenced by the application.

The sequence of steps is detailed after Figure 6.

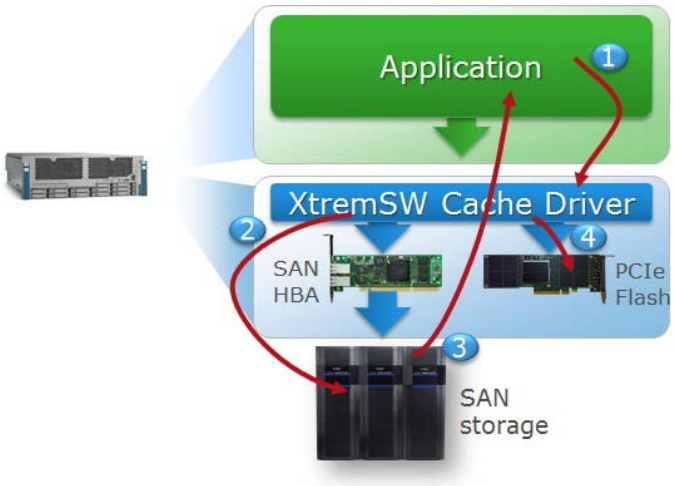

Figure 6: Read Miss example with XtremSW Cache

1. The application issues a read request that is intercepted by the XtremSW Cache driver.
2. The XtremSW Cache driver determines that the requested data is not in XtremSW Cache and forwards the request to the back-end storage.
3. The data is read from the back-end storage and returned back to the application.
4. Once the application read request is completed, the requested data is written by the XtremSW Cache driver to the XtremSFcard. This process is called "Promotion". This means that when the application reads the same data again in future, it will be a "Read Hit" for XtremSW Cache, as explained previously.

If all cache pages in XtremSW Cache are already used, XtremSW Cache uses a *least-recently-used* (LRU) algorithm to write new data into itself. If needed, data that is least likely to be used in future is discarded out of XtremSW Cache first to create space for the new XtremSW Cache promotions.

Write example

In this example, you can assume that the application has issued a write request.

The sequence of steps is detailed after Figure 7.

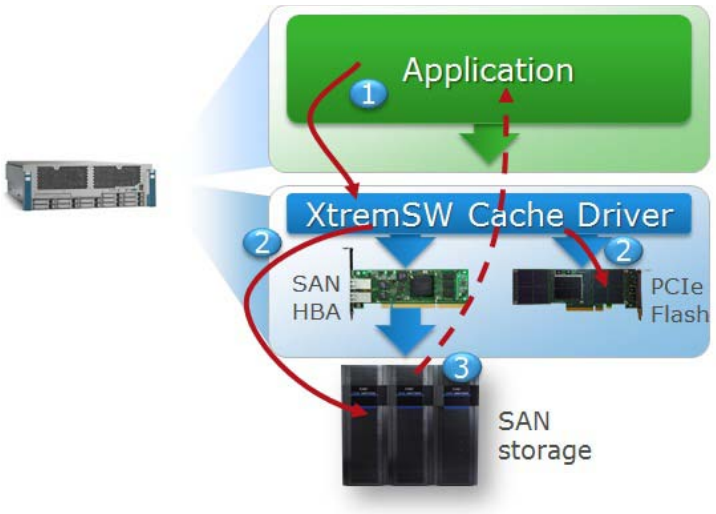

Figure 7: Write example with XtremSW Cache

1. The application issues a write request that is intercepted by the XtremSW Cache driver.
2. Since this is a write request, the XtremSW Cache driver passes this request to the back-end storage for completion, and the data in the write request is written to the XtremSFcard in parallel. If the application is writing to a storage area that has already been promoted into XtremSW Cache, the copy of that data in XtremSW Cache is overwritten. The application therefore will not receive a stale or old version of data from the XtremSW Cache. XtremSW Cache algorithms ensure that if the application writes some data and then reads the same data later on, the read requests will find the requested data in XtremSW Cache.
3. Once the write operation is completed on the back-end storage, an acknowledgment for the write request is sent back to the application.

The process of promoting new data into XtremSW Cache as explained in the previous two examples is called “Cache Warmup”. Any cache needs to be warmed up with the application working set before the application starts seeing the performance benefits. When the working set of the application changes, the cache will automatically warm up with the new data over a period of time.

VMware implementation

The implementation of XtremSW Cache in a VMware environment is slightly different from the implementation in a physical environment. In a virtualized environment, multiple virtual machines on the same server may share the performance advantages of a single XtremSW Cache card. This is shown in Figure 8.

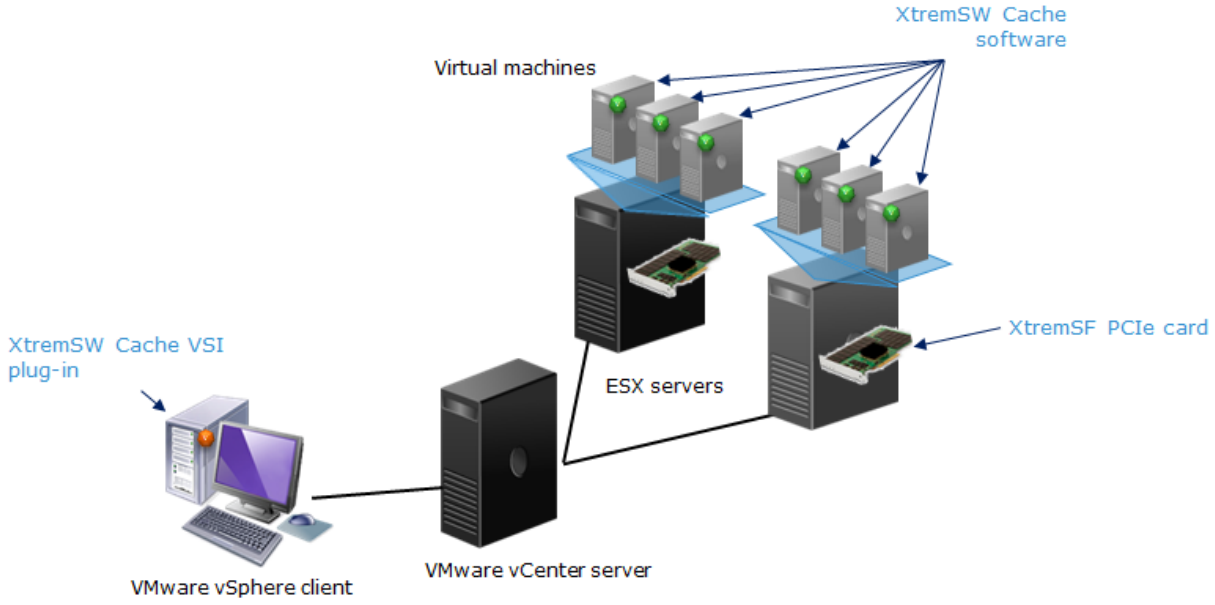

Figure 8: XtremSW Cache implementation in a VMware environment

XtremSW Cache implementation in a VMware environment consists of the following components:

- Physical XtremSF card on the VMware ESX® server
- XtremSF card driver on the ESX server

  XtremSW Cache software in each virtual machine that needs to be accelerated using XtremSW Cache. In a VMware environment, the XtremSW Cache software includes the XtremSW Cache driver, CLI package, and XtremSW Cache Agent. The XtremSW Cache software does not need to be installed on all the virtual machines in the server. Only those virtual machines that need to be accelerated using XtremSW Cache need to have XtremSW Cache software installed.
- XtremSW Cache VSI Plug-in for XtremSW Cache management in the VMware vCenter™ client
- This is usually the laptop that the administrator uses for connecting to the vCenter server.

You have to create a datastore using the XtremSFhardware on the ESX server. Once the XtremSW Cache datastore has been created, the rest of the setup can be managed using the XtremSW Cache VSI plug-in. In order for a virtual machine to use the XtremSW Cache datastore, a virtual disk (vDisk) for the virtual machines cache device must be created within the XtremSW Cache datastore. vDisks can be created either through the XtremSW Cache VSI plug-in or directly using the vSphere client. This virtual disk needs to be added to the virtual machine.

The cache configuration and management steps from this point on are similar to the steps that you would follow in a physical server environment. These can be done

using either the CLI on the virtual machine or the VSI plug-in on the vCenter client. More details on installation of XtremSW Cache in VMware environments can be found in *XtremSW Cache Installation Guide for VMware* and *XtremSW Cache VMware Plug-in Administration Guide* available on EMC Online Support.

Depending on the cache size required on each virtual machine, an appropriate sized cache vDisk can be created from the XtremSW Cache datastore and assigned to the virtual machine. If you want to change the size of XtremSW Cache on a particular virtual machine, you need to do the following:

1. Shut down the virtual machine.
2. Increase the size of the cache vDisk assigned to the virtual machine.
3. Restart the virtual machine.

XtremSW Cache is a local resource at the virtual machine level in the ESX server. This has the same consequences as any other local resource on a server. For example, you cannot configure an automatic failover for a virtual machine that has XtremSW Cache. You cannot use features like VMware vCenter Distributed Resource Scheduler (vCenter DRS) for clusters or VMware vCenter Site Recovery Manager (vCenter SRM) for replication.

You cannot use XtremSW Cache in a cluster that balances application workloads by automatically performing vMotion from heavily used hosts to less-utilized hosts. If you are planning to use vMotion functionality, you can do that in an orchestrated, manageable way from the VSI plug-in. Choose the source ESX server and the target ESX server, and a live migration operation will be performed automatically.

Both RDM and VMFS volumes are supported with XtremSW Cache. NFS file systems in VMware environments are supported as well.

Cache data deduplication

When you enable data deduplication on the PCIe cards, this creates several benefits:

- A card with data deduplication can hold more data because it is does not hold duplicate copies of the same data. The effective cache size is therefore larger than the physical cache size.
- Because duplicate data does not have to be written to the Flash card, there is a reduction in the number of writes to the card, which translates to lower card wear-out.
- XtremSW Cache data deduplication is an inline deduplication method, which means that the deduplication process is done dynamically. Deduplication is done on an 8 Kb fixed block size, and is performed on the entire card level. This increases the chance of discovering duplications in the data.
- If the XtremSW Cache-accelerated application has a high ratio of duplicated data chunks, the data deduplication gain will be higher.

- Deduplication can be managed and monitored through the CLI and through the VSI plug-in (in a VMware environment).

Active/passive clustered environments

XtremSW Cache can operate in several common clustered environments that work in active/passive or active/standby mode. When an application within a cluster is accelerated using XtremSW Cache, the data is written to the XtremSW Cache device and to a shared LUN. If there is a failover, the application can be moved to a different node and will continue to write to the shared LUN, but will not write to the cache device of the previously active node. When the application fails back to the original node, the application retrieves data from the cache device, but this device can contain stale data, as shown in Figure 9.

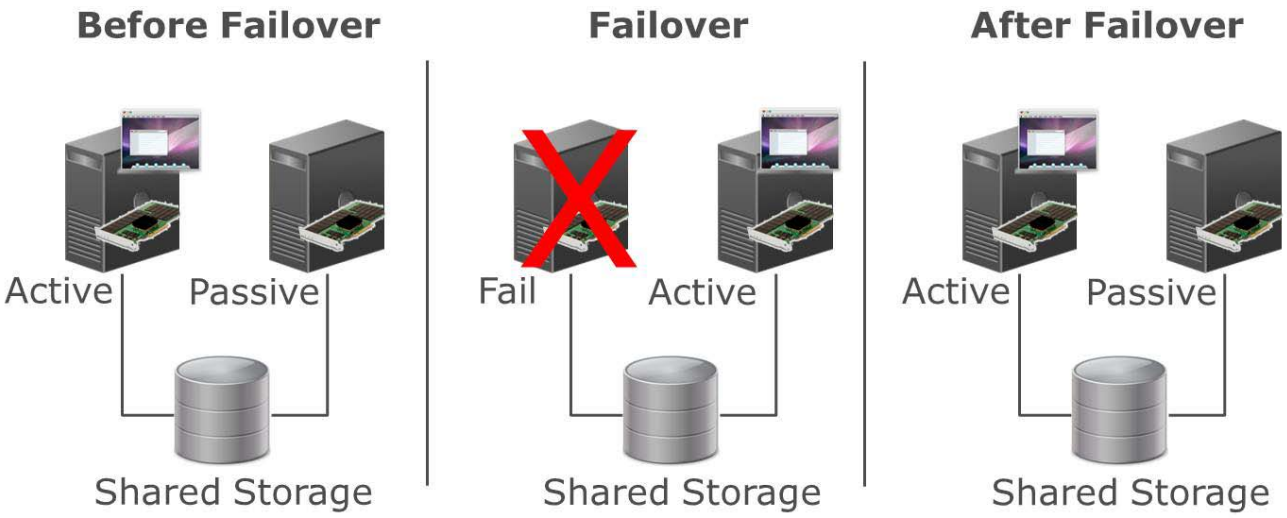

Figure 9: XtremSW Cache in an active/passive environment

During failover and failback operations between the nodes, XtremSW Cache active/passive cluster scripts automatically flush the old data from the cache to ensure that the applications never read stale data. You must ensure that the cluster scripts are installed on all nodes in the cluster – irrespective of whether that node has XtremSW Cache installed or not.

Split-card feature

EMC XtremSW Cache has a unique "split-card" feature, which allows you to use part of the server Flash as a cache and another part of the server Flash as DAS. When using the DAS portion of this feature, both read and write operations from the application are done directly on the PCIe Flash capacity in the server.

The contents of the DAS portion do not persist to any storage array. Therefore, EMC highly recommends that you use the DAS portion only for temporary data, such as operating system swap space and temp file space. This feature provides an option for you to simultaneously use the card as a caching device and as a storage device for temporary data.

When this functionality is used, the same Flash capacity and PCIe resources are shared between the cache and DAS portions. Therefore, the cache performance may be less compared to when the PCIe card is being used solely as a caching solution.

XtremSW Cache management

XtremSW Cache does not require sophisticated management software. However, there is a CLI for management of the product. There is also an option of using a VSI plug-in for XtremSW Cache management in VMware environments.

Performance considerations

XtremSW Cache is a write-through caching product rather than a Flash storage solution, so there are certain things that need to be considered when evaluating XtremSW Cache performance.

Locality of reference

The key to maximizing XtremSW Cache performance is the locality of reference in the application workload. Applications that reference a small area of storage with very high frequency will benefit the most from using XtremSW Cache. Examples of this are database indexes and reference tables. If the locality of reference is low, the application may get less benefit after promoting a data chunk into XtremSW Cache. Very low locality will result in few or no promotions and thus no benefit.

Warm-up time

XtremSW Cache needs some warm-up time before it shows significant performance improvement. Warm-up time consists mostly of promotion operations into XtremSW Cache. This happens when the XtremSW Cache has been installed and is empty. This also happens when the working dataset of the application has changed dramatically, and the current XtremSW Cache data is no longer being referenced. During this phase, the XtremSW Cache read-hit rate is low, so the response time is more like that of the SAN storage. As the XtremSW Cache hit rate increases, the performance starts improving and stabilizes when a large part of the application working set has been promoted into XtremSW Cache. In internal tests using a 1.2 TB Oracle database, the throughput increased to more than twice the baseline values in 30 minutes when TPC-C-like workload was used.

Among other things, warm-up time depends on the number and type of storage media in the back-end SAN storage. For example, a setup of 80 SAS drives will have a shorter warm-up time than a setup with 20 SAS drives. Similarly, a setup with SAS hard-disk drives (HDDs) in the back end will warm up faster than with NL-SAS HDDs in the back end. This is because NL-SAS drives typically have a higher response time than SAS drives. When you are designing application layouts, it is important to remember that there is a warm-up time before stable XtremSW Cache performance is reached.

In a demo or a Proof of Concept, the warm-up time can be speeded up by reading sequentially through the test area in 64 KB I/O size. Once the working set has been promoted, the benchmark test can be run again to compare the numbers with the baseline numbers. CLI commands can be used to find out how many pages have been promoted into XtremSW Cache. This gives you an idea of what percentage of the working set has been promoted into the cache.

If you are comparing the performance against PCIe Flash DAS solutions, the initial performance numbers of XtremSW Cache will be less because the cache needs to warm up before the stable performance numbers are shown. For DAS solutions, all read and write operations happen from the PCIe Flash and there is no warm-up phase.

Therefore, initial performance numbers should not be compared between a caching and a DAS solution.

Workload characteristics

The final performance benefit that you can expect from XtremSW Cache depends on the application workload characteristics. EMC recommends that you do not enable XtremSW Cache for storage volumes that do not have a suitable workload profile. This enables you to have more caching resources available for those volumes that are a good fit for XtremSW Cache. For example:

- **Read/write ratio**

  XtremSW Cache provides read acceleration, so the higher the read/write ratio of the workload, the more performance benefit you get.

- **Working set size**

  You should have an idea of the working set size of the application relative to the cache size. If the working set is smaller than the cache size, the whole working set will get promoted into the cache and you will see good performance numbers. However, if the working set is much bigger than the cache, the performance benefit will be less. The maximum performance benefit is for those workloads where the same data is read multiple times or where the application reads the same data multiple times after writing it once.

- **Random versus sequential workloads**

  An EMC storage array is efficient in processing sequential workloads from your applications. The storage array uses its own cache and other mechanisms such as "prefetching" to accomplish this. However, if there is any randomness in the workload pattern, the performance is lower because of the seek times involved with accessing data on mechanical drives. The storage array cache is also of limited use in this case because different applications that use the storage array will compete for the same storage array cache resource. Flash technology does not have any latency associated with seek times to access the data. XtremSW Cache will therefore display maximum performance difference when the application workload has a high degree of random component.

- **Concurrency**

  Mechanical drives in the storage array have only one or two read/write heads, which means that only limited number I/Os can be processed at any one point in time from one disk. So when there are multiple threads in the application that tries to access data from the storage array, response times tend to go up because the I/Os need to wait in the queue before they are processed. However, storage and caching devices using Flash technology typically have multiple channels internally that can process multiple I/Os simultaneously. Therefore, XtremSW Cache shows the maximum performance difference when the application workload has a high degree of concurrency. The application should request multiple I/Os simultaneously.

- I/O Size

  Large I/O sizes tend to be bandwidth-driven and reduce the performance gap between Flash technology and non-Flash technologies. Applications with smaller I/O sizes (for example, 8 KB) display the maximum performance benefit when using XtremSW Cache.

Throughput versus latency

There are some applications that can "push" the storage environment to the limit to provide as many IOPS as possible. Using XtremSW Cache in those application environments will show very high IOPS at very low response times. However, there are also applications that do not require very high IOPS, but they require very low response times.

You can see the benefit of using XtremSW Cache in these application environments. Even though the application issues relatively few I/Os, whenever the I/Os are issued, they will be serviced with a very low response time. For example, a web application may not have a lot of activity in general, but whenever a user issues a request, the response will be quick.

Other bottlenecks in the environment

XtremSW Cache helps improve throughput and reduce latencies for the applications. However, any drastic improvement in application throughput may expose new underlying performance bottlenecks or anomalies. Addressing these may include application tuning, such as increasing buffer cache sizes or other changes that increase concurrency. For example, in a typical customer deployment, a Microsoft SQL Server administrator should not enable XtremSW Cache on the log files. An inefficient storage layout design of the log files may be exposed as a bottleneck when XtremSW Cache improves the throughput and latency of the SQL Server database.

Write performance dependent on back-end array

XtremSW Cache provides acceleration to read I/Os from the application. Any write operations that the application issues still happens at the best speed that the back-end storage array can provide. At a fixed read/write ratio from an application, this tends to limit the net potential increase in read throughput. For example, if the storage array is overloaded and is processing write operations at a very slow rate, XtremSW Cache will be unable to accelerate additional application reads.

Once XtremSW Cache has been enabled on a particular source volume, every I/O from the application needs to access the XtremSF card, whether it is a read or a write operation. Usually, the processing capability of XtremSW Cache will be much greater than what the storage array can provide, therefore XtremSW Cache will not be a performance bottleneck in the data path. However, if a very large number of disks on the storage array are dedicated to a single host application, and they are fully utilized in terms of IOPS, the throughput that the storage array could provide without XtremSW Cache can be more than what XtremSW Cache can process. In this scenario, XtremSW Cache may provide minimal performance benefit to the application.

Usage guidelines and characteristics

This section provides some of the usage guidelines and salient features of XtremSW Cache.

- Because XtremSW Cache does not store any data that has not already been written on the storage array, the application data is protected and is persisted on the storage array if anything happens to XtremSF card on the server. However, the cache would need to be warmed up again after the server starts up.
- In a physical environment, you can enable or disable XtremSW Cache at the source volume level or LUN level. In a virtual environment, the XtremSW Cache capacity needs to be partitioned for individual virtual machines, as applicable. This allocated cache capacity inside the virtual machine can then be configured at vDisk-level granularity. The minimum size for the cache vDisk is 20 GB.
- There is no hard limit on the maximum number of server volumes on which XtremSW Cache can be enabled. However, if you enable it on a very large number of volumes, that may create resource starvation for those volumes that could benefit from XtremSW Cache. EMC recommends that XtremSW Cache not be enabled for those volumes that are least likely to gain any performance benefit from XtremSW Cache. This allows other volumes that are a good fit for XtremSW Cache to get the maximum processing and cache capacity resources.
- PowerPath optimizes the use of multiple data paths between supported servers and storage systems, which provides a performance boost by doing load balancing between the paths. XtremSW Cache improves the application performance even further by helping to move the most frequently accessed data closer to the application by using PCIe Flash technology for write-through caching.

  PowerPath and XtremSW Cache are complementary EMC products for scaling mission-critical applications in virtual and physical environments, including cloud deployments. Additionally, because XtremSW Cache sits above the multipathing software in the I/O stack, it can work with any multipathing solution on the market. PowerPath and XtremSW Cache are purchased separately.
- XtremSW Cache is complementary to FAST VP and FAST Cache features on the storage array. However, it is not required to have FAST VP or FAST Cache on the storage array to use XtremSW Cache.
- XtremSW Cache only accelerates read operations from the application. The write operations will be limited by the speed with which the back-end array can process the writes.
- If multiple XtremSF cards are being used in the same server, XtremSW Cache software creates individual cache devices on each PCIe card. EMC

recommends that you spread the anticipated workload from different applications evenly between various cache devices to get the maximum performance benefit.

- Customers can select the maximum I/O size from the application that XtremSW Cache will intercept and cache in the PCIe card. If this parameter is set to a very large value, more cache pages are reserved for caching larger-size application I/Os. You should do a proper analysis before changing the default values.

Specifications

- The cache page size that is used internally in XtremSW Cache has defulat value of8 KB, but it will work seamlessly with applications where the predominant I/O size is other than 8 KB. The cache page size is customizable, which enables further performance optimization for applications with larger predominant I/O size
- One instance of the XtremSF Cache software is required on a server, even if multiple XtremSFcards are used in the server. However, it is not possible to combine the capacities of both cards to create one large cache.
- XtremSW Cache supports the following connection protocols between the server and the storage array:
    - 4 Gb/s Fibre Channel
    - 8 Gb/s Fibre Channel
    - 1 GB/s iSCSI
    - 10 GB/s iSCSI
    - FCoE
- XtremSW Cache is compliant with the Trade Agreements Act (TAA). The following main requirements are certified as not applicable to XtremSW Cache:
    - FIPS 140-2
    - Common Criteria
    - Platform Hardening
    - Research Remote Access

Constraints

- XtremSW Cache does not provide connectivity between the server and the SAN storage array. You still need to use an HBA card to connect to the back-end storage array where the data is eventually stored.
- Blade servers require a customized version of the card and therefore do not support half-height half-length form-factor XtremSF cards. Such customized version of the card (called Mezzanine card) is currently available for Cisco UCS

blade servers, in 400 GB and 800 GB capacities. The XtremSW Cache software can then be used to provide performance and protection.
XtremSW Cache can be installed also on servers that enable a PCIe expansion card to connect to their chassis can work with an XtremSF card. For the most current list of supported operating systems and servers, refer to *E-Lab Interoperability Navigator*.

- XtremSW Cache is currently not supported in shared-disk environments or active/active clusters. However, shared disk clusters in VMware environments are supported since XtremSW Cache is implemented at the virtual machine level rather than the ESX server level.
- The default I/O size is set to 64 KB, which XtremSW Cache driver intercepts. It can be enlarged to 128 KB or reduced to 32KB from the user interface if needed. Any I/O larger than 128 KB will not be intercepted by XtremSW Cache. Applications with larger I/O sizes are typically bandwidth sensitive and have sequential workloads, which would not benefit from a caching solution like XtremSW Cache.
- Due to Oracle's database implementation, cache deduplication shows little to no benefit at all on such environments.

Stale data

- **Stale data because of storage array snapshots**

  If any operations modify the application data without the knowledge of the server, it is possible to have stale data in XtremSW Cache. For example, if a LUN snapshot were taken on the array and later used to roll back changes on the source LUN, the server would have no knowledge of any changes that had been done on the array. This would result in XtremSW Cache having stale data that had not been updated with the contents from the snapshot. As a workaround in this case, you need to manually stop and restart the XtremSW Cache software driver for the source volume.

  Note The whole cache device does not need to be stopped, only the source volume on which the snapshot operations are being done needs to be stopped. When you restart the XtremSW Cache software driver on the source volume, a new source ID is automatically generated for that source volume, which invalidates the old XtremSW Cache contents for the source volume and starts caching the new data from the snapshot. The application then gets access to new data from the snapshot.

- **Stale data in VMware environments**

  If you use VMware, you should be careful when the VMware snapshot feature is being used. XtremSW Cache metadata is kept in the virtual machine memory, therefore it will be a part of the virtual machine snapshot image when a virtual machine snapshot is taken. This means that when this snapshot image is used to

roll back the virtual machine, the old metadata is restored and potentially causes data corruption.

You must clear the XtremSW Cache before the virtual machine “suspend” and “resume” operations. This is handled using scripts that are automatically installed when the XtremSW Cache Agent is installed in the virtual machine. These scripts are automatically invoked when these virtual machine operations are run.

In Windows environments, you should take care to ensure that other programs or installations in the virtual machine do not change the default suspend/resume scripts in such a way that the XtremSW Cache scripts are not executed on those events. XtremSW Cache can also be cleared manually before suspend and resume operations in the virtual machine, if needed.

Application use case and performance

XtremSW Cache helps you boost the performance of your latency and response-time sensitive applications — typically applications such as database applications (for example, Oracle, SQL Server, and DB2), OLTP applications, web applications, and financial trading applications. XtremSW Cache is not suitable for more write-intensive or sequential applications such as data warehousing, streaming media, or Big Data applications. Use cases are shown in Figure 10.

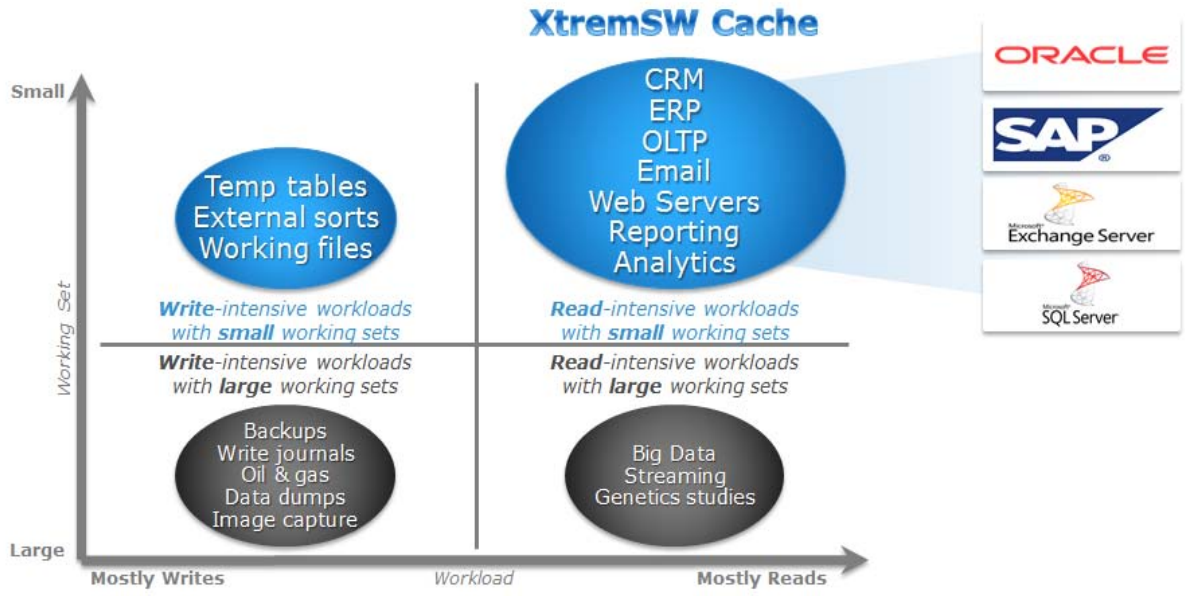

Figure 10: XtremSW Cache Use Cases

The horizontal axis represents a typical read/write ratio of an application workload. The left side represents write-heavy applications such as backups. The right side represents read-heavy applications such as reporting tools.

The vertical axis represents the locality of reference or "skew" of the application's workload. The lower end represents applications that have very low locality of reference, and the top side represents applications where a majority of the I/Os go to a very small set of data.

You will achieve the greatest results with XtremSW Cache in high-read applications and applications with a highly concentrated skew of data.

Test results

EMC conducted application-specific tests with XtremSW Cache to determine potential performance benefits when this product is used. Here is a summary of the XtremSW Cache benefits with a couple of applications (all test were done with an XtremSF card):

- Microsoft SQL Server

With a TPC-E like workload in a 750 GB Microsoft SQL Server 2008 R2 environment, the number of transactions increased three times and the latency was reduced by 87 percent when XtremSW Cache was introduced in the configuration.

- Oracle
  - With a TPC-C-like workload in a 1.2 TB Oracle 11*g* R2 physical environment, the number of transactions increased three times and the latency was reduced by 50 percent when XtremSW Cache was introduced in the configuration. The test workload had 70 percent reads and 30 percent writes.
  - In a VMware setup with 1.2 TB Oracle Database 11*g* R2 and TPC-C-like workload, the number of transactions increased by 80 percent when XtremSW Cache was introduced in the configuration. The test workload had 70 percent reads and 30 percent writes.

For more information on application-specific guidelines and test results, refer to the list of white papers provided in the References section.

Conclusion

There are multiple ways in which Flash technology can be used in a customer environment today, for example, Flash in the server or the storage array, Flash used as a cache or a tier. The key, however, is the software that brings all of this together, using different technologies at the right place and time for the right price.

- XtremSW Cache uses EMC FAST technology in the storage array and FAST in the server to provide this benefit most appropriately, as simply and as easily as possible.
- XtremSW Cache dramatically accelerates the performance of read-intensive applications.
- XtremSW Cache software caches the most frequently used data on the server-based PCIe card, which puts the data closer to the application. It extends FAST technology into the server by ensuring that the right data is placed in the right storage at the right time.
- The intelligent caching algorithms in XtremSW Cache promote the most frequently referenced data into the PCIe server Flash to provide the best possible performance and latency to the application.
- XtremSW Cache provides you with the flexibility to use the same PCIe device as a caching solution and a storage solution for temporary data.

XtremSW Cache suits many but not all customer environments, and it is important that you understand the application workload characteristics properly when choosing and using XtremSW Cache.

XtremSW Cache protects data by using a write-through algorithm, which means that writes persist to the back-end storage array. While other vendors promise the performance of PCIe Flash technology, EMC XtremSW Cache provides this performance with protection.

References

The following documents are available on EMC Online Support:

- *EMC XtremSW Cache Datasheet*
- *XtremSW Cache Installation and Administration Guide for Windows and Linux*
- *XtremSW Cache Release Notes for Windows and Linux*
- *XtremSW Cache Installation Guide for VMware*
- *XtremSW Cache Release Notes for VMware*
- *XtremSW Cache VMware Plug-in Administration Guide*
- *Considerations for Choosing SLC versus MLC Flash*
- *EMC XtremSW Cache Accelerates Oracle - EMC XtremSW Cache, EMC Symmetrix VMAX and VMAX 10K, Oracle Database 11g*
- *EMC XtremSW Cache Accelerates Virtualized Oracle - EMC XtremSW Cache, EMC Symmetrix VMAX and VMAX 10K, VMware vSphere, Oracle Database 11g*
- *EMC XtremSW Cache Accelerates Oracle - EMC XtremSW Cache, EMC VNX, EMC FAST Suite, Oracle Database 11g*
- *EMC XtremSW Cache Accelerates Microsoft SQL Server - EMC XtremSW Cache, EMC VNX, Microsoft SQL Server 2008*
- *EMC XtremSW Cache Accelerates Virtualized Oracle - EMC XtremSW Cache, EMC Symmetrix VMAX, EMC FAST Suite, VMware vSphere, Oracle Database 11g*