

NEW FEATURES IN EMC ENGINUITY 5875 FOR OPEN SYSTEMS ENVIRONMENTS

Abstract

This white paper introduces new features made available by Enginuity™ 5875 on EMC® Symmetrix VMAX™ storage arrays. Sub-LUN tiering concepts are especially noteworthy in this Enginuity release, although many other important features are also included. This white paper completes a picture of the intelligence and functionality of Enginuity 5875, which is designed to satisfy the demands of modern IT infrastructures.

January 2011

Copyright © 2011 EMC Corporation. All Rights Reserved.

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

The information in this publication is provided “as is”. EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware, ESX, and vSphere are registered trademarks of VMware, Inc. All other trademarks used herein are the property of their respective owners.

Part Number h8092.1

Table of Contents

Executive summary	4
More efficiency	5
FAST VP	5
FAST VP algorithms.....	8
Data movement.....	8
Virtual LUN VP Mobility.....	8
Federated Live Migration	11
More scalability	12
Bandwidth improvements	12
10 Gb/s Ethernet support.....	12
Dynamic back-end configuration changes	12
Security and interoperability	13
Symmetrix Data at Rest Encryption	13
vStorage APIs for Array Integration.....	13
Improved application compatibility	14
Virtual Provisioning	14
Duplicate TimeFinder/Snaps	15
Concurrent SRDF/A.....	15
SRDF general enhancements	15
Space reclamation with SRDF	15
Zero space reclamation	16
Copy QoS	16
Conclusion	16

Executive summary

Enginuity™ 5875 is the latest intelligent, multitasking, preemptive storage operating environment (SOE) released for EMC® Symmetrix VMAX™ systems. As with previous Enginuity versions this release is devoted to storage operations and optimized for service levels required in high-end environments. As expected, this Enginuity version on EMC Symmetrix® VMAX systems further advances the ability of EMC self-optimizing intelligence to deliver performance, array tiering, availability, and data integrity that now define advanced storage functionality.

The EMC Symmetrix VMAX series with Enginuity incorporates a scalable fabric interconnect design that allows the storage array to seamlessly grow from an entry-level configuration to a 2 PB system. Symmetrix VMAX systems provide predictable, self-optimizing performance and enable organizations to scale out on demand in private cloud environments. It automates storage operations to exceed business requirements in virtualized environments, with management tools that integrate with virtualized servers and reduce administration time in private cloud infrastructures.

Symmetrix VMAX users are able to achieve “always on” availability with maximum security, fully nondisruptive operations and multisite migration, recovery, and restart to prevent application downtime.

Enginuity 5875 for Symmetrix VMAX systems extends customer benefits in the following areas:

- **More efficiency:** Zero-downtime tech refreshes with Federated Live Migration, and lower costs with automated tiering
- **More scalability:** Up to 2x increased system bandwidth, with the ability to manage up to 10x more capacity per storage admin
- **More security:** Built-in Data at Rest Encryption
- **Improved application compatibility:** Increased value for virtual server and mainframe environments, including improved performance and faster provisioning for z/OS servers

Enginuity 5875 carries the extended and systematic feature development forward from previous Symmetrix generations. This means all of the reliability, availability, and serviceability features; interoperability and host operating systems coverage; and application software capabilities developed by EMC and its partners continue to perform productively and seamlessly even as underlying technology is refreshed.

This white paper describes the additional feature enhancements relevant to open systems that are made available with the Enginuity 5875 operating environment on Symmetrix VMAX systems:

- Fully Automated Storage Tiering with Virtual Pools (FAST VP)
- Virtual LUN VP Mobility

- Federated Live Migration
- Bandwidth improvements
- 10 Gb/s support for Ethernet, SRDF TCP/IP, and iSCSI
- Additions to dynamic configuration changes
- Symmetrix Data at Rest Encryption
- vStorage APIs for Array Integration
- Virtual Provisioning enhancements
- Concurrent SRDF/A
- Thick-to-thin migration with SRDF®
- SRDF general enhancements
- Duplicate TimeFinder®/Snaps
- Copy QoS improvements

More efficiency

FAST VP

FAST enables customers to realize the benefits of tiered storage with less time and effort spent on performance tuning, management, and monitoring. FAST VP extends these benefits, optimizing cost and performance, by placing the right thin data extents, on the right tier, at the right time. The FAST VP system enables a storage administrator to decide how much SATA/Fibre Channel/Flash capacity is given to a particular application and then automatically place the appropriate busiest thin data extents on the desired performance tier and the least busy thin data extents on a capacity tier. The administrator's input criteria are assembled into FAST policies. The FAST VP system uses policy information to perform extent data movement operations within two or three drive tiers in the Symmetrix VMAX array. Because the unit of analysis and movement is measured in *thin* extents, this sub-LUN optimization is extremely powerful and efficient. FAST VP made available in 5875 is an evolution of the existing FAST and EMC Optimizer technology.

There are two components of FAST VP: the FAST controller and the Enginuity 5875 microcode. The microcode is responsible for collection of performance statistics, at both the LUN and sub-LUN level. The FAST controller is responsible for analyzing performance data collected by the microcode.

The resulting data analysis generates a data movement policy that contains promotion and demotion thresholds for each tier included in a FAST policy.

The microcode applies this movement policy to all thin devices under FAST VP control to determine the appropriate tier for the data, and will generate and execute movement requests to relocate thin extents to the appropriate tier.

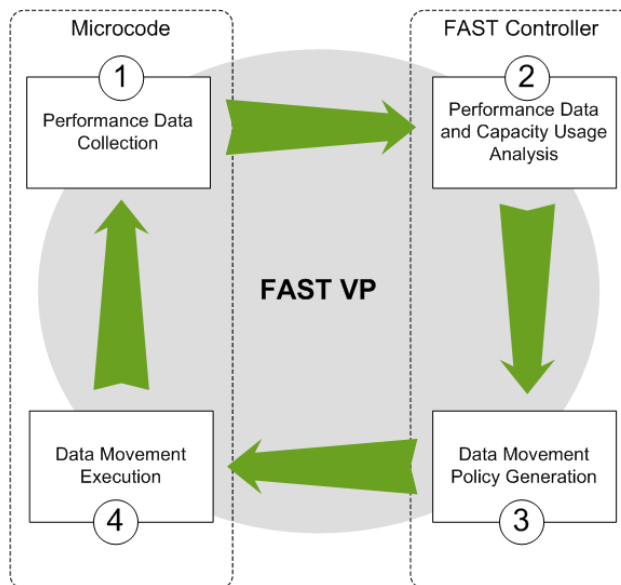


Figure 1. FAST VP components

FAST VP requires three control objects —storage groups, FAST policies, and storage tiers.

- **Storage groups** are a logical collection of Symmetrix volumes that are to be managed together, typically associated with an application.
- **FAST policies** contain a set of tier usage rules that can be applied to one or more storage groups.
- **Storage tiers** contain one to four virtual pools of a matching drive technology (Enterprise Flash Drive, FC, or SATA) and a RAID protection type.

The storage group definitions are shared between FAST and Auto-provisioning Groups. However, a Symmetrix device may only belong to one storage group that is under FAST control.

A FAST VP policy groups one to three tiers, and assigns an upper usage limit for each tier. The upper limit specifies how much allocated capacity of the thin devices in an associated storage group can reside on that particular tier.

The upper capacity usage limit for each tier is specified as a percentage of the allocated capacity of the thin devices in the associated storage group. The usage limit for each tier must be between 1 percent and 100 percent. When combined, the upper usage limit for all Symmetrix tiers in the policy must total at least 100 percent, but may be greater than 100 percent up to 300 percent.

A tier will contain at least one thin storage pool from the Symmetrix but can include up to four. If more than one thin pool is contained in a tier, the thin pools must be of the same drive technology type and RAID protection.

Figure 2 shows two storage groups, Thin_ProdApp1 and Thin_Development. Each storage group is associated with one policy, Platinum and Bronze, respectively. These policies associate the storage groups with up to three tiers.

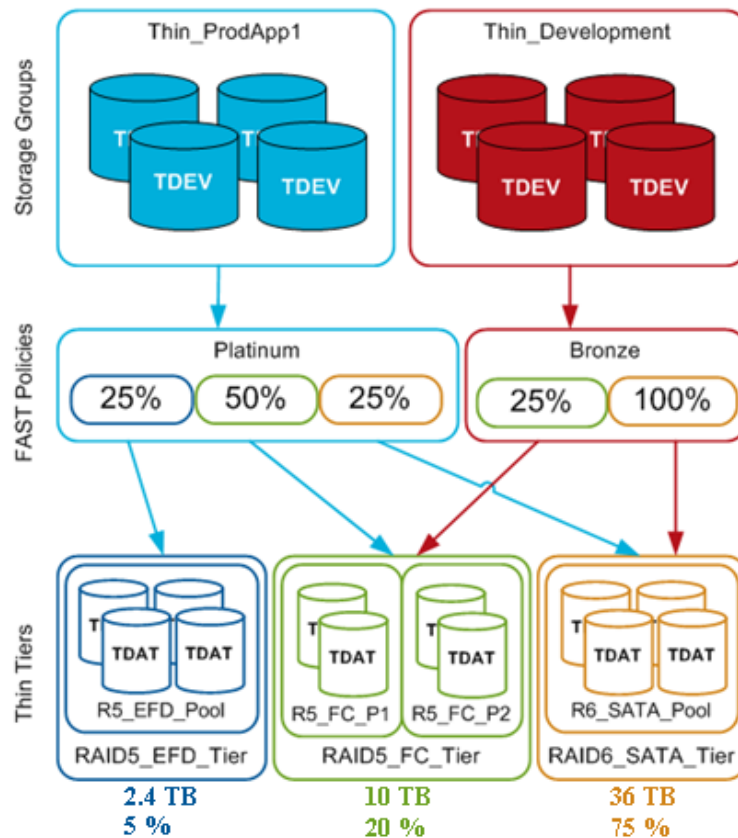


Figure 2. FAST policy association

Based on the Platinum policy, FAST VP will place 25 percent of the allocated capacity of the Symmetrix thin devices in the Thin_ProdApp1 storage group, in one or more thin pools configured as RAID 5 (3+1) on EFD, and 50 percent on RAID 5 (7+1) thin pools on FC, with the remaining 25 percent in RAID 6 (14+2) thin pools on SATA drives.

In the Thin_Development storage group, all of the allocated capacity of the thin devices can exist in RAID 6 (14+2) thin pools on SATA drives. However, depending on performance needs and utilization, up to 25 percent of the allocated thin device capacity may be relocated by FAST to RAID 5 (7+1) thin pools on FC.

Performance time windows can be defined to specify when the FAST VP controller should collect performance data, upon which analysis is performed to determine the appropriate tier for devices. Also, defined data movement windows will determine when to execute the data movements necessary to move data between tiers.

FAST VP has two modes of operation, Automatic or Off. When operating in Automatic mode, data analysis and data movements will occur continuously during the defined data movement windows. In Off mode, performance statistics will continue to be collected, but no data analysis or data movements will take place.

FAST VP algorithms

FAST VP uses two distinct algorithms, which act in tandem and counter-balance one another, when determining the appropriate tier on which data should reside. The algorithms are described below:

- The **intelligent tiering algorithm** uses sub-LUN-level performance metrics to determine the appropriate tier for the data on each VP volume under FAST VP control. This is done by calculating a “prioritized access score,” based on long-term and short-term activity rates. Data with higher scores will be promoted to higher performing tiers, while lower scores will be relocated to more cost-effective tiers.

The result of the intelligent tiering algorithm is a set of promotion and demotion thresholds for each tier, defining recommendations for data placement.

- The **allocation compliance algorithm** detects when the allocated capacity of a storage group exceeds the capacity allowed in a single tier, based on the usage limit specified within the FAST VP policy.

The result of the allocation compliance algorithm is a set of specific data movements that will bring the allocated capacity storage group back within the boundaries specified by the associated policy.

Data movement

Each FAST VP extent on a VP volume, for which sub-LUN-level statistics are collected, is divided into FAST VP sub-extents of contiguous space. Activity bitmaps are maintained at the sub-extent level to precisely identify data contributing to the performance metrics.

When promoting data to a higher tier, only active sub-extents will be relocated. However, when a FAST VP extent is considered to be idle enough to be demoted to a lower tier, the entire FAST VP extent will be relocated. The result is both focused efficient promotion of performance data and large efficient demotion of idle data.

Data movements take place by moving the allocated tracks groups of each FAST VP sub-extent from one thin pool to another. When the relocation completes, the track groups from the original thin pool are deallocated.

Virtual LUN VP Mobility

Symmetrix Virtual LUN technology enables the seamless movement of volumes within a Symmetrix without disrupting the hosts, application, or replication sessions. Prior versions permitted the relocation of fully provisioned (fat) FBA and CKD devices across drive types (capacity or rotational speed) and RAID protection types. VLUN VP now provides “thin-to-thin” mobility, enabling users to meet tiered storage requirements by migrating thin FBA LUNs between virtual pools in the same array (CKD devices are not yet supported as virtually provisioned volumes). Virtual LUN VP Mobility gives administrators the option to “re-tier” a thin volume or set of thin volumes by moving them between thin pools in a given FAST configuration. This

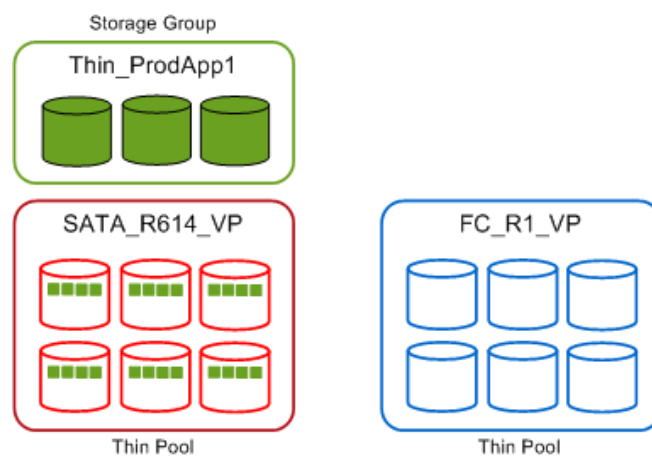
manual “override” option helps FAST users respond rapidly to changing performance requirements or unexpected events.

Virtual LUN VP migrations are session-based – each session may contain multiple devices to be migrated at the same time. There may also be multiple concurrent migration sessions. At the time of submission a migration session name is specified. This session name is subsequently used for monitoring and managing the migration.

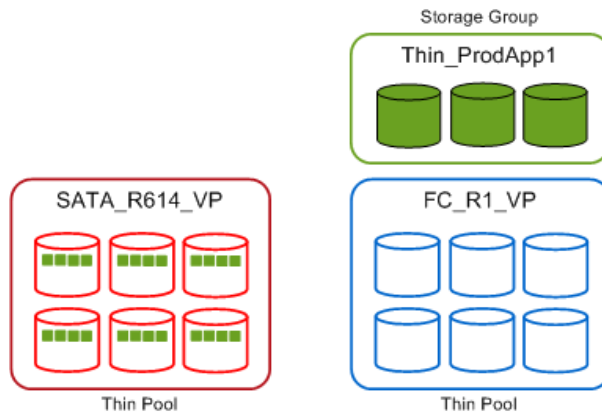
While an entire thin device will be specified for migration, only thin device extents that are allocated will be relocated. Thin device extents that have been allocated, but not written to (for example, pre-allocated tracks), will be relocated but will not cause any actual data to be copied.

New extent allocations that occur as a result of a host write to the thin device during the migration will be satisfied from the migration target pool.

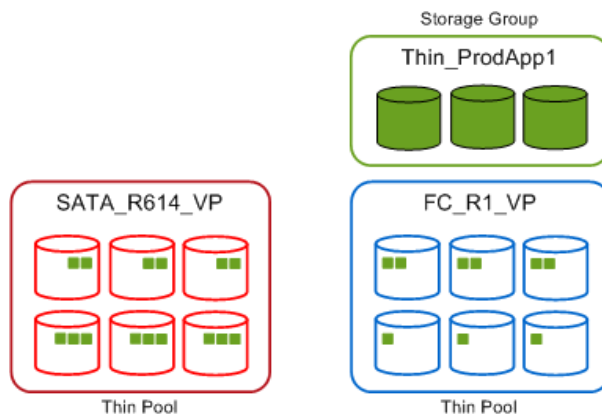
The following process occurs when migrating a group of thin devices from one thin pool to another. In this example, all the devices in a storage group will be migrated from one thin pool, configured as RAID 6 (14+2) on SATA drives, to a target thin pool, configured as RAID 1 on FC drives.



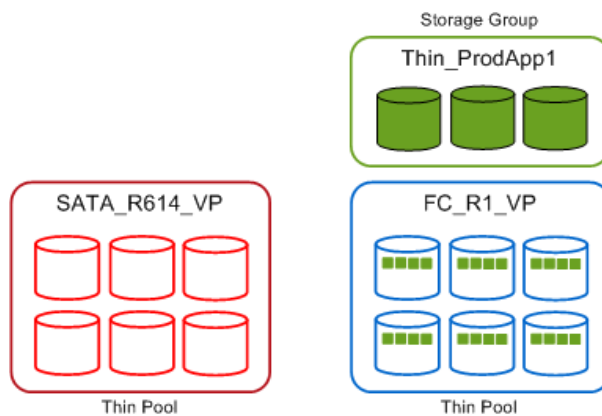
1. Storage group Thin_ProdApp1 is shown with all devices bound to pool SATA_R614_VP.



- When migrated to another pool, all devices in the storage group will be rebound to the target pool, FC_R1_VP.



- Once the devices have been rebound, all allocated tracks will be moved from the source pool to the target pool. As tracks are relocated, they are deallocated from the source pool.



- When all allocated tracks have been relocated, the migration is complete.

Federated Live Migration

Engenuity 5875 introduces a feature known as Federated Live Migration (FLM), which allows data movement from an older Symmetrix system to a Symmetrix VMAX system running Engenuity 5875 without downtime to applications and without loading software on any connected hosts. No host interruption is required to load virtualization software. This feature makes use of Open Replicator to move the data between the Symmetrix arrays and EMC PowerPath® (or other multipathing solutions) to manage host access to the arrays while the migration is taking place. Federated Live Migration enables a device in the Symmetrix VMAX array to “impersonate” a device in the old Symmetrix array – making it assume the complete identity and geometry of the old device – and then performing an ORS hot pull, donor_update operation. Impersonation can be maintained indefinitely or removed at a time convenient for the host.

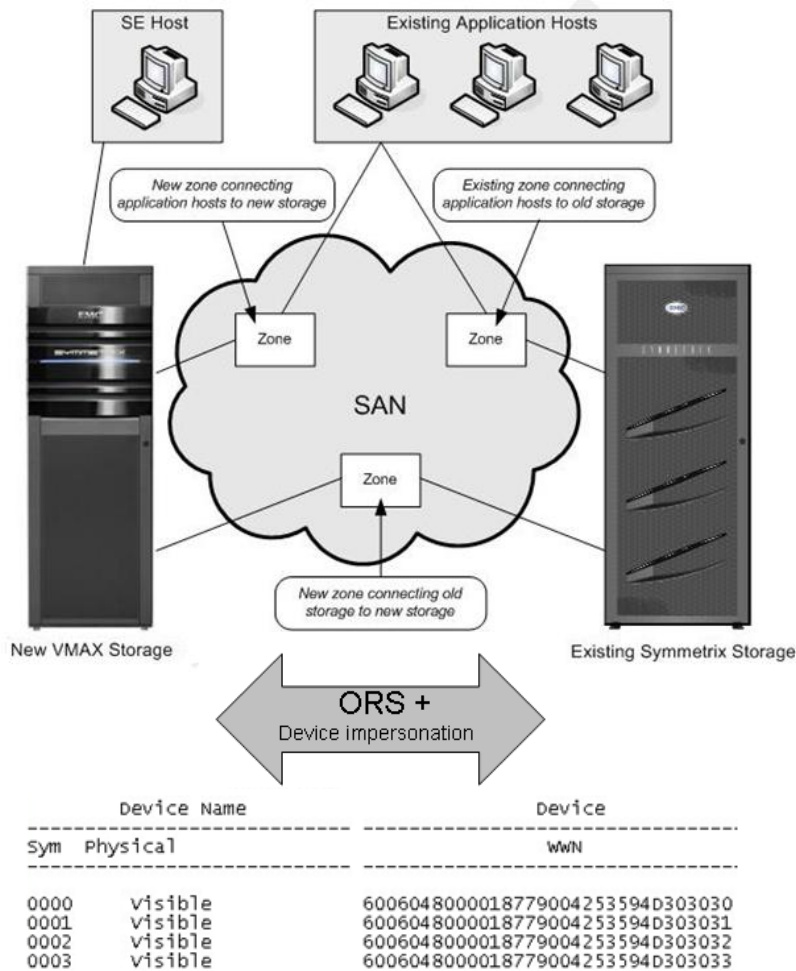


Figure 3. A Federated Live Migration environment

The original device must be in a Symmetrix and cannot be involved in any type of local or remote replication. (The ORS pull session will not be able to copy consistent data if new data is written to the remote device while the session is running.) The SYMCLI interface for Federated Live Migration handles all of the data movement.

However it does not automatically set up the required SAN zones. This includes zones from the application hosts to the new storage array and from the old storage array to the new storage array. When the migration is complete, removing old zones is also a task outside the SYMCLI interface. At a convenient future maintenance window after the migration, time can be scheduled to perform a rescan (or reboot hosts on some operating systems) and undo the “impersonation” of the ORS migration target devices that allowed the migration to occur without installation of a host virtualization layer. If desired, device impersonation can continue indefinitely.

More scalability

Bandwidth improvements

With Engenuity 5875, Symmetrix VMAX arrays allow selected data to be “fast tracked” in the Virtual Matrix™, enabling front-end and back-end I/O to bypass buffer memory. With this intelligent optimization in place the Symmetrix VMAX at Engenuity 5875 benefits with up to two times more I/O bandwidth for large block reads.

10 Gb/s Ethernet support

New front-end I/O modules supported in Engenuity 5875 represent the latest 10 Gb/s Ethernet. 10 Gb/s Ethernet offers improved single-stream performance for iSCSI-connected hosts and increases SRDF bandwidth. In summary the 10 Gb/s modules apply to:

- SRDF TCP/IP
- iSCSI

Dynamic back-end configuration changes

Ideal configuration changes should cause absolutely no interruption to array processing, should be performed in parallel with other configuration changes, and should be finished in a short time. Previous Engenuity versions on Symmetrix VMAX systems offered configuration changes that approached this ideal. Front-end mapping and masking, metavolume creation, and setting device attributes were defined as dynamic changes that used a new mechanism with no interrupt to the Symmetrix array. Up to four operations could be performed simultaneously and operations were completed in less than 30 seconds.

With Engenuity 5875 many back-end configuration changes have been planned to take advantage of the dynamic methodology. Initial additions to dynamic capability are listed below, with more functionality scheduled for future releases.

As with other dynamic changes the following back-end operations cause no interrupt, can participate in parallel operations, and are fast to complete:

- Adding and removing drives
- Permanent sparing operations

Security and interoperability

Symmetrix Data at Rest Encryption

Symmetrix Data at Rest Encryption with Engenuity 5875 utilizes Data Encryption Keys to encrypt/protect data on drives within a Symmetrix VMAX storage array. Data at rest encryption eliminates security risks when drives are removed from an array because of normal drive replacement or media theft and when arrays are repurposed. Compliance to industry encryption requirements can also be satisfied with this solution. Symmetrix Data at Rest Encryption is a “set-and-forget” feature for the entire Symmetrix VMAX array that is enabled at array installation.

Encryption keys are managed from a key server and secure repository that reside on the Symmetrix VMAX service processor and in this implementation no key management is required of the end user. Multiple copies of the encryption keys are kept on the VMAX for redundancy and recoverability, and all key copies are themselves securely encrypted. There is a different encryption key for each drive in the Symmetrix VMAX, and the array does *not* require specialized encryption-enabled drives. All drive types are supported for Symmetrix Data at Rest Encryption and hence Fully Automated Storage Tiering technology is available for Enterprise Flash Drives, and Fibre Channel and SATA drives in use with encryption.

The encryption process occurs on the back end of the array in the I/O module. The encryption chip has been designed to be just as fast as existing non-encrypting back-end I/O modules. This specialized hardware ensures that there is *no* negative effect on performance associated with the Symmetrix VMAX encryption solution. Encryption on the disk is the 256-bit AES-XTS block cipher standard. Encrypted drives are always unreadable unless plugged in to the correct Symmetrix VMAX array. All data formats are accepted for encryption on the Symmetrix VMAX. Enterprise arrays consolidating open system (FBA), mainframe (CKD), and iSeries data participate fully in encryption benefits.

vStorage APIs for Array Integration

This functionality supports VMware’s vStorage APIs for Array Integration (VAAI) initiative. Supporting this initiative improves performance at the VMware® ESX® Server level. The first enhancement is Hardware Assisted Locking. This new function locks at the block level and not the LUN, with an efficient SCSI compare and swap command. Block-level locks have no effect on other virtual machines. The second enhancement is the Block Zero command where multiple SCSI writes are avoided by using an efficient command to write the same data for a range of blocks. The last improvement is Full Copy. This SCSI extended copy command avoids read write operations to the host and identifies data for the array to copy within itself. These API enhancements are particularly helpful in accelerating clones of virtual machines initiated by VMware vSphere™.

Improved application compatibility

Virtual Provisioning

Symmetrix Virtual Provisioning™ enhancements continue to improve ease of use and create a similar user experience as standard provisioning. Engenuity 5875 Virtual Provisioning enhancements include:

- The ability to rename thin pools.
- Virtual Provisioning rebalancing was released with a pool variance measure of ± 5 percent. Upon request this variation value was made user-definable from 1 percent to 50 percent. The default value is ± 1 percent variation, which delivers a very balanced result but consumes more array resources. The new range allows user choice for a higher skew between pool members, reducing the overall amount of rebalancing but also reducing resource consumption. The opportunity exists for rebalancing to be run with successive reductions in the variance value to manage the final variance result with resource consumption spread over a longer timeframe.
- Another user-requested parameter for rebalancing Virtual Provisioning pools was the maximum number of concurrent devices that participate in the rebalance. In Engenuity 5875 this can be set anywhere from two devices to the entire pool.
- Users are now able to reclaim “no longer used” space in virtual pools for file systems that comply with the T10 space reclamation standard. This standard enables host-based file systems to notify storage arrays that certain tracks can be safely deleted. The T10 SBC-3 committee has finalized standards for two new SCSI commands for thin devices. The UNMAP command advises a target device that a range of blocks are no longer needed. If the range covers a full Virtual Provisioning extent Engenuity 5875 can return that extent to the pool. If the UNMAP command range covers only some tracks in an extent, those tracks are marked Never Written by Host (NWBH). The extent is not returned to the pool but those tracks are not read from disk to return all zeros; don't have to be copied for snap, clones, or rebuilds; and will not take bandwidth with SRDF transfers.
- The Block Zero command instructs a Symmetrix VMAX to write the same block of data to a specified number of sequential logical blocks, greatly accelerating the provisioning process. Operating systems like VMware's will use this command to zero (write zeros) to a Symmetrix LUN. Without this command the VMware host would need to use multiple write requests to the LUN. But with Block Zero support in Engenuity 5875, a host can do the same format using just one instruction.
- Virtually provisioned devices can now participate in Generic Save Write and Oracle Check Sum.
- Virtually provisioned thin striped metadevices can now be expanded. A “thick” BCV is established to preserve data during the expansion. The reconfigured thin striped meta will remain thin, not become fully allocated.
- Thin cascaded clones are now supported.

Duplicate TimeFinder/Snaps

Duplicate TimeFinder/Snap offers the capability to capture TimeFinder/Snap replicas from another TimeFinder/Snap point-in-time copy. This functionality is targeted mainly at SAP and Oracle environments in which copies of production environments are repurposed for testing, QA, or development. Work can proceed against an existing TimeFinder/Snap while duplicate copies can be created for additional downstream processes or checkpoint backups. With Enginuity 5875 snap copies can be taken from another snap source, adding even more disk space savings and flexibility through this track sharing technology.

Concurrent SRDF/A

Concurrent SRDF/A expands the SRDF multisite topology offering by allowing two separate asynchronous links from a Symmetrix VMAX to Symmetrix systems located at remote data centers. This configuration exploits the core benefits of SRDF/A for improved application response times while replicating at extended distances. Enginuity 5875 offers the flexibility to change Concurrent SRDF/S and SRDF/A disaster restart topologies to Concurrent SRDF/A and SRDF/A. Such flexibility is designed to:

- Meet performance goals during planned and known workload spike periods
- Offer a new migration option for data center relocation
- Provide additional disaster restart protection

SRDF general enhancements

SRDF continues to evolve to meet user requirements for ease-of-use improvements and respond to the changing needs of high-end business environments. The Enginuity 5875 general SRDF enhancements are:

- The ability to configure simultaneously multiple static SRDF groups
- The ability to throttle host write I/O response time up to a user-defined limit
- Allow TimeFinder/Snap off an SRDF/A R2 device
- When full-device TimeFinder/Clone, extent-level TimeFinder/Clone, or TimeFinder/Snap is run off an SRDF/A R2 device a host write I/O response time throttle is available for resource management

Space reclamation with SRDF

“Space reclamation” maintains flags marking which tracks have “never been written by host.” Currently this allows TimeFinder/Clone to replicate only tracks that have been written, greatly reducing the time needed to perform copy operations and to reclaim that zero space. With Enginuity 5875, this capability has been extended to include support for SRDF remote replication. Now when performing an SRDF remote synchronization process, the amount of data that must be replicated can be significantly less (zero data is not replicated), greatly reducing the amount of bandwidth consumed by the SRDF synchronization. This is particularly useful upon

completion of a Federated Live Migration operation, when administrators need to re-create the SRDF R1/R2 relationship of the new R1 and conduct a full resynchronization – the time to resynchronize and be fully protected is significantly reduced. Space reclamation when migrating standard volumes to VP volumes via SRDF/DM provides dual benefits of both reduced bandwidth requirements and smaller R2 volumes on the target array. Users can migrate or replicate standard (non-VP) volumes to standard volumes via SRDF (any mode) to reduce bandwidth requirements, but the R2 volume consumes the same amount of space as the R1.

Supported Enginuity combinations include:

- 5671 thick => 5875 thin
- 5773 thick => 5875 thin
- 5875 thick => 5875 thin

Zero space reclamation

Zero space reclamation allows customers to reclaim all-zero blocks. In 2009, Symmetrix released zero space reclamation, which allows customers to reclaim zeroed blocks within VP LUNs. The release of Enginuity 5875 for Symmetrix VMAX includes zero space reclamation as part of a migration process using Open Replicator for Symmetrix (Open Replicator) or Federated Live Migration (FLM). As data arrives at the destination Symmetrix VMAX, all zero blocks are stripped out and the remaining data is destaged to disk as part of a VP device.

Zero Space Reclamation for Open Replicator can be used when migrating from Symmetrix DMX™, CLARiiON®, and third-party arrays to Symmetrix VMAX and with FLM when migrating from Symmetrix DMX to VMAX. Continue to check [EMC E-Lab Interoperability Navigator](#) for the latest list of supported arrays.

Copy QoS

Local and remote replication have become integral to today's business practices. In Enginuity 5875 with Symmetrix VMAX, back-end copy operations are controllable at a very granular level to keep one replication task from overtaking other replication tasks. Individual priorities can be set for TimeFinder/Clone, SRDF, and VLUN device-level copy operations. This quality-of-service (QoS) implementation provides 16 priority-of-service levels. In addition, within these features it is possible to set QoS priorities on each independent feature session. Finally, the QoS mechanism is intelligent enough to allow dynamic resource utilization, meaning when back-end resources are idle, all operations will proceed more aggressively than the QoS setting. If the system returns to a busy state, the user-set QoS values will be reasserted.

Conclusion

Enginuity 5875 delivers a major advancement in automated storage tiering with sub-LUN FAST. Many enhancements are direct responses to user requests. New capabilities also improve ease of use, business continuity, and security for Symmetrix VMAX systems in an open systems environment.