THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2012012 May 9, 2012

Speeding Up Apps Using EMC's New VFCache

Analyst: Jim Baker

Management Summary

The idea is very simple, actually: *Mix and match server and disk technology to speed up your IT applications.* For some time, rotating disk technologies have become ever denser but they have not picked up any appreciable end-to-end I/O speed (for quicker information transfer times). This means they can store more, but getting the data returned to the server for processing is not that much faster. Primarily, this phenomenon is due to the physics associated with spinning magnetic media and the need for a disk's read/write heads to be positioned over the location where the desired data is located. Rotating storage has many real constraints: size of the media, how fast the media can spin, density of the media, plus several built-in limits (heat creation and dissipation), and the inherently mechanical nature of rotating storage devices.

EMC, as a leading storage vendor, took the first step in addressing the mismatch of capacity with I/O speed in 2008, by adding SSDs (Solid State Devices or Drives) to their *VMAX* storage array[1]. The result was blazing fast transfers for Tier 1 data, the organization's most critical (and also usually most valuable) data. However, they did not just employ faster SSD hardware. They also added *FAST (Fully Automated Storage Tiering)* at the same time. This software automatically balances data placement across hardware resources in the system using autonomic promotion and demotion of data across tiers. If demand for specific data is "hot" or in intense demand by the server, it is ***pro***moted automatically and moves to the highest (and fastest) tier of storage available, in this case usually SSDs inside the array. Conversely, if a data's demand appears to "cool," it is autonomically ***de***moted to a lesser tier, usually one with slower access and transfer times. Below the SSD tier, you may find high-performance spinning disks (rotating at 15,000 RPMs). Right below this tier, you will find slower-spinning disks, such as 7200 RPM SAS or SATA drives. This slower tier has very high capacity but slower throughput, making it appropriate for many backup and archival applications, where slightly slower response times are acceptable.

Unfortunately, many modern applications have a continuing, seemingly voracious need for even faster data transfers. Likewise, modern high-speed processor technology, especially now with many cores, means that storage may not be feeding data fast enough to the waiting applications, resulting in a mismatch between the processor's appetite for data and the I/O subsystem's ability to fetch and deliver it in a timely manner. EMC's most recent innovation is to put *VFCache* on a server's high-speed PCIe bus. The result is even faster performance (microsecond response times versus millisecond response times associated with using rotating media.) That is three orders of magnitude faster! The speed improvement comes from not having to jump on the I/O bus and through the network between the server and its external storage, for the purposes of a data transfer, and not having to wait for the rotating disks to access the needed data. Instead, the anticipated (hot) data remains largely inside the VFCache within the server and thus is ready for work. To find out why this technology is more than just an interesting IT evolution and more like a game changer, please read on.

IN THIS ISSUE

- What is VFCache? 2
- How Does VFCache Work? 4
- Additional Insights 5
- Conclusion 6

[1] See **The Clipper Group Navigator** dated January 14, 2008, entitled *EMC Symmetrix DMX to Offer Flash Drives for Ultra-High Performance*, available at http://www.clipper.com/research/TCG2008002.pdf.

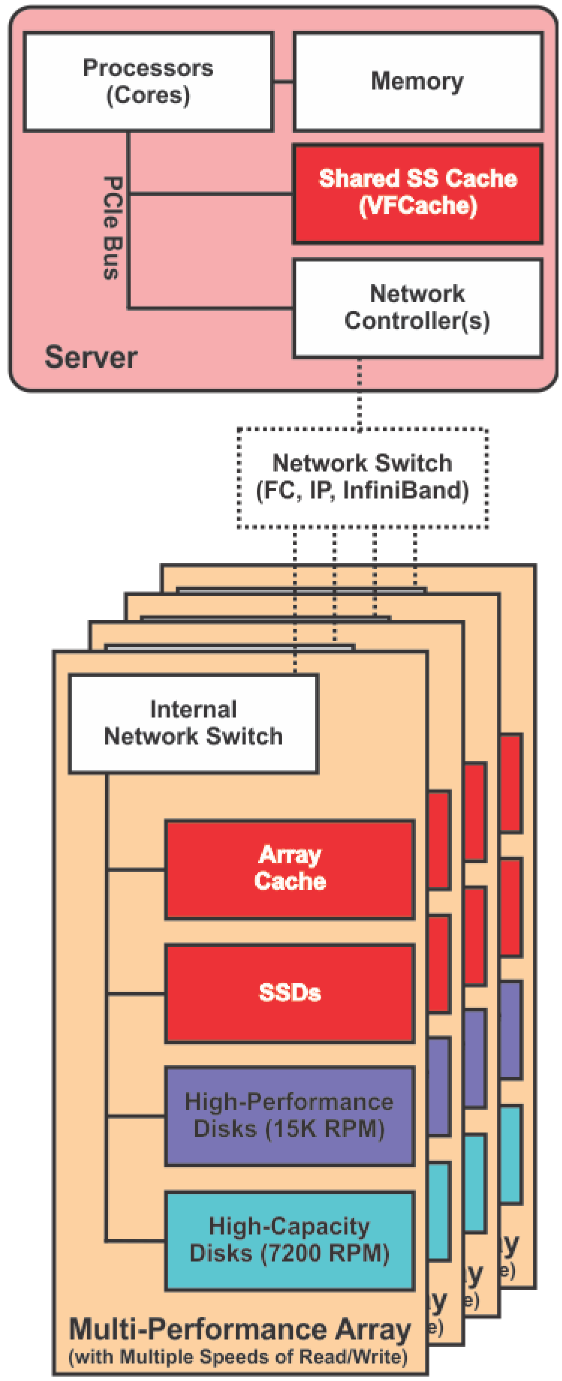
What is VFCache?

VFCache is an EMC hardware/software solution that sits on an application or database server's native PCIe bus, the path connecting the server processor core/server memory complex with external peripherals, such as network switches and storage. As such, *VFCache* is physically and logically closer to the server core running an application needing the data versus data out on the peripheral data store. (See Exhibit 1, to the right, that illustrates the nearness of VFCache to the processor's cores and the much greater distance that continually must be traversed to reach external storage.) While this diagram is only a representation, you can see why transfers between the VFCache and the cores are extremely fast; VFCache avoids jumping onto the I/O bus, then through at least one switch and then on the InfiniBand, Fiber Channel, or IP network, and then through at least another switch to reach the external storage (with all of its own built-in delays). Thus, VFCache avoids all of that switching and networking overhead once the needed data is in the cache. **Essentially, VFCache creates a new class of storage infrastructure, which we are calling *Tier N* (for "near", which will be explained shortly).**

VFCache is part of EMC's end-to-end storage tiering strategy. It extends the *FAST* hierarchy beyond the array and into the server. In most data centers, there are multiple, hierarchical tiers of storage appropriate for differing types and uses of data, often called *Tiers 0, 1, 2,* and *3.*[2] For simplicity's sake and for purposes of this paper, let's just say that Tiers 0 and 1 usually contain high-value (or mission-critical) data that runs the business, such as online transaction processing and database access, where speedy response time is required. EMC's VFCache is a new, even faster tier, Tier N, where the most critical data can be placed and whose applications and use will be accelerated by its nearness to the processor cores. With VFCache,

Exhibit 1 — Data Residing on Shared Solid State Cache Sits Much Closer to the Processor Cores than when Stored in an External Disk Array

Takeaway – It's all about speed. It takes much longer to read from or write to the Array (how long depends on the media where the data sits or is going) than it does to read/write to the SSD cache on the PCIe bus inside the server.



Source: The Clipper Group, Inc.

[2] The numberings of tiers is not universal. It used to be that Tier 1 was the top tier and consisted of 15K RPM hard disks. When speedier solid-state drives were added, many called these "Tier 0" (zero). This numbering scheme has inherent problems, because there may be storage tiers above the SSD tier especially if you consider the presence of caches, such as processor internal caches, hardware disk drive caches, global array memories preceding rotating storage, plus the fact that disks have differing densities, spin speeds, and transfer times. With VFCache, this "something faster" needed a meaningful tier name, hence we chose "Tier N" (rather than trying to use negative numbers to label what is hierarchically above Tier 0).

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

this additional tier has been implemented cleverly so that it does not detract or get in the way of established data flows and protection protocols nor does it interfere with the usual I/O behavior and performance of other tiers and other applications.

Primary Use Cases for VFCache

As a Cache Accelerator

The primary use case for VFCache is that of a reading performance accelerator. VFCache is located much closer to the processor cores than an external array, thus significantly shortening the time needed to send and receive data from the processor. Even when targeting the same tier of media (think "SSDs", for this example), VFCache is faster because it does not need to go across the server's bus and through two switches (at least) before getting to the storage on the backend array.

Additionally, VFCache behaves as a *write-through* cache to EMC's *VMAX and VNX* storage arrays. Thus, VFCache satisfies the seemingly insatiable need for more and more data access speed to keep up with expanding applications, more users, and more performant processors with multi-core technology, and does so unobstrusively.

The big idea here is to capture changed data inside the application server before it is written out to a back end store. If that changed data is needed by the application right away, it can be transferred to the core at near memory speeds rather than at external I/O speeds. The choice of these top-of-line EMC arrays is not accidental. These have storage-associated software that enhances data persistence and protection. They are in service at some of the world's most demanding customers running the most challenging applications.

Not as a Replacement for Rotating Disks

If VFCache is so fast, then why not avoid spinning disks entirely? That is a fair question. However, EMC understands that today's realities dictate against this approach for three important reasons:

1. the cost of flash memory (SSDs),
2. the need for data permanency, and
3. the need for data protection, including remote copies.

(Each is discussed below.) Thus, VFCache is an accelerator of array-based storage. It turbo-chargers what is farther away, as illustrated in Exhibit 1 on the previous page. Also, note that VFCache can service multiple external arrays.

(1) Cost of Flash Memory (SSDs)

Yes, the price for flash-based disks is indeed coming down. However, it is still a technology that is rare enough to command a premium price, and thus is not suitable for all of your data, especially your parked data that is used infrequently.[3]

Rotating disk technology is far more mature, with generations of improvements in both technology and price performance. For now and the foreseeable future, all-flash (or all-memory) solutions will be much more expensive than some cache plus rotating storage. This is true whether you are talking about SSDs inside a server or SSDs inside an array, but the differences are worth noting.

When you add an SSD to a server or directly connect to it in an array, it's just another disk to be managed by storage administrators. It doesn't do anything automatically (i.e., dynamically changing to meeting current needs in real time).

When you add SSDs to an array, it also can be an invisible accelerator of the rotating disks in the array, i.e., it can serve an invisible array cache rather than visible, specified tier. Most applications can meet their higher performance requirements with a combination of spinning disks with, say, 5-to-10% SSD array cache. 100% SSDs would be overkill and expensive. Nonetheless, everything in the array still is much farther "down the pike" from the server cores, thus capping the benefit of the acceleration.[4]

(2) Need for Data Permanency

Data sitting in the VFCache is not permanent. Of course, the data has to be made permanent, lest it might not be available when demanded by the applications or users. Better stated, **data in VFCache is transient (like a buffer) and only is used to accelerate performance. It is not a *storage* solution; it is a *buffering* solution.** That is why VFCache relies on the external array to hold a permanent instance of that data. These arrays use techniques like RAID, replication, and backup technologies to protect what is stored within the array. A simple way to visualize this is "the array stores and protects while VFCache significantly accelerates the use of what has been stored in the array." Additionally, this use (of VFCache) fits in nicely (and transparently) with

[3] This is true for both types of flash: *MLC*, as well as *SLC* (multi-level versus single-level cell).

[4] Yes, it is faster than only using high-RPM disks, but the delays due to hops (in and out of switches) on the network and the distance limit the acceleration potential.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

disk-based storage that already has been procured.

(3) Need for Data Protection, including Remote Copies

Cache itself does not protect data against catastrophic loss of power that could result in data loss or at least data delay. Rotating disks and SSDs in arrays retain their magnetic storage characteristics, even after powering down. Better arrays offer software for high-availability use cases, wherein mirrors of primary data can be hundreds to thousands of miles away in secondary and tertiary locations. Should the primary location be compromised, there is always the mirrored data at the "other" site. Cache like VFCache does not have such remote (copy) protection characteristics.

How Does VFCache Work?

Why SLC over MLC in Today's VFCache?

Because VFCache sits in front of the external arrays, it sees a lot more traffic than the array would see. This is because it is holding copies of the most actively used data. Because what is most active can and often does change frequently, it has more writing to do in order to keep what is in highest demand within its flash memory. So, while an array might move data to its resting place just once, the transient data within VFCache may write the same data into its cache many times, whenever it becomes "hot". This amount of writing being done dictates the use of SLC flash over MLC, because SLC offers about a ten-fold endurance advantage over MLC. Both kinds of flash technologies erode with repeated use, the so called "wear-leveling" problem. As the vendor, EMC has to manage uneven wear leveling by paying close attention to where data is written, so as not to unevenly wear out the flash memory. The SLC's better endurance makes the most sense for VFCache. Neither EMC nor the data center wants to see the flash memory of VFCache being replaced due to wearing out.

What is the Overhead Associated with VFCache?

VFCache offers extremely powerful I/O inspection technology with little or no overhead. Theoretically, worst case overhead occurs when an application is characterized as 100% reads (a rare real world scenario), where the "price" to use VFCache is 5% of CPU overhead, plus at most only .5 GB to 1 GB of real memory spent for this purpose. This translates to 20 microseconds per I/O operation and almost is unnoticeable.

Why Flash Technology?

Cache is a much more reliable technology than disk. When was the last time you heard of a memory or cache failure in comparison to the failure of rotating disk? However, just in case there is a problem with a particular area of cache, VFCache includes the ability to fence off offending areas of the cache. If there is good data already located in that area, EMC will relocate data automatically without administrator intervention. This includes updating all of the necessary internal tables.

Using VFCache has yet another positive side effect. If target data is found in cache, then used by processors for the task at hand, and later written to permanent disk storage, fewer read requests are reaching the external array, thus lightening its load. This offloading of I/O frees the array to perform its remaining work more effectively, because the most demanding I/Os (think "hottest data") are being handled by VFCache and not reaching the external array. In theory, all workloads will benefit, not just the ones using the hottest data.

VFCache is Hardware, Software and Application Agnostic

So long as a server has a free PCIe slot, VFCache can be used. (These days, most servers support the PCIe bus architecture.) This makes the VFCache solution very general purpose and likely to fit into your IT environment, even if you have never used EMC hardware or software before, especially if your existing storage is not EMC branded.

EMC FAST auto-tiering software likewise is transparent to VFCache operations, because it is being done within the EMC array. Also, at the end user level, all applications and middleware are transparent to the presence and use of VFCache. This means no reprogramming on your part as well as easy introduction of the technology to your users.

Additionally, there are no operational differences when using VFCache in pure-play physical environments or when it is running on virtualized servers. This flexibility ensures long-term product life as well as product reuse and repurposing.

Accordingly, VFCache is easy to install and transparent to use, thereby resulting in a shorter time to value.

VFCache Packaging and Warranty

VFCache is sold as a packaged solution containing both hardware and software. It is

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

ordered using just one SKU (order number), making procurement, installation, and configuring very straightforward.

VFCache's warranty terms match the typical duty cycle of most commercial servers, which are replaced every three or four years not so much because they wear out, but because the next generation usually has very compelling technology enhancements and/or attractive cost reductions. Similarly, EMC offers a standard three-year warranty with an option to buy coverage for a fourth year.

Management of VFCache

The management console for VFCache gives users a full stack view of data in the system. From a single pane of glass, operators can diagnose issues, configure installations, and manage system-wide resources in real time. Now that Tier 0 data is in the equation, it is important that the management console open the door to gain an end-to-end view of the health and throughput of the system.

Additional Insights

Overall IT Performance

The question of IT application performance has long been characterized as an onion. Depending on the onion, a new technology (such as VFCache, a novel new scripting technique, or "the next whatever") may have a large or small effect. The external onion skin is almost transparent in its thinness but the first chunky layer of onion may be substantive. Nonetheless, we can state with certainty that there undoubtedly will be yet another performance barrier once the current one is mitigated. This is the unending challenge of trying to keep up with the next bottleneck, and there always is one. VFCache will be a performance-enhancing solution with major positive effects, but it will move the next biggest problem to the top of the list, no doubt.

VFCache as a Stressor (in a Good Way)

Thus, VFCache will stress your system – but in a good way. Besides matching performance to data availability, you may consider VFCache as an innovative system-level diagnostic tool. By installing VFCache in your server, determining the cause of your present (or additional) I/O performance issues may become obvious and quickly overcome. However and possibly more importantly, VFCache also may expose the source of delays in other *non-I/O* related areas, such as these.

A Special Use Case for VFCache

As explained elsewhere in the body of this report, EMC does not recommend using an all-flash solution as the ultimate resting place for all of your data. Rather, VFCache should be used primarily as a performance accelerator with data ultimately written to persistent back end storage. When used in this way, the user can take advantage of multi-location replication technology, such as *SRDF*, normal backup and restore, and archiving applications, and also enjoy the lower costs associated with spinning disks.

However, using a portion of VFCache for low-level work files, temporary files, and data that can be reconstructed from other repositories definitely is an option. Accessing such ephemeral data will accelerate due to the inherent throughput advantage of cache versus full read/write operations to a rotating disk. And, there is no worry about the data's safety because it can be resurrected, if this becomes necessary.

This is a win-win story for users. Mission critical data is passed through the accelerator portion of the card to tried and true enterprise-class storage (with all of its HA and value-added software features), while temporary data enjoys performance increases that can actually supercharge needy applications.

- **Overall system design** – Is it still accurate and relevant?
- **System orchestration** – Has it been optimized and is it working as advertised?
- **Task sequencing** – Are any tasks out of order? Are any tasks no longer necessary? Are all tasks optimized?
- **Script design** – Are scripts optimized? Is the organization's default scripting tool modern and efficient?
- **Script order** – Just as tasks may be out of sequence, the same might be true for scripts.
- **Correct and timely data feeds from other systems** – Legacy systems are usually "left alone" if they are working, but they may also contain invalid assumptions about data arrivals from other applications or systems.
- **Preprogrammed wait states or loops** – Lose them, if they no longer are necessary.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

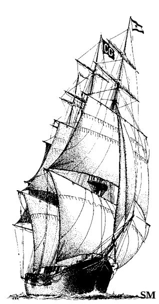
- **Identification of the need for new wait states or loops** – Insert them, if they now are necessary.
- **Operator errors and/or deficiencies** – These latter people-oriented issues may point to the need for more frequent or more rigorous operator training, new or revised training designs, or identification of missing key skills to be filled by internal transfers or new hires.

All of the above favorable by-products of stressing the system-wide performance should be considered "no additional charge" bonuses of installing VFCache. They likely are not insignificant and they most definitely are not uncommon.

Conclusion

Every organization has its "performance-starved" application. Often we hear, "if it only would go faster!" However, especially in today's market, there probably is no budget in time or money to rewrite that application. Instead, you continue to cope and/or compromise. Well, cope or compromise no more. Widespread deployment of VFCache offers a viable alternative that quite probably will fix not only your "troubled" application, but might speed many applications system-wide.

You owe it to your organization to perform a thorough evaluation of what VFCache can do for a relatively small investment in your server. In addition, this solution is from a company that is an industry leader with a proven track record. Unlike untested young companies, EMC has its reputation on the line. EMC would not be delivering VFCache if it wasn't ready for prime-time use on your most important applications. Check it out!



Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twentieth year, is an independent publishing and consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

***Jim Baker* is a Senior Contributing Analyst for The Clipper Group.** Mr. Baker specializes in storage hardware and software, the relationship between technology and its application in business, and emerging technological trends affecting users and vendors in the Information Technology community. He recently joined The Clipper Group after three decades in hardware and software product marketing and product management at companies as large as EMC and Honeywell Information Systems and as small as IPL Systems and Datapoint Corporation. He most recently served as Research Manager for Storage Software at IDC in Framingham, MA. Mr. Baker earned a Bachelor of Science in Marketing degree from the University of Illinois at Urbana-Champaign and an MBA, also from UIUC.

- ***Reach Jim Baker via e-mail at jim.baker@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.