



Symmetrix VMAX circa 2011 — A Continuing Drumbeat of Innovation

Analyst: Jim Baker

Management Summary

VMAX, the present generation of the EMC *Symmetrix* Storage Array, was enhanced yet again by a series of announcements drum-rolled in New York City on January 18, 2011. Even though it is the storage flagship of the company, VMAX still had to share the stage with a set of new offerings of Unified Storage, announcements from the Backup and Recovery Systems division, and executives reporting in on progress of customer companies on the journey to the private cloud. Dramatizing the theme of the conference (“Record Breakers”), they also shared the stage with a “hard to believe it unless you see it” demonstration of 26 people fitting inside a Mini Cooper, as observed by the audience and an adjudicator from the *Guinness Book of World Records*. You can find it on *YouTube*.

The flagship VMAX was well represented in the afterglow of the announcements. Nineteen of the forty-one announcements featured at the event were about Symmetrix VMAX. The *Engenuity* operating environment (that is, the software directing the storage) is now at level 5875 microcode. It offers previous Symmetrix generations a non-disruptive upgrade path so that they do not have to take an outage to migrate their data or jump to the next technology refresh when moving to a VMAX. Indeed, one of the new features detailed below called *FLM* (or *Federated Live Migration*) further addresses the need never to take down the storage system.

It was an across-the-board announcement touching virtually every aspect of VMAX, the hosts (including mainframe) that attach to it, and the environments in which it operates (both physical and virtual). However, two subtle trends were noted amongst all the demos, verbiage, and other hoopla.

1. Storage administrators are finding their lives far easier because of increased automation. Good news but it has an eye-opening corollary. These administrators now should strive to elevate their contributions (to, say, the IT policy level or specialize in yet-to-be-solved problems, such as security or multi-tenancy in the cloud), else they will be replaced by lesser-skilled and lesser-paid staff.
2. We are starting to see evidence that vendors are now working better together for the benefit of mutual customers. Interoperability is enhanced further by widespread vendor acceptance of new SCSI standards. Virtualization itself is a normalization factor, as multiple vendors are supported as VMs or as endpoints from different manufacturers. Customers’ IT infrastructures can use best-of-breed solutions with the expectation that they will operate properly and be supported by cooperating vendors.

Inside, there is a quick overview of the announcement and a drill down on its innovations. If all you seek are the details about various features, feeds and speeds, or pricing information, you would be better off discussing this subject with EMC directly. However, if you are interested in the features that kicked the VMAX up a notch and, more importantly, what they mean, please read on.

IN THIS ISSUE

➤ Highest of the Highlights	2
➤ Other Hardware Advancements	4
➤ Significant Software Advancements	5
➤ Customer Enablement and Ease of Use.....	6
➤ Conclusion	6

Highest of the Highlights

FAST VP (Fully Automated Storage Tiering for Virtual Pools)

In a phrase, *FAST VP* is automatic sub-LUN movement of data according to user-defined policies for promotion and demotion to other storage tiers (Enterprise Flash Drives (EFDs), Fibre Channel (FC) Drives, and SATA Drives). EMC has chosen three different tiers, each based on different drive technology. The architecture supports as many as 16 tiers but three appears to be a reasonable and manageable number. Historically, the first EMC implementation of FAST was at the LUN level, whereas today's FAST VP now works at the sub-LUN level.

Using FAST VP on demand and as needed, promotion and demotion of data is based on user-defined and site-specific policies. To accomplish this laudable goal, EMC uses the concept of dual windows for analysis. The first window is the data collection observation window that samples I/O patterns in time increments from 2 hours to 4 weeks. Relocation is performed at varying rates during a second window with the user specifying one of 10 rates to use. Skew factors determine where to place data dynamically once the user sets the policy and goals. Skews are classified as heavy, moderate, or low, based on hits of I/Os against the percentage of the data being held. Hottest and heaviest are candidates for Enterprise Flash Drives.

FAST VP is part of a *virtuous circle*. Consider the fact that hottest and heaviest data is promoted to EFDs, assuming there is space available, with the resulting very quick performance. However, the very act of moving that data up to flash has positive benefits for the lower tiers as well, since they are no longer thrashing to service the super hot READs and WRITES now being serviced quite nicely in flash. This means that the FC layer and the SATA layer also perform to their best ability because they are no longer trying to service data unrealistically. In fact, they too are being optimized, but for their (different) roles.

As you can imagine, there is a plethora of factors involved in the decision where to place data on which tier of storage. First, what should be the size of the increment to move? Rather than guess or use lab simulations, EMC collected traces of real customer data from 3500+ workloads. They observed a wide variety of environments – test, development, live production, sequential, random, concentrated I/O in several

industries, and with various applications. The result is that the optimal size of the moveable increment is 8 MB (bytes, not bits). Second, EMC has a long history of WRITE optimization based on locality of reference and various pre-fetch algorithms. These algorithms also inform the decision of how much to move and where to put it. Third, metadata capture is automatic and is used for deciding where to post I/Os. Remember that promoting data to EFD implies that there is room for the newer data. This means that no longer will hot data need to be demoted to some other tier in order to make room for newly hot data. (There certainly are many decision points to this process.)

EMC says the average and optimal segment for movement is 8 MB based on the analysis of 3500+ customer workloads, historical EMC expertise in pre-fetch algorithms, as well as metadata FAST VP is using. The IBM solution is quite different. First, IBM *EasyTier* has just two tiers: EFD and spinning disks. Next, IBM moves much bigger chunks – each 1 GB in size. This can lead to moving more than necessary and consuming valuable flash with less than important data, which is wasteful. Unimportant data can slip into the flash tier while important data can be stuck in the lower tier.

EMC claims that VMAX is the purpose-built engine for FAST and that the VMAX hardware was developed in response to the needs of FAST VP software and not the other way around. Predictably, EMC claims that competitors simply bolted their own FAST-like functionality onto their hardware as an afterthought.

FAST VP is policy driven. Decisions on where to place data are made in conjunction with user-developed goals, for example, optimization for high throughput versus optimized for capacity. To help users make these and other trade-offs, EMC offers *Tier Advisor*. Tier Advisor is a tool used in conjunction with the advice of EMC pre and post-sales experts. Also included in the tool is “what if modeling” that enables the user to see results before actually effecting the changes. Tier Advisor is free.

EMC was the first major vendor to use Enterprise Flash Drives and therefore has the most experience with it. They know, for example, that flash excels on READ operations rather than WRITE operations. In fact, writing to a flash drive and to a spinning FC channel drive is quite similar in duration. This means you do not want to waste flash on optimized WRITES but you do

want to use flash for optimized READs.

FAST VP is smart. It does not waste resources moving empty data, such as can be found in spreadsheets, and yet unpopulated DBMS structures.

Here is the key to success with FAST VP – you must align storage groups with applications that are in the system while keeping an eye out for scheduling best practices. For example, unless you really want to optimize backup at the expense of other applications, you would close the Movement Window to exclude BU from the analysis, lest the following unintended consequence occurs. For example, if you left the movement window open over the weekend and ran only backup, what would be in flash as of first thing Monday morning? The answer is “backup stuff” that would not be useful to the new workweek. Using EMC tools, users can avoid batch operations (such as backup) that could cause less desirable data to be promoted to flash.

Finally, *ControlCenter StorageScope Charge Back for FAST VP* offers the ability to charge back to the user community, according to which assets are consumed.

Federated Live Migration

Federated Live Migration (FLM) is a tool to use for technology refreshes because it allows applications and their hosts to stay live during the process of migrating to VMAX from DMX. There is no need to take an outage to move the data or to insert the virtualization layer into the infrastructure. This eliminates host and apps downtime, thereby keeping users and their families happy. Families are happy because employees no longer have to come in on weekends or holidays to perform the migration.

FLM has open systems support only and currently requires *PowerPath* multi-pathing software, as a prerequisite. Future plans call for *MPIO (MultiPath I/O)* and *DMP (Dynamic MultiPathing)* standards support, as an alternative to *PowerPath*. FLM itself is included in the price of VMAX although there would be a separate charge for multi-pathing software. Although FLM has no *z/OS* (mainframe) support, EMC already has another live migration tool for migration called *LDMF (Logical Data Migration Facility)* for this.

Here is how FLM works. FLM is a peer-to-peer LUN migration facility. All I/O operations for a given LUN are intercepted and pointed to the new location. The VMAX essentially stops

the alternate path for one I/O and forces a retry. Thereafter, all I/Os go to the new location. Key to making FLM work is that the same worldwide name is spread over the old and new ports. This migration is usually a one-way trip but FLM could reverse the process mid-stream, if necessary. Source platforms can be *DMX-1, DMX-2, DMX-3, DMX-4*, or another VMAX. The end result is that migrations now take days and not months. EMC estimates a 75% reduction in elapsed time for major migrations.

Virtualization, VMware, and Hyper-V Support

VMware VAAI stands for *vStorage APIs for Array Integration*. Support for VAAI subdivides the work of writing data to a virtual LUN into APIs that are optimal for the attached storage to perform, but each storage vendor must agree to integrate them for their own platforms. *Engenuity Microcode 5875* is the contemporary EMC software for this integration. In combination with *Virtual Storage Integrator (VSI) for vCenter, vShield*, and *vSphere*, EMC’s VMware solution is industry leading. It certainly helps to be part of the VCE coalition where designs and schedules are coordinated, access to gurus is immediate, and joint planning is always underway. As a proof point, EMC quoted an example of 5,000,000 VMs (Virtual Machines) on one VMAX, an impressive figure.

EMC’s support for Microsoft’s *Hyper-V* follows a similar model. VSI features have been extended into Microsoft Service Console (or MSC) plug-ins, with the result being a VMAX that can support either a VMware or a Hyper-V environment or both simultaneously.

Data@Rest Encryption

This feature answers the call put out by all concerned that their data can be compromised while it is “at rest” sitting on some kind of magnetic media within the VMAX. Regulatory and privacy concerns have prompted the development of *D@RE*. There is also a cloud play at work here, as storage-as-a-service customers do not want their data intermixed with that of other users sharing cloud storage and feel more comfortable if, somehow, it is encrypted so that outsiders cannot see it. This contribution from EMC’s RSA Division guarantees that data is encrypted by employing a new purpose-built hardware engine, which means it is lightning fast and does not affect the speed of reading or writing data on the VMAX. All data is encrypted regardless of the drive technology, including

EFD, FC drives, and low-cost SATA drives. Key management is handled via RSA with protections against misuse by all personnel whether from the user or from EMC (say, on a trouble call). RSA guarantees that keys are not published in any form, thereby increasing the security of data at the media (or drive) level. The encryption algorithm is unique to each array, each drive, and each slot in which the drive is located. Before this feature, those who were extremely paranoid about the data and its residual ones and zeros left on the magnetic strata of drives went to extreme means. They would contract to have the drives demagnetized and or shred when they took them out of service. With D@RE and no knowledge of the location of keys nor the encryption algorithm used, residual data is meaningless gobbledygook.

Other Hardware Advancements

Native 10 Gigabit Ethernet Connectivity

This feature refers to the front (or processor facing) end of the array and is supported in iSCSI environments. EMC claims it is eight times faster than traditional Fibre Channel connectivity running at 1 Gigabit per second speeds. This higher speed is also supported by EMC's industry-leading data replication, disaster recovery product, *SRDF (Symmetrix Remote Data Facility)*, resulting in faster connectivity to the remote location.

Enginuity Performance Optimization

Although Enginuity easily could be considered software, EMC positions it as the hardware's microcode and discusses it as part of the hardware. Every release of microcode attempts to "up" the performance of the array. This release is no exception as it will be necessary for customer installs that are getting so large (now growing to as large as 2 PB). One of the techniques used in this release is to increase the speed of large-block read-miss transfers (by factor of 2.5 times over the previous microcode generation). In Symmetrix arrays, a "read miss" means that the required data was not in cache (a.k.a. *global memory*), thereby forcing the system to reach down and access the spinning disk or flash drive to retrieve it. Such misses incur a performance penalty, so the system attempts to avoid them in the first place and to make a quick transfer, if they do indeed happen.

Virtual LUN v3

Because this is the third generation of this feature, many users have had experience with it

in a variety of applications and situations. Its purpose is to relocate a virtual LUN in one pool to another pool. Normally, FAST VP will move data as necessary in compliance with policies, but some customers wanted to have another lever to pull, should there be a performance, workload, or even an organizational reason to make the change.

Reducing System Locks During Upgrades

Enginuity 5875 also now greatly reduces system locks when upgrading the Symmetrix VMAX. Previously, customers would be prohibited by the array from making configuration changes (including provisioning storage) for up to a week before new drives or engines could be added to the array. With Enginuity 5875, this configuration change lock is now required only for a short time (less than an hour) to capture the current configuration prior to the upgrade, and then again for a brief period while the new drives or engines are added. This addresses a key customer concern by allowing administrators to manage and allocate storage more flexibly, even when an upgrade is in process. This feature removes an unnecessary constraint on the system, thus making the scheduling of changes far easier for the organization, in general, and the storage administrator, in particular.

Thick to Thin In-Line Zero Space Reclaim

One of the advantages of EMC's virtual provisioning is its space efficiency. Only true data is captured initially, while space is allocated just in time, as it is needed as the data grows. If the user is migrating from a traditional or "thick-provisioned" LUN, this feature reclaims space previously allocated as "head room" or growth space during the original provisioning process. After all, why unnecessarily consume the space that is not yet being used? Simultaneously, there is an efficiency argument for less but denser storage resulting in greener storage, taking less power to operate and cool, and less floor space.

T10 UNMAP & WRITE SAME Standard Support

EMC's adoption of these two SCSI standards from the T10 working committee is quite beneficial. The T10 Technical Committee is responsible for defining SCSI interface standards. In this case, the new commands are for space recovery of scarce disk space and for speeding the performance of arrays writing repetitive data (such as empty spreadsheet cells or empty rows and columns in a database). The

SCSI WRITE SAME function allows the same character (usually zero, but not necessarily so) to be written to the disk as an array function thereby not consuming processor cycles or channel bandwidth between the host and its storage. The SCSI UNMAP function allows previously consumed space on the disk to be unmapped and, therefore, recovered for re-use by the next application. This is an important feature for virtual LUNs because it is not known where in the pool data will eventually land.

EMC adopts standards when they foresee practical benefit for their products or their customers. Sometimes they blaze the trail by offering up their own solutions for consideration by the standards organization. Leaders do that and EMC is obviously a storage leader. Meanwhile, they are not wasting time awaiting formal and often lengthy adoption of a feature that can be put to immediate use by their customers.

Significant Software Advancements

TimeFinder Copy Quality of Service

Some copies are more important than others, so why slow down the system for an unimportant copy? The challenge is how to communicate which copy task has higher or lower priority than another. With this feature, EMC provides the user with another lever to pull if wanted. Sixteen priority levels can be assigned, allowing the user to choose either fine or gross control. Higher numbered tasks trump lower numbered ones. It is a creative approach.

TimeFinder Duplicate Snap

Snaps are space efficient replicas vis-à-vis *TimeFinder BCVs (or Business Continuanance Volumes)* because they are pointer-based and capture and consume only changed data. With this feature, the user may want another snap to run another simulation or the next test cycle. What other uses are there for duplicate snaps is a question for the user community.

Concurrent SRDF/A Sessions

Symmetrix Remote Data Facility/Asynchronous is different than traditional SRDF, which is synchronous in character. This solution is preferred over longer distances (think hundreds of miles or more) to minimize the latency associated with the speed of light and the time it takes to go long distances. EMC had previously offered the concurrent combination of one synchronous leg and one asynchronous leg off the same source volume. This new feature allows

two asynchronous legs from a single source volume. Every SRDF configuration is different but an example of how to use this feature would be a Florida-based company with offices in Nashville, TN, and London, England. The Florida-to-Tennessee link would protect the firm from a regional disaster, such as a hurricane, while the longer link could provide almost-instant access to data for London workers and also could serve a truly remote backup location.

Vault-to-Flash

Vaulting is an EMC data protection technique used on Symmetrix family arrays, including VMAX. If there is a power failure that affects the array, it is necessary to save user data and active metadata that is in global memory to a safe place for recovery, once power is restored. As a “cache first” array, it is highly likely that the application will have received confirmation from the VMAX that all writes have been written to the array even if they are still in cache waiting to be de-staged to spinning or flash disk.

With the latest release of microcode (version 5875), the vault feature is made available in an all Enterprise Flash Drive VMAX system. In this case, such user and metadata must be placed into an EFD-only vault. Previously, in an all Fibre Channel disk configuration, the vaulting would be spread over several spinning drives in order to accomplish the task in a minimum amount of time using the parallelism of multiple heads and platters. In a mixed EFD and Fibre Channel drive storage configuration, the EFD data would have landed on multiple save areas across several spinning drives.

The purpose of this feature is the quick de-staging of data and saving battery power during an electrical crisis. The new all-EFD vault process also spreads out the saved data to multiple drives to gain the speeds afforded by parallel write activities. Looking forward, we expect that certain hot applications that need terrific performance but not a whole lot of capacity will use a VMAX configuration with 100% EFD storage. And, as the price curve drops for EFDs versus spinning disks, this scenario is much more likely. EMC is ready now for customers fitting this profile.

Automated Port Settings

In an ease-of-use addition to 5875 microcode, EMC provides automated port settings. By using a list of most common hosts during the set up, the provisioning process can be more automatic. If a host is not on the list, then the setting

can be manually inserted by the storage administrator. Meanwhile, much time will be saved because of all of the settings that he or she does not have to configure. Less human interaction means fewer opportunities to make careless errors. This feature further exemplifies one of the underlying themes that permeate this release: *Take away manual tasks so that there is less for storage administrators to do.* Fewer errors. Fewer downside consequences. Faster time to operational status. All good things.

Customer Enablement and Ease of Use

Customer Self-Service Drive Replacements

Sophisticated enterprises and those located in remote locations often want to replace disk drives themselves and avoid the wait times associated with placing a call to EMC or one of their servicing partners and then standing by for their arrival onsite. Because drive replacement is now so easy, users will be able to replace drives at the time of their choosing and convenience. A byproduct of this feature is the ever-increasing self-sufficiency of the EMC customer, a good step forward for both vendor and user.

Customer Self-Service Health Check

This feature builds and maintains user confidence in the system. It is especially important, given the kind of environments that are supported by VMAX (the world's largest commercial businesses, central governments, large universities, complex engineering and medical institutions, etc.) With this tool, IT managers can verify the operational status and overall health of their storage subsystem.

E-Licensing Management

Electronic licensing management is a win-win for the customer and for the vendor. On the vendor side, license renewals are quicker and easier, as is tracking of which customer has installed which modules. Usage patterns can be tracked along with other interesting analytics for both support and marketing organizations. On the enterprise side, the IT department can manage their own infrastructures better, eliminating duplicate payments, controlling which users have which modules, documenting compliance efforts, and returning unneeded or unwanted copies of chargeable software. EMC offers a reporting feature for visibility and easier decision-making.

Conclusion

EMC has invested wisely in this extensive set of new features for the VMAX. The list of enhancements in 5875 microcode is even longer than described herein. In short, there are enhancements for every class of customer so that each can see “what’s in it for me?” and find something meaningful. VMAX now defines the state-of-the-art for 2011. Take the time to see what the new VMAX can do for your enterprise.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent 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.