

Module 2: Best Practices for Symmetrix VMAX

Upon completion of this module, you should be able to describe the best practices for:

- Configuring for capacity and performance
- Symmetrix Priority Controls and Dynamic Cache Partitioning
- Virtual Provisioning
- FAST and FAST VP



This module focuses on best practices for Symmetrix VMAX.

Module 2: Best Practices for Symmetrix VMAX

Lesson 1: Configuring Symmetrix VMAX for Capacity and Performance

During this lesson the following topics are covered:

- Best practices for configuring Symmetrix VMAX



This lesson covers considerations and best practices for configuring a Symmetrix VMAX for capacity and performance.

Symmetrix VMAX Features

- A Symmetrix VMAX array can have between 1 and 8 engines:
 - ▶ Each Engine has a pair of directors – an array can have from 2 to 16 directors
 - ▶ The directors are connected through a Virtual Matrix
 - ▶ Each engine provides front-end and back-end connectivity
 - ▶ Each engine is configured for 8 redundant Fibre Channel loops
- Each array can support between 48 to 2400 drives:
 - ▶ Maximum of 10 storage bays
 - ▶ Maximum of 240 drives per storage bay
- Front-end connectivity to hosts can be Fibre Channel, FICON, iSCSI or GigE



The Symmetrix VMAX array doubles the number of front-end and back-end ports as compared to the Symmetrix DMX-4 array. Symmetrix VMAX array supports heterogeneous host connectivity – Mainframe, Open Systems, IBM iSeries. Configuration changes, software and hardware updates can be performed online and non-disruptively. Capacity expansion and performance upgrades can be done online while the Symmetrix VMAX array is running production applications.

Start with a Strong Foundation

- Each resource has a limit of I/Os per second and MBs per second:
 - ▶ Disk drives
 - ▶ Back-end controllers (DAs)
 - ▶ Front-end controllers (Fibre, FICON, GigE)
 - ▶ SRDF controllers
- Configure enough components to support your workload:
 - ▶ Tools are available to help determine how many are needed
- Use those resources as uniformly as possible:
 - ▶ Skewed workloads are limited by most heavily utilized components

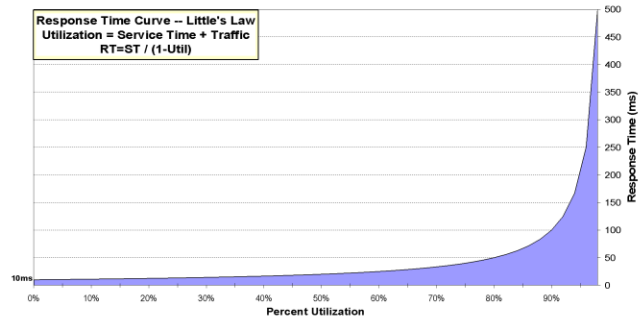


The Performance Engineering group conducts extensive tests to determine the limits of each component. These results are available to the members of the SPEED community. It is important to determine the expected workload (peaks) and map them to system resource limits right at the outset. SymmMerge is the tool of choice to design system components based on workload. As stated earlier SymmMerge is available for members of the SPEED community. It can model a target VMAX based on existing Symmetrix BIN and BTP files or based on application characteristics.

What Size System Would be Needed?

Utilization Limits

- The Symmetrix system and its components have performance limits.
 - It is important to understand the mixture of your workload and configure enough resources so that the array does not get close to 100% utilization on a regular basis.
- As a rule of thumb, a limit of no more than 50% utilization is good if response time is very critical
 - A higher utilization can be tolerated if only IO/s or total throughput matters



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

5

It is important to configure enough system resources so that no component hits a 100% utilization. As the utilization level increases, the response time increases dramatically. For performance critical configurations, utilization levels should be maintained at 50% or less.

How Many Disk Drives Should be Configured?

- Safely assume the following for small random I/O: *
 - ▶ Each 15 K FC disk can do 150 IOPS
 - ▶ Each 10 K FC disk can do 120 IOPS
 - ▶ Each 7200 SATA disk can do about 70 IOPS
 - ▶ Each EFD disk can do 2,000+ IOPS
- 1 host write can create many I/Os on the drive, depending on RAID protection.
- Balance all drives across all available DAs where possible:
 - ▶ Ideal configuration is multiples of 8 drives per VMAX engine
 - ▶▶ Of the same type/speed, not including spares
 - ▶ More important for EFDs

* With reasonable response time



Copyright © 2011 EMC Corporation. All Rights Reserved.

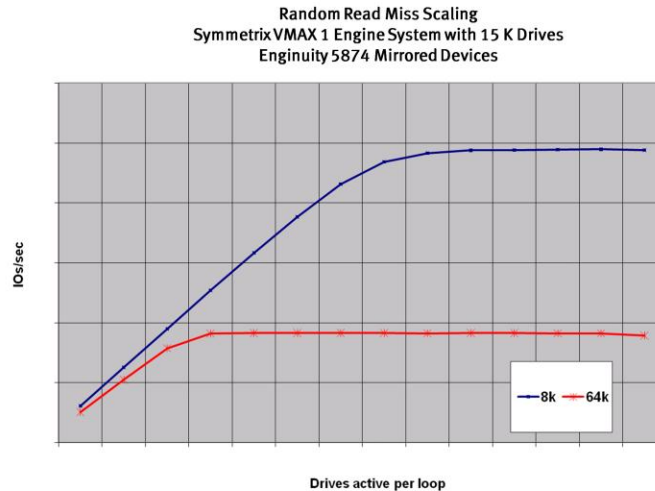
Module 2: Best Practices for Symmetrix VMAX

6

The number of disk drives should be driven by performance and not by capacity. The workload should be mapped to the desired drive type to determine the number of drives required to support it. RAID protection type should also be factored into determining the number of drives required.

Drive Scaling

- Adding drives to a system will add performance linearly, to a point:



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

7

As shown here, adding disk drives to a single loop scales the performance linearly to a point. Beyond this, the system limitations such as FA and DA utilization will come into play. Hence adding more drives beyond this point of saturation does not scale the performance.

What RAID Protection Should be Used?

- Performance of reads similar across all protection types:
 - ▶ If you are comparing the same number of drives!
- Major difference is with random write performance:
 - ▶ RAID 1: 1 Host Write = 2 Writes
 - ▶ RAID 5: 1 Host Write = 2 Reads + 2 Writes
 - ▶ RAID 6: 1 Host Write = 3 Reads + 3 Writes
 - ▶ Sequential writes for RAID 5 or RAID 6 can be more efficient than RAID 1
- Cost is also a factor:
 - ▶ RAID 5 or RAID 6 are best at 12.5% or 25% protection overhead
 - ▶ RAID 1 has a 50% protection overhead



Choosing the RAID protection is a balance between cost, performance and availability.

Front-end Connections

- Go wide before you go deep:
 - ▶ Use all the “zero” ports on director first and then the “one” ports
 - ▶ Spread across directors first, then on the same director
 - ▶ You also get increased availability if you spread wide first
- Each front-end port/slice has performance limits (see table below).
- Two active ports on one FA CPU do not generally do more IO/s.
- May need more paths if migrating from some other array such as Unified Storage:
 - ▶ Symmetrix has dedicated CPUs, while CX and VNX do not
- Front-end slices have different performance for different workloads:
 - ▶ For example (all with small I/Os on one “slice” normalized to hits)

FA Slice Performance	DMX-3/4	VMAX
Read Hit I/O	1	1
Read Miss	1/2	1/2
Overwrite	1/2	1/2
Write New Slot	1/4	1/4

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

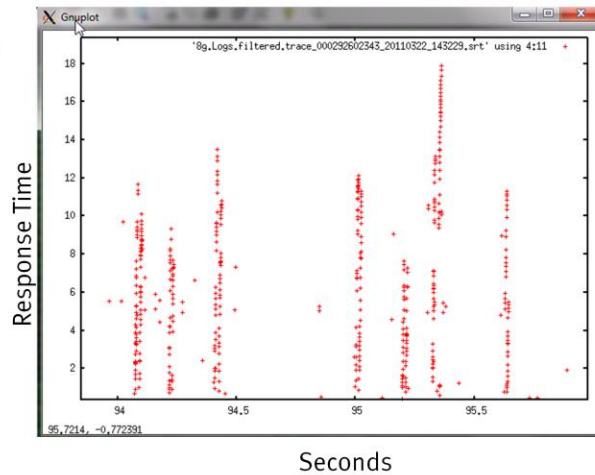
Module 2: Best Practices for Symmetrix VMAX

9

These are the recommendations for using the Front-end ports.

Front-end Performance

- Planning to use the average I/O rate may not be good enough.
- Writes tend to come in bursts.
- You need enough front-end CPUs to absorb the bursts.



EMC²

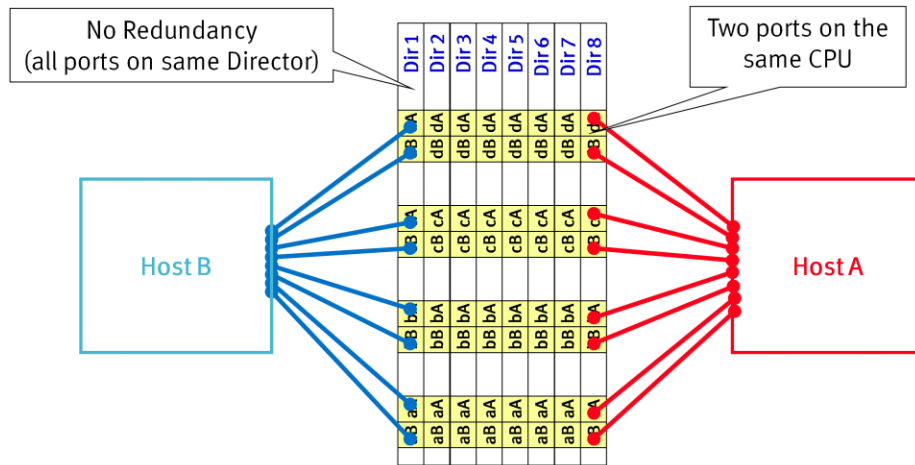
Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

10

As writes tend to come in bursts, using average I/O rates may not be good enough. A clear understanding of the I/O pattern through the days/weeks is necessary to plan for the peak activity.

Avoid Configuring Two Hosts With Eight Fibre Channel Adapters Each



EMC²

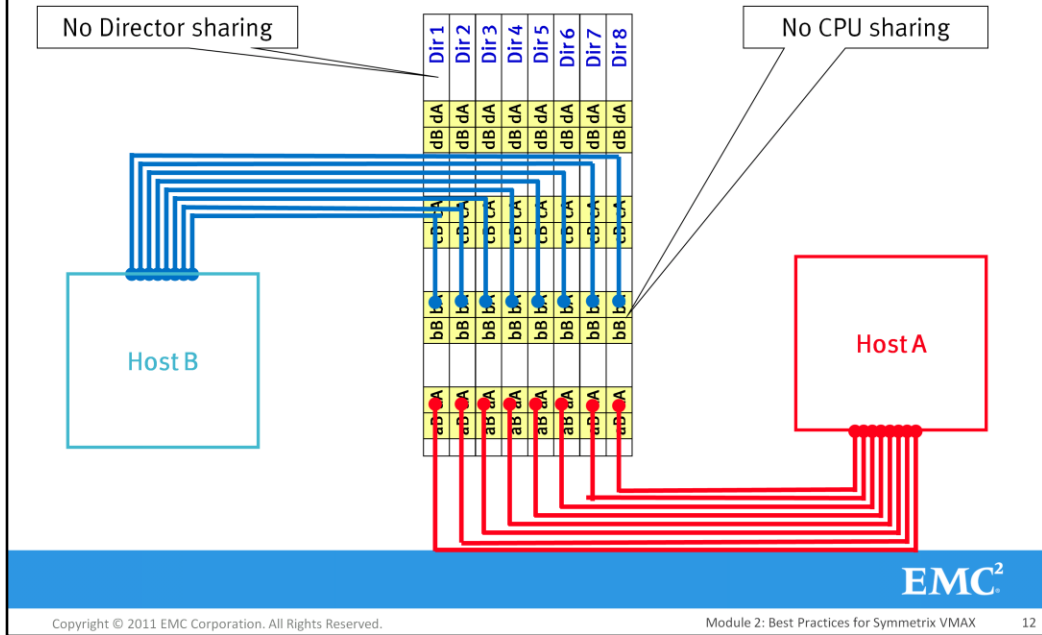
Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

11

For redundancy and performance all ports on one director and both ports on the same CPU should not be used to configure a multi-pathing solution.

Best Case: Two Hosts With Eight Fibre Channel Adapters Each



This configuration is the best way of ensuring redundancy and performance in this example. Each host's access is spread over all directors and the ports on the same CPU are not shared.

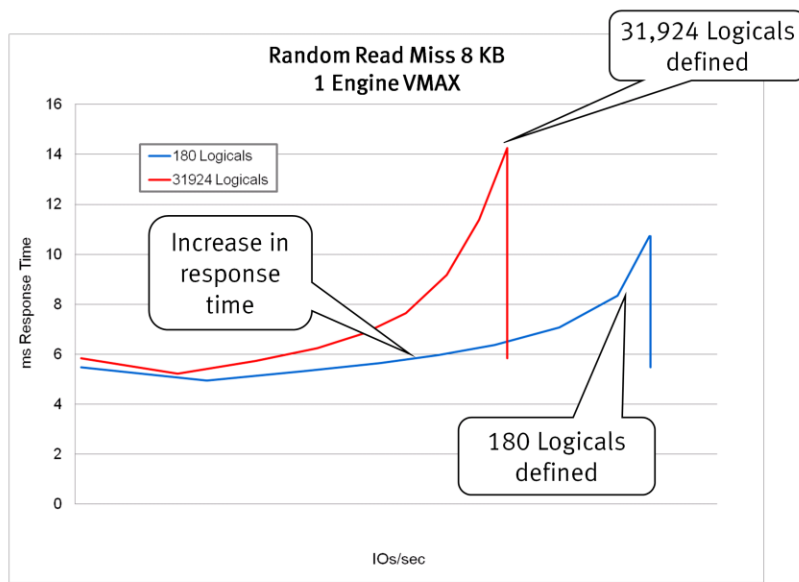
Logical Volumes and “Hypers”

- As a general rule-of-thumb, fewer and larger hypers will give better overall system performance:
 - ▶ There is system overhead to manage a logical volume, so it makes sense that more logical volumes could lead to more overhead
- Frequently, legacy hyper size is carried forward because of migration strategy:
 - ▶ However, the same hyper size that made sense 10 years ago on a 9 GB disk drive, does not make sense on a 2 TB disk drive today



Due to the overhead of management of logical volumes, fewer logical volumes with larger hyper sizes will give better overall system performance. When migrating to a new VMAX, consider changing the hyper sizes rather than carrying over legacy hyper sizes.

Effect of Adding More Logical Volumes



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

14

In both of these tests, the same 120 RAID5 logical volumes are active. In one case only 180 logical volumes are defined in the bin file, and in the other 31924 logical volumes are defined (31,744 added). The performance difference can be attributed to the extra overhead to manage all the additional logical volumes.

Meta Volumes Considerations

- Meta member counts should be “reasonable”:
 - ▶ Enginuity supports up to a 255 member meta
 - ▶ Symmetrix DMX-3/4 systems: ~15 TB volume
 - ▶ Symmetrix VMAX systems: 60 TB volume
 - ▶ Reasonable sizes for meta member counts are something like 4, 8, 16 or 32
- Even numbers are preferred:
 - ▶ Powers of 2 fit nicely into the Symmetrix system back-end configurations
 - ▶ Powers of 2 are not important for Virtual Provisioned (thin) meta volumes



Even though the system can support enormous sizes for meta volumes, the actual sizes should be kept reasonable. Powers of two are preferred for the meta member count.

Meta Volumes Consideration (cont)

- Getting enough I/O into a very large, striped meta can be a problem:
 - ▶ 32-way RAID 5 (7+1) meta volume would need at least 256 I/Os queued to have 1 I/O per physical disk
- Do not place striped meta volumes on the same RAID group (wrapped).
- Concatenated meta volumes can be placed on the same RAID group for some corner cases:
 - ▶ This can create an emulation of large RAID volume DMX systems or VMAX systems



Additional considerations for meta volumes are provided in this slide. As shown here, a large number of I/Os per second would be required to ensure at least one I/O per physical disk.

Virtually Provisioned Meta Volumes

- Meta volumes may still be needed with VP.
- For LUNs larger than 240 GB.
- For any size host LUN that requires very high performance:
 - ▶ LUNs doing 10s of MB/s or 100s of IO/s most likely would not need a striped VP meta volume for performance
 - ▶ LUNs doing 100s of MB/s or 1,000s of IO/s may need a striped VP meta volume for performance
- If expansion is never needed, use a striped VP meta volume.
- If ease of expansion is more important than performance, use a concatenated meta volume.



Virtually Provisioned striped meta volumes can be expanded starting with Enginuity 5875 Q4 2010 service release. The choice of striped versus concatenated depends on the requirements for ease of expansion or performance.

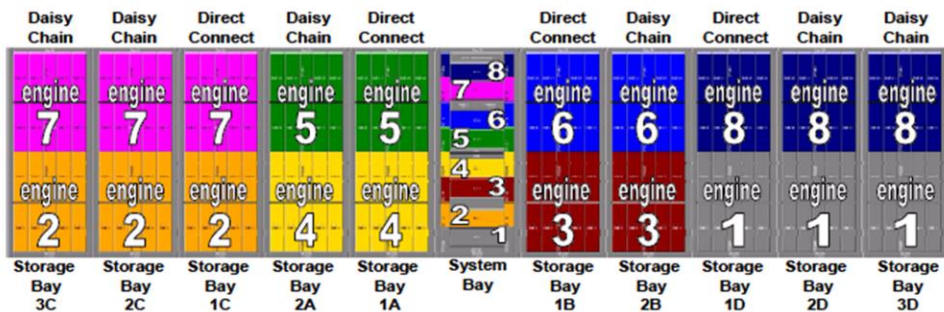
I/O Size and Striping

- For most host-based striping, 128 KB or 256 KB is good:
 - ▶ This is in actual capacity
 - ▶ If stripe is in 'blocks' you need to divide by block size
- May want to consider a smaller stripe size for database logs:
 - ▶ 64 KB or maybe smaller as advised by a Symmetrix performance guru
- I/O sizes between 64 KB and 256 KB are optimal for most workloads:
 - ▶ I/O sizes above 256 KB may actually decrease throughput
 - ▶ Try to avoid I/O sizes above 256 KB on all Symmetrix systems



When configuring host based striping, the stripe size should be kept at or below 256 KB.

Symmetrix VMAX – System and Storage Bays



- System Bay houses up to 8 Engines.
- Storage Bays are divided into an upper and a lower half.
- Engines in the upper portion of the System Bay are connected to the upper portions of the Storage Bays.
- Likewise, engines in the lower portion of the System Bay are connected to the lower portions of the Storage Bays.

EMC²

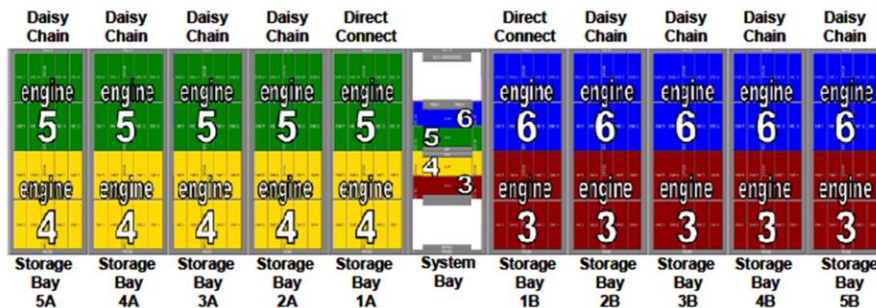
Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

19

Front view of a fully populated Symmetrix VMAX array is shown here. In this example the drives are cabled in a Standard Configuration. In this configuration there are 4 Storage Bays that are direct connected and 6 Storage Bays that are connected in a Daisy Chain. The division of the Storage Bays into upper and lower halves and the connection of the corresponding engines greatly simplifies the cabling requirements. The Engines in the System Bay are populated from the center to top and bottom. The Storage Bays are added from inside out. The Standard Configuration provides the maximum amount of resources – directors, processors, ports, cache.

Symmetrix VMAX – Extended Drive Loop Configuration



- Enables capacity expansion to 10 Storage Bays with 4 Engines
- Useful in instances where additional capacity is required, but additional resources are not – directors, processors, ports, cache
- Will require less power and cooling

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

20

Extended Drive Loop configuration enables expansion to the maximum of 10 Storage Bays but with 4 Engines. In this case there is one pair of Storage Bays that are direct connected and the other 4 pairs are Daisy Chained. Standard Configurations with up to 4 Engines and 4 Storage Bays can be converted to Extended Drive Loop Configuration. Standard Configurations with more than 4 Engines cannot be converted.

Tiered Storage

- Tiering using dedicated Symmetrix VMAX arrays based on application performance and criticality requirements:
 - ▶ Ultra-performance applications on a dedicated array with Enterprise Flash Drives (Tier 0)
 - ▶ Mission Critical applications on a dedicated array with Fibre Channel drives (15k RPM) and RAID 1 protection (Tier 1)
 - ▶ Business Critical applications on a dedicated array with Fibre Channel drives (10k RPM) and RAID 5 protection (Tier 2)
 - ▶ Backup, Recovery and Archive on a dedicated array with SATA drives (7.2k RPM) and RAID 6 protection (Tier 3)



Tiering using dedicated arrays will certainly eliminate interaction between competing or disparate application workloads. Tier 0/1 applications have high I/O rates. These workloads tend to push the backend performance. Use of EFDs for Tier 0 applications eliminates the limitations imposed by magnetic hard drives. For Tier 1 applications one can use smaller capacity, higher speed Fibre Channel drives. Tier 3 applications can be deployed on larger capacity, slower speed drives.

Tiered Storage (cont)

- Tiering within a Symmetrix VMAX array:
 - ▶ Segregating tiers
 - ▶▶ Dedicated Engines with appropriate drive types to meet application needs
 - Delivers predictable performance by segregating workloads on to different Disk Groups
 - Overall system performance is reduced as each tier has access to a restricted set of resources
 - ▶ Mixing drive types throughout the array
 - ▶▶ Drive types are mixed and spread throughout the array
 - Maximizes the availability of all resources for all tiers
 - FAST, Symmetrix Priority Controls (SPC), Dynamic Cache Partitioning (DCP) along with the appropriate drive type and RAID protection can be used to meet performance requirements for any tier



Tiering within the array can be achieved by isolating drive types into separate groups of disk enclosure and assigned to dedicated engines. This will provide predictable performance but the overall system performance will be lowered. For optimal performance different drive types should be mixed and spread throughout the array. Automated features such as FAST and FAST VP can be used in conjunction with SPC and DCP to fine tune application performance and resource requirements.

Should Volumes be Segregated?

- For the most *optimal* system performance, you should not segregate applications/BCVs/Clones/Thin Pools onto separate physical disks, DAs or engines.
- For the most *predictable* performance, you should segregate applications/BCVs/Clones/Thin Pools onto separate physical disks, DAs or engines.
- When segregating, make sure the components are able to handle the total workload.
- You should **not** segregate if the workloads on the various sets of volumes in question are not concurrent with one another, even if you desire the most predictable performance.



Segregating volumes is generally not recommended. Not segregating will provide optimal system performance. However if predictable performance is desired for different application, then segregating the physical disks, DAs and engines can be done. If for example an application is segregated to a set of physical disks and engines, then the capacity of these components to handle the application workload must be evaluated.

Module 2: Best Practices for Symmetrix VMAX

Lesson 2: Symmetrix Priority Controls and Dynamic Cache Partitioning

During this lesson the following topics are covered:

- Symmetrix Priority Controls
- Dynamic Cache Partitioning



This lesson covers Symmetrix Priority Controls which can be used to manage preferential disk resource allocation for different application workloads. This lesson also covers the concepts of Dynamic Cache Partitioning.

Symmetrix Priority Controls

- Allows prioritizing I/O requests in the disk adapters and SRDF directors.
- Disk adapters prioritize read requests.
- SRDF directors prioritize SRDF/Synchronous data transfers over the link.
- Ensures that high-priority workloads have minimal disk and SRDF/S queue time in congested environments.
- Does not penalize low priority workload when disk or SRDF/S adapter utilization is low.
- At high utilization levels priority (range of 1-16; 1=normal, 16=lowest) will dictate servicing of the I/Os.
- Priorities can be set via SMC or SYMCLI, at a Symmetrix Logical Volume level.
- Symmetrix Priority Controls are not available for Virtual Provisioning.



By giving preference to high-priority workloads at the disk level, Symmetrix Priority Controls helps minimize disk queue time in congested environments. SPC is will become effective only when the utilization levels at the disk or the SRDF queue are high. Under low utilization levels both the high and the low priority workloads will have equal service times.

Conditions for Effective Prioritization

- Prioritization is most effective when different application service classes share the same physical disks.
- For SRDF/S prioritization is most effective when different application service classes share the same SRDF directors.
- If application service classes are isolated on different disks/storage tiers, or are assigned to different SRDF directors, SPC cannot deliver any differentiation.
- Even when physical disks and/or SRDF directors are shared, prioritization is ineffective if applications are active at different times.
- If applications sharing the same spindle have different I/O profiles, such as high read miss for one and majority of writes for another, prioritization will not deliver any substantial benefit.
- If disk or SRDF director utilization is low, SPC will not introduce any unnecessary delays in data transfer.



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

26

If different service classes do not share the same spindle(s) then SPC cannot be effective. For example if workloads have been isolated to different tiers of storage then each workload has the same pattern of access its own tier. Prioritization cannot deliver preferential access in this case.

If different service classes for SRDF/S workload are already isolated on different SRDF directors, again SPC cannot be efficient.

Further more, if the applications share the same spindle(s), but are active at different times, this will lead to normal utilization levels. SPC becomes effective only when utilization levels are high. So in this case, SPC will not be effective as well.

SPC and SRDF/S, SPC and Flash Drives

- Data transmission across the SRDF links from primary to secondary will use SPC values to re-order the tasks in the SRDF director queue.
- Additionally SPC values can be set on the secondary devices.
- In the event that the primary has to read from the secondary, SPC values on the secondary will be used to prioritize these requests.
- As Flash drives deliver millisecond read/write response times, SPC should not be used for these.
- SPC should be reserved for use with traditional spinning hard disks.



Because Flash Drives deliver millisecond response times, SPC should not be set on SLVs that reside on these drives. When different service classes share the same SRDF directors for SRDF/S, then the higher priority workload will get preferential data transfer over the links, only when the SRDF director utilization is high. For SRDF/S the assigned SPC value applies equally to both local disk reads on the R1 devices as well as to data transfers to the R2 Symmetrix. If data has to be read from the R2 device, the SPC values on the R2 devices will dictate the priority of data transfer from R2 to R1.

Dynamic Cache Partitioning

- Provides control over cache resource allocation to different applications sharing the same Symmetrix Array.
- Up to eight cache partitions are supported in a Symmetrix Array:
 - ▶ One default partition and an additional seven named (31 characters) partitions can be created
- Partitions are logical and are evenly distributed across physical cache resources.
- Partitions can be:
 - ▶ Flexible
 - ▶ Static
- Provides an Analyze mode to observe the effects of partitions prior to enabling DCP.
- Allocation can be changed dynamically using Symmetrix Management Console (SMC) or SYMCLI.



When different dissimilar workloads are consolidated on the same Symmetrix array, it becomes difficult to achieve efficient resource utilization and maintain guaranteed service levels for each application. Dynamic Cache Partitioning allows Storage Administrators to segregate different applications into their own Cache Partitions. The partitions are logical entities and are spread over all available physical cache resources. Flexible or static cache partitions can be created. Management of the cache partitions can be performed using SYMCLI or SMC.

DCP Configurable Parameters

- Allocation as a percentage of total available cache:
 - ▶ Minimum [Min%] 0 – 100%
 - ▶▶ Partition will not go below this slot allocation
 - ▶ Target [Target%] between 10% and 100%
 - ▶▶ Desired slot allocation. Sum of all partitions should equal 100%
 - ▶ Maximum [Max%] 10% – 100%
 - ▶▶ Maximum slot allocation for the partition
- Donation age:
 - ▶ If a slot is not accessed within the donation age, it becomes eligible for donation to another partition
 - ▶ Valid for flexible partitions only
- Write Pending Limit [WP%] as a percentage of the partition size.



The minimum and maximum allocation parameters control Flexible partitions. If a partition is allowed to donate or accept cache to or from other partitions (based on the donation age), then these two parameters set the flexibility of this partition. Cache used by the partition will not go below the minimum allocation and will also not exceed the maximum allocation. Target defines the starting partition size. As noted in the slide, the sum of all targets should equal a 100% of the available of cache.

A cache slot will become eligible for donation to flexible partitions if its fall through time becomes greater than the donation age. Fall through time is the time since the slot was last accessed. So if a slot is not reused before the donation age is reached, it becomes a candidate for donation to a different partition. A high donation age will decrease the chance of donation and conversely a low donation age will increase the chance of donation.

Each partition can have its own Write Pending Limit. The limit can be set between 40% and 80% of the partition size. When the write pending limit is approached, host write activity will be throttled at the Symmetrix front-end adapters. When the write pending limit is reached, high priority destage operations will begin at the Symmetrix back-end adapters. Write pending limit can be used to curtail the cache utilization of write intensive workloads. By setting a lower write pending limit, host write activity can be contained.

Flexible and Static DCPs

- Flexible partitions:
 - ▶ Have overlapping minimum and maximum settings
 - ▶ Can donate cache slots to other partitions if donation age is reached
 - ▶ Can accept cache slots from other partitions when workload varies
- Static partitions:
 - ▶ Min%=Max%=Target%
 - ▶ Partition size is fixed
 - ▶ Cannot donate/accept slots to/from other partitions
 - ▶ Provides absolute resource segregation
 - ▶ Results in predictable and controlled service levels
 - ▶ Donation age is irrelevant
- Flexible and Static partitions can co-exist.



Static partitions are very useful for storage service providers as these guarantee cache resources. They also simplify chargeback calculations. Flexible partitions optimize cache utilization and can improve overall performance. If a Flexible partition has cache slots that have not been accessed within the donation age, it can donate them to another Flexible partition that requires it due to its sudden variation in workload.

With Flexible partitions cache slots can migrate between different Flexible partitions based on demand and usage or access. When cache slots are migrated to a different partition, their usage calculation begins anew. So their time since last access is reset and recalculated in the new partition before further migration. Higher the donation age, the lesser the chance of cache slot migration. This is because it requires a long interval between cache slot reuse before donation can take place. A lower donation age will increase the possibility of migration to a different partition. Cache slots will be accepted by a partition only until the Maximum Percentage is reached.

Module 2: Best Practices for Symmetrix VMAX

Lesson 3: Virtual Provisioning

During this lesson the following topics are covered:

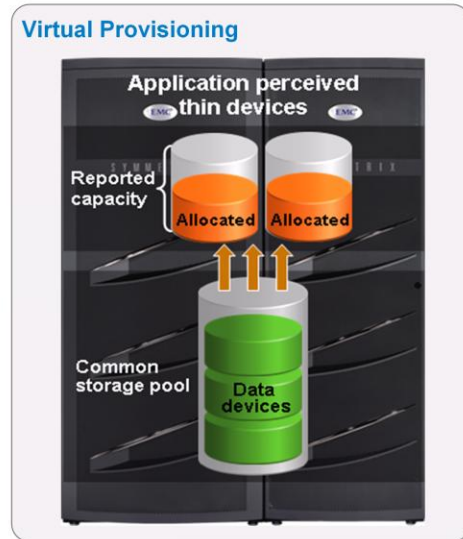
- Virtual Provisioning Overview
- Benefits of Virtual Provisioning



This lesson covers considerations and best practices for Symmetrix Virtual Provisioning.

Virtual Provisioning – High Level Description

- Known in the industry as Thin Provisioning.
- Large volume presented to a host consumes physical storage from a shared pool only as needed.
- Managed and monitored with Symmetrix Management Console (SMC) and the Solutions Enabler command line interface (SYMCLI).



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

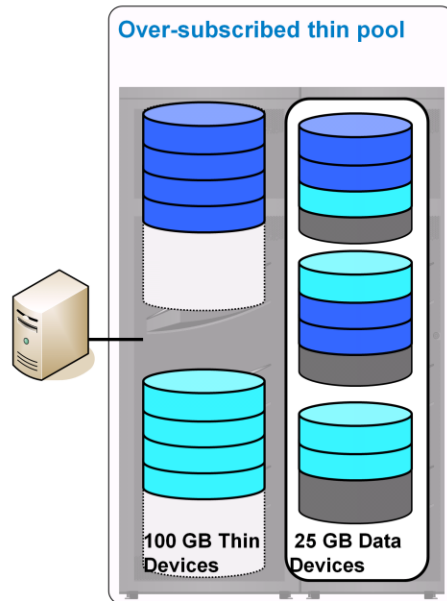
Module 2: Best Practices for Symmetrix VMAX

32

Symmetrix Virtual Provisioning, also known in the industry as “thin provisioning”, is the ability to present a host and therefore an application, with more storage capacity than is physically allocated to it in the storage array. The physical storage is then allocated to the application “on-demand” as it is needed from a shared pool of storage.

Over-subscription

- Bound TDEV capacity is greater than available capacity in the thin pool:
 - ▶ Addresses inefficiency of under utilized storage
- Oversubscription has risks:
 - ▶ If the pool becomes full, writes to unallocated extents on TDEVs fail and an IO error is returned
 - ▶ Storage administrator must monitor pool to prevent “pool full” condition
- Data devices can be dynamically added to the pool:
 - ▶ Best practice is to add multiple devices to the pool to prevent bottlenecks
 - ▶ Pool Rebalancing is also recommended



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

33

Over-subscription is where the total capacity of all TDEVs bound to a Thin pool is greater than the aggregate capacity of all data devices in the pool. This is a normal practice but does require that a storage administrator monitor the pool to prevent a “pool full” condition. When the pool is full and a host attempts to perform a write that requires a new extent to be allocated, a write error is returned to the host.

Different hosts react differently when this error is experienced. The Host Connectivity Guide for each operating system describes the behavior of the specific operating system, volume manager and file system. These Guides are available on Powerlink.

When additional capacity is needed in the pool, data devices can be dynamically assigned to the pool. It is recommended practice to add multiple devices to the pool at the same time to ensure that the allocations to a thin device are spread across multiple data devices. It is also advisable to initiate a pool rebalancing operation.

By default there is no limit to oversubscription. To reduce the risk of over subscription, a limit can be set for a pool that would limit the total capacity of the thin devices that are bound to a thin pool as a percent of the pool capacity.

Benefits of Virtual Provisioning – 1

- Improved capacity utilization:
 - Reduce the amount of allocated but unused physical storage
 - Avoid over allocation of physical storage to applications
 - Reduces energy consumption and footprint
- Ease and speed of provisioning:
 - Provision independently of physical storage infrastructure
 - Minimize the challenges of growth and expansion
 - Simplifies data layout
 - Saves costs by simplifying procedures to add new storage



One of the issues with the traditional way of storage provisioning is the amount of allocated but unused storage. When a host is allocated a lun it has exclusive access to that lun. Any unused space on the lun can not be accessed by any other host. It becomes unused storage. With Symmetrix virtual provisioning, storage is allocated from a shared Thin pool. Multiple hosts and applications can access this shared storage. This means the utilization of storage capacity can be optimized when using Symmetrix Virtual Provisioning. There will be less unused storage.

Another advantage is the ease and speed of provisioning, specifically re-provisioning. While the initial provisioning process is the same for thin devices as it is for regular “thick devices”, adding additional capacity is simply a matter of adding data devices to the Thin pool. By presenting to a host a larger device than is initially required, re-provisioning operations are not required at the host level.

Thus, Symmetrix Virtual Provisioning can simplify storage management. Storage provisioning can be done independently of the physical storage and it can reduce the re-provisioning steps required to support capacity growth.

Benefits of Virtual Provisioning – 2

- Pool based view of back-end layout allows efficient use of available resources:
 - ▶ Wide striping distributes I/O across spindles
 - ▶▶ No requirement for LVM striping or striped meta volumes except sometimes with SRDF/S
 - ▶ Reduces disk contention and enhances performance
 - ▶ Maximize return on investment



An added benefit is improved performance for certain workloads. There is the overhead of the initial allocation, however because the Thin pool utilizes wide striping, there is a performance benefit for subsequent access.

With Synchronous SRDF, Enginuity allows one outstanding write per thin device per path. With concatenated metadevices, this could sometimes cause a performance problem by limiting the concurrency of writes. This limit will not affect striped metadevices in the same way because of the small size of the metavolume stripe (1 cylinder or 1920 blocks).

Symmetrix Virtual Provisioning: Components

- Thin Device
- Data Device
- Thin Pool

These components take part in various operations such as binding, draining, etc.



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

36

Symmetrix Virtual Provisioning uses a type of Symmetrix device called a Thin Device. It is called a thin device because no actual data is saved on the device. The thin device is presented to a host like any Symmetrix volume and can be used by a host application to write data to it. Thin devices can be presented to hosts before they're bound, but they'll appear as "not ready" to the host.

Data does not get written to the thin device. It is written to a collection of symmetrix devices, called data devices.

The data devices are grouped together into a storage pool called a Thin Pool. A Thin Device must be bound to a Thin Pool before it can be used. This Thin Pool provides the shared storage. The storage capacity in the Thin Pool can be shared among multiple hosts and applications.

The action of binding a thin device to a thin pool makes the device ready and associates storage with the thin device. It also allocates one extent (twelve 64 KB tracks or 768 KB) from the pool.

Configuring Thin Pools

- Different pools for different classes of storage.
- Different pools for different application families.
- All data devices should reside on drives with the same rotational speed.
- All data devices should be spread evenly across as many DAs and drives as possible.
- Data devices should be the same size, if possible:
 - ▶ Using different sizes could result in uneven data distribution
- Fewer pools with many data devices are better than many pools with fewer data devices.



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

37

When planning a configuration using thin devices the first step involves determining how many separate Thin pools are needed and the required composition of each thin data pool. Typically this will involve conceptually organizing disk storage into separate classes, with further subdivision as needed to allow the back-end resources (drives and DAs) used by the pools to be isolated from one another. Depending on the mix of applications to be placed on thin devices, it will often be necessary to create multiple thin pools but generally, the most efficient use of resources will be achieved by using a minimal number of pools.

Typically, a Thin pool should be designed for use by a given application, or set of related applications, aligned with a given business group. The applications sharing a Thin pool will compete for back-end resources, including Thin pool storage capacity, so applications should not share the same pool if this is not acceptable. The devices comprising a Thin pool will have the same performance and protection properties, so the applications sharing a Thin pool should have the same performance and protection requirements.

Meta Devices Considerations

- Meta devices are used when the host needs larger devices:
 - ▶ Maximum Symmetrix device size limit of 262668 cylinders (240 GB)
- Striped metadevices (recommended in most cases):
 - ▶ Better performance with Synchronous SRDF
 - ▶ Better read performance with pre-5875 Enginuity
 - ▶ More IOs can be queued to striped metas because of higher Write pending limits caused by each member getting a percentage of cache
- Advantages of concatenated metadevices:
 - ▶ Non meta devices can be converted to meta devices without destroying existing data
 - ▶ Concatenated thin meta devices can be expanded without destroying data



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

38

In most cases, EMC recommends using striped rather than concatenated metadevices with Virtual Provisioning because they offer the following advantages:

•With Synchronous SRDF, Enginuity allows one outstanding write per thin device per path. With concatenated metadevices, this could cause a performance problem by limiting the concurrency of writes. This limit will not affect striped metadevices in the same way, however, due to the small size of the metavolume stripe (1 cylinder or 1920 blocks).

•In Enginuity releases prior to 5875, the code allows eight read requests per path per thin device. This may limit the number of read requests that can be passed through to the thin pool regardless of the number of data devices it contains. This can cause slower performance in environments with a high read miss rate. A striped metadevice offers improved throughput, because each metamember is allowed to queue requests to the back end in parallel.

Even with the limit removed in Enginuity 5875, there are still potential performance benefits to using striped thin metas instead of concatenated thin metas.

•The Symmetrix Enginuity operating environment has a logical volume write pending limit to prevent one volume from monopolizing writeable cache. Because each meta member gets a small percentage of cache, a striped meta is likely to receive more writeable cache for the metavolume.

However, concatenated thin metadevices do offer two important operational advantages over striped thin metadevices:

•Non-metadevices can be converted to concatenated metadevices without destroying the existing data. And, beginning with Enginuity 5875 and Solutions Enabler 7.2, it is possible to convert a thin device to a concatenated meta without first unmapping the device.

•Concatenated thin metadevices can be expanded without destroying existing data by adding metamembers to an existing metadevice. And, beginning with Enginuity 5875, a striped metadevice can also be expanded while preserving existing data.

Performance Considerations

- Extent allocation is striped across all enabled data devices in pool:
 - ▶ With regular Symmetrix volumes, wide striping can be achieved using striped meta devices
 - ▶ Thin devices are widely striped automatically
 - » Makes it easier to avoid hot spots
 - » Potential performance improvement for some workloads
- Pool-based view of storage layout:
 - ▶ Thin pools should contain storage of a particular class (or tier) with a defined business purpose
 - ▶ Planning consists largely of deciding what tier of storage is required and which business group's resources should be used
 - ▶ Storage is viewed at a higher level when planning new devices
 - ▶ By managing Thin pools with FAST VP storage can be quickly and dynamically relocated to a different tier when needed



Thin device performance can be better than regular devices. However, real world application performance is highly dependent on the applications that are sharing the Thin pool.

In addition to potential contention for access to shared data device, response time and throughput can be impacted by the overheads incurred the first time a write is performed on an unallocated region of a thin device. This will disappear once the “working set” of a thin device has been written.

When managed with FAST VP, thin pool extents can be dynamically migrated to higher or lower performing storage tiers depending on the performance needs of an application.

Thin Friendly Environments

- Applications with predictable and controlled growth.
- Applications that do not format storage before it is needed, e.g.
 - ▶ Databases that use “auto-extend” capabilities during page formatting or table creation
- Applications with tolerance for performance variability are suitable for VP such as:
 - ▶ Document and Media Repositories
 - ▶ Test Development and QA
- Flash drives:
 - ▶ Compared to rotating drives, flash drive performance does not degrade as disks fill up



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

40

Applications with a predictable and controllable growth rate are good fits for VP because they do not run the risk of unexpectedly consuming most or all of the shared storage.

Virtual Provisioning is appropriate for applications for which some performance variability can be tolerated. Some workloads will see performance improvements from wide striping with Virtual Provisioning. However, when multiple thin devices contend for shared spindle resources in a given pool, and when utilization reaches higher levels, the performance for a given application and thin device might become more variable.

Document repositories can be a strong fit for Virtual Provisioning, provided the environments meet the performance, growth, and other criteria outlined previously, as rapidly rising capacity requirements for these repositories create the opportunity for improved capacity.

Many organizations will see an opportunity to lower TCO by improving ease of use and capacity utilization for storage associated with test, development, and QA activities, which often can tolerate some level of performance variability.

Virtual Provisioning can help ensure that Flash drive resources are highly utilized. Flash drives can offer excellent response times as well as significant IOs/GB. With a traditional drive, larger seek times are encountered during reads as the drive fills up. Read response times in Flash drives do not decline as utilization of the drive increases.

Thin Hostile Environments

- Any application that formats storage without using it causes unused capacity to become fully allocated, such as:
 - ▶ File systems that append data to file system instead of reusing free space
 - ▶ Some utilities that copy “raw” data such as “dd” in Unix
 - ▶ Administrative operations such as bad block detection or file system checks that cause dense write patterns on all reported storage



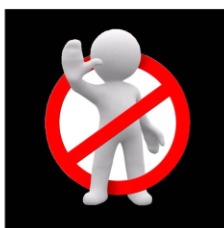
There are a variety of factors that can contribute to making a given application environment thin-hostile. One step or a combination of steps involved in simply preparing storage for use by the application may force all of the storage that is being presented to become fully allocated.

If the storage space management policies of the application and underlying software components do not tend to reuse storage that was previously used and released, that will increase the speed in which underlying thin devices become fully allocated.

Data copy operations such as disk balancing operations and de-fragmentation operations that are carried out as part of the administration of the environment.

If there are administrative operations such as bad block detection operations or file system check commands that perform dense patterns of writes on all reported storage.

Avoid Pool Full Condition by Proper Monitoring



- When pool fills up:
 - ▶ Next write to a logical block on an unallocated thin device extent fails
 - ▶ Data is still readable
- Different Operating Systems react differently to a pool full condition:
 - ▶ For details check the host connectivity guides on Powerlink
- To free up pool space:
 - ▶ Add data devices
 - ▶ Migrate thin devices to another pool using VLUN VP
 - ▶ Run zero space reclamation on thin devices
 - ▶ Unbind thin devices

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

42

Given the nature of Virtual Provisioning, there is always the danger that storage space in a Thin pool is exhausted due to oversubscription. This condition should be avoided at all costs by proper monitoring and taking preventive action. However, should this circumstance occur, this is what happens:

Once space in the pool is exhausted, no new writes to the thin devices are allowed. The reaction to the failed write will depend on the application and the Operating System. The Host Connectivity Guides for all operating systems are available on EMC Powerlink and describe the behavior when the “No space on Device” condition is encountered.

The data in the Thin pool is still available for reads.

To free up pool space you can add data devices to an existing pool or migrate selected thin devices to another pool.

Running zero space reclamation on the thin pool will free up space if thin devices have allocated but unused space.

If there are bound thin devices whose data is no longer needed, those devices can be made not ready and unbound from the pool, thus freeing up pool space.

Monitoring Thin Pools

- Thin pool warning thresholds depend on application environment:
 - ▶ Default EMC warning thresholds start when the pool is 60% full
- Pools should be configured to absorb data from occasional “runaway” applications.
- Monitoring can be done with:
 - ▶ Symmetrix Management Console (SMC)
 - ▶ Solutions Enabler `symcfg monitor`
 - ▶ Best to monitor with event daemon `storeventd`



Virtual provisioning simplifies, but does not entirely eliminate, the need for proper planning. When implementing Virtual Provisioning it is important that realistic utilization objectives are set. Generally, organizations should target no higher than 60 percent to 80 percent capacity utilization per pool. A buffer should be provided for unexpected growth or a “runaway” application that consumes more physical capacity than was originally planned for. There should be sufficient free space in the Thin pool equal to the capacity of the largest unallocated thin device.

Thin pool storage can be managed with SMC, Solutions Enabler and the event daemon. The event daemon offers the most flexible way of monitoring free space in a Thin pool.

Module 2: Best Practices for Symmetrix VMAX

Lesson 4: FAST and FAST VP

During this lesson the following topics are covered:

- Recommended best practices for FAST
- Recommended best practices for FAST VP



This lesson covers considerations and best practices for implementing FAST and FAST VP in a Symmetrix VMAX.

Two Variations on FAST

- FAST (also referred to as FAST DP) supports disk group provisioning for Symmetrix VMAX:
 - ▶ Full LUN movement of disk group provisioned (“thick”) devices
 - ▶ Supports FBA and CKD devices
 - ▶ Introduced in Enginuity 5874
- FAST VP supports virtual provisioning for Symmetrix VMAX:
 - ▶ SubLUN movement of thin devices
 - ▶ Supports FBA devices only
 - ▶ Available in Enginuity 5875



FAST moves application data at the volume (LUN) level. Entire devices are promoted or demoted between tiers based on overall device performance.

FAST VP adds finer granularities of performance measurement and data movement. The data from a single thin device under FAST control can be spread across multiple tiers. The FAST controller is free to relocate individual sub-extents of a thin device, based on performance data gathered at the extent level.

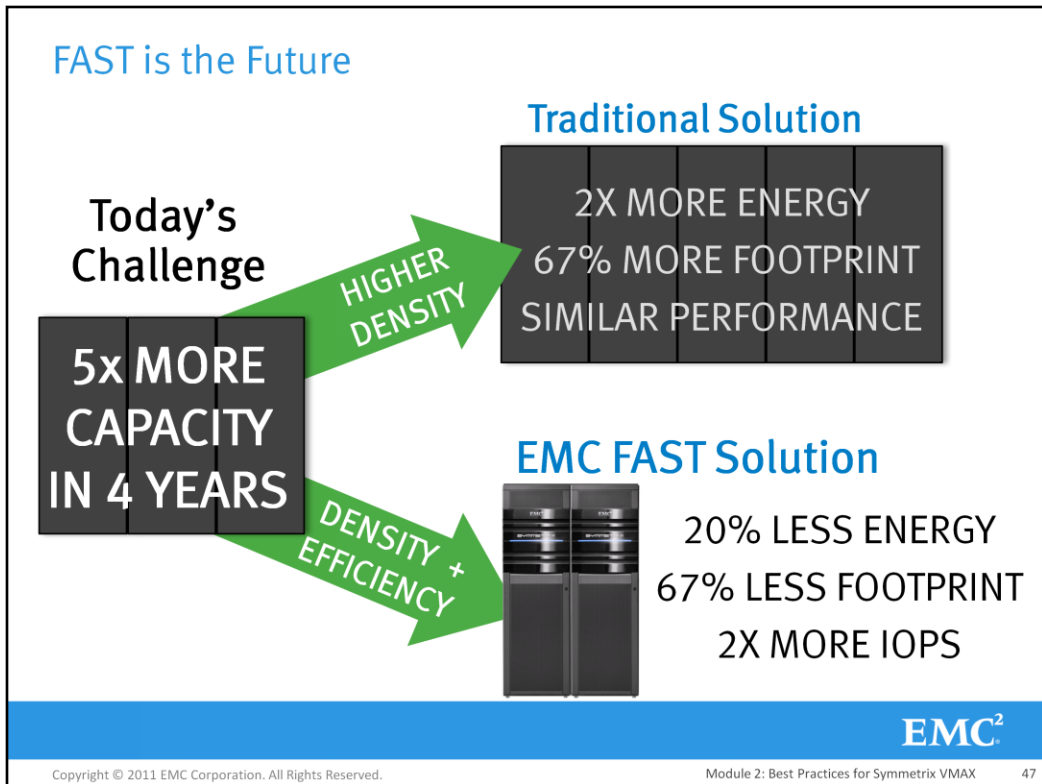
When to use FAST and FAST VP

- Workloads with a higher skew will benefit more from FAST or FAST VP:
 - ▶ Workloads with skew above 80/20 are considered good candidates
 - ▶ Unbalanced workloads direct a higher percentage of I/O to a small percentage of the storage allocated
 - ▶ Heavily utilized devices are moved to faster technologies, to reduce response time
 - ▶ Under utilized devices are moved to less expensive technologies, to reduce cost
- Workloads with a lower skew may not benefit:
 - ▶ Workloads with a skew closer to 50/50 (uniform workload) are less likely to contain candidates for promotion/demotion



FAST is most effective in an environment where the workload tends to be skewed and the skew patterns are dynamic. For example, if devices that are heavily accessed this month may be underutilized next month, a suitable FAST policy will continually perform the necessary relocations to locate the heavily used devices on faster performing drives.

FAST is not beneficial if the workload is not skewed or if the I/O workload is static, since workload skews can then be addressed using manual methods such as Symmetrix VLUN Migration.



Customers continue to experience massive growth of their retained data capacities. As this grows 60+% per year, we can look at the next 5 years and see where the problems will be. If nothing changes, the required storage space will be huge, even when accounting for progress in disk drive and array hardware technology. We expect to see:

- 67% larger physical system than today.
- Double the energy usage than today.
- Roughly the same performance per GB as current systems.

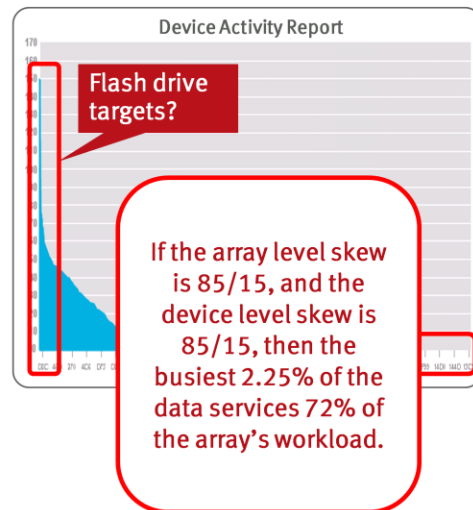
With FAST VP, we see a very different future. Taking advantage of the coming hardware advances AND combining that with FAST VP, we expect to see:

- Lower power consumption than today.
- Less than half the physical size of today.
- Double the iops per GB compared to current systems.

The future state without FAST assumes the move to 2.5" drives for all platforms to get reduced size, power, and weight. As we look out 4 years, it is clear that shift will be present for most higher performance spinning media.

Workload Skew

- Skew defines an asymmetry in device usage over time:
 - ▶ 85% of the workload serviced by 15% of the devices
- FAST accounts for array level skew:
 - ▶ Relocates busiest devices to higher performing tiers
 - ▶ “Idle” devices moved to more cost-effective tiers
- FAST VP accounts for device level skew:
 - ▶ Relocates busiest thin device extents to higher performing tiers
 - ▶ “Idle” data moved to more cost-effective tiers



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

48

The chart on the left is a Device Activity Report that shows a small number of active devices that are candidates for higher performance drives and a large number of inactive devices that are candidates for lower performance drives.

FAST moves complete logical devices to higher or lower performing tiers. In contrast, FAST VP moves thin device extents as necessary.

FAST DP Usage Considerations – 1

- Use Symmetrix Performance Analyzer (SPA) snapshot views to determine applicable workloads for FAST.
- Use Tier Advisor to assist with recommendations for tiering.
- Use many SATA drives spread across the back end to reduce IO traffic on individual drives.
- FAST device moves and swaps can impact to disk response times.



Symmetrix Performance Analyzer is a graphical tool that is bundled together with SMC. SPA snapshot views can provide a good idea of environments that would benefit from FAST.

Tier Advisor is a modeling tool that estimates the performance and cost of mixing different types of disk drive technology within EMC storage arrays. It helps you determine whether to add, upgrade, or replace existing storage to achieve optimal performance and cost effectiveness in your storage environment.

SATA drive technology is suited for data that is not accessed often, or data that is accessed mostly sequentially. It is best for I/O bandwidth to be shared by the larger pool of SATA disks (in comparison to FC disks). In this way, on an average, each SATA disk will be accessed less often.

FAST device move and swap operations may result in a short term impact to disk response time for a very limited period. In the case of swaps, the impact will be greater due the greater number of devices involved. This impact is not specifically caused by FAST; however, it is a consideration for FAST data movement.

FAST DP Usage Considerations – 2

- Configure a single standard size of device on all tiers:
 - ▶ Increases probability of like-sized devices being available to perform swap
 - ▶ Smaller LUN sizes will benefit more from FAST
 - ▶ Larger LUN sizes may also benefit, if workload is sufficiently skewed
- Meta devices are moved/swapped as complete entity:
 - ▶ Meta device members cannot exist across physical disk groups
 - ▶ Members of striped meta devices may not exist on same physical disks for FC and SATA
- For optimal system performance use Optimizer in concert with FAST:
 - ▶ Optimizer will balance load within a tier



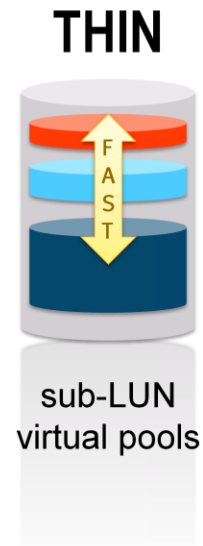
Using a single size of device within the array allows greater flexibility in executing device swaps. Smaller LUNs tend to experience greater workload skews and can benefit more from the implementation of FAST.

Meta device members are not permitted to exist on different disk groups. FAST will not cause a striped metavolume to end up with two members of the stripe on the same physical disk.

Symmetrix Optimizer complements the functionality of FAST. The former does load balancing within disk groups, whereas the latter performs load and capacity balancing across disk groups.

Customer Acceptance of FAST VP

- FAST VP is working very well in production at customer sites.
- Over 50% of all new VMAXs are shipping with a FAST VP licence.
- FAST VP is driving the demand for EFDs and SATA disks:
 - ▶ 92% of VMAXs with FAST VP are shipping with EFDs
 - ▶▶ Over 80% of EFD's shipped in Q2 were 200GB EFDs
 - ▶ 91% of VMAX with FAST VP are shipping with SATA disks



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

51

Since its release in Enginuity 5875 FAST VP has overtaken FAST in customer acceptance, because of its capacity to quickly move small chunks of data.

Data movements performed by the microcode are achieved by moving allocated extents between tiers. The size of data movement can be as small as 768 KB, representing a single allocated thin device extent, but will more typically be an entire extent group, which is 7,680 KB in size.

What Makes FAST VP so Different?

- FAST VP reacts quickly:
 - ▶ Initial review period as short as 2 hours
 - ▶▶ Ongoing reviews every 10 minutes
 - ▶ Ongoing movement when thresholds are crossed
- FAST VP tiers easily:
 - ▶ Up to 3 tiers of storage (EFD, FC, ATA)
 - ▶ Tiering across discrete pools
 - ▶ Tiering controlled at the storage group (or LUN) level
- FAST VP moves small chunks:
 - ▶ 7.5 MB moves are very efficient
 - ▶ New performance profiles are more readily noticed



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

52

FAST VP is dramatically different from other tiering tools.

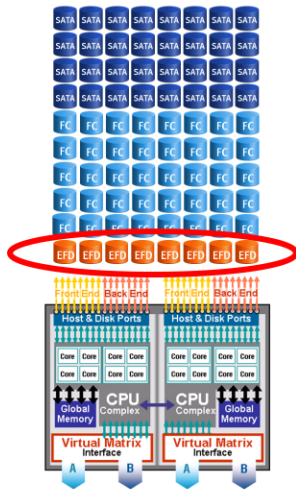
First, FAST VP is more responsive. After the initial review period, there are ongoing reviews to set the thresholds every 10 minutes.

More importantly, the system looks at performance continuously. So when the performance of a sub-LUN data chunk crosses one of the set thresholds, it can be moved immediately – there is no waiting for a new analysis period.

FAST VP moves data easily across 3 tiers of storage. By having EFD, FC, and ATA tiers available, FAST VP can place the right data in the most effective storage class. And by keeping the pools physically separate, different policies can easily be applied at the storage group level, or even down to a single LUN.

FAST VP also moves data in very small chunks. By moving 8 MB (really 7.5, but close enough) chunks of capacity, the system is more agile. High I/O levels are able to be much more isolated, and the system can take action more quickly.

Balancing EFDs across the Backend



- A balanced back-end EFD configuration is a recommended part of a successful FAST VP deployment with even distribution of EFD's across DA's and back-end FC loops.
- Multiples of 8 EFD's per VMAX engine:
 - ▶ i.e. 1, 2, 3 etc EFD's per DA
- All engines to be configured with same number and size of EFD's:
 - ▶ Total of 16, 32, 48 etc EFD's with 2 engine VMAX
 - ▶ Total of 32, 64, 96 etc EFD's with 4 engine VMAX
 - ▶ Total of 64, 128, 192 etc EFD's with 8 engine VMAX
 - ▶ This does NOT include EFD spares !
- Same size of EFD's to be used, assuming EFD's to be configured in a single pool and used by one or more FAST policies.

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

53

A balanced back-end EFD configuration is an essential part of a successful FAST VP deployment with even distribution of EFD's across DA's and back-end FC loops .

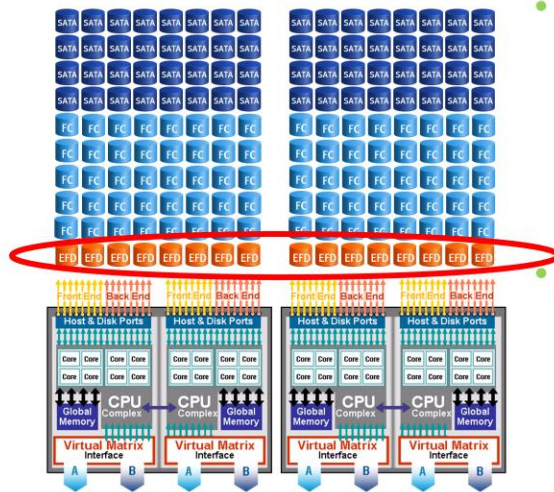
The recommendation is to implement 8 EFDs per VMAX engine, or multiples of 8, across all VMAX engines, as is shown in the diagram highlighted within the red oval.

Below are the recommended number of Flash drives for various size VMAXs:

<u>1 engine</u>	<u>2 engine</u>	<u>4 engine</u>	<u>8 engine</u>
8, 16, 24, 32 ...	16, 32, 48, 64 ...	32, 64, 96, 128 ...	64, 128, 192, 256 ...

The above does NOT includes spare EFDs. The use of the same size EFDs within a single pool is critical to ensure even loading of EFDs and Das. Also as shown a balanced configuration of FC and SATA disks is recommended but is less critical, as each FC and SATA can only place a relatively small load on a DA, whereas a single EFD can place a significant load on the DA.

Example of a Recommended EFD Configuration



- 2 Engine VMAX with 16 EFDs:
 - ▶ Balanced configuration
 - ▶ Even DA Utilization
 - ▶ This number excludes spares
- Optimal configuration is 8 or 16 EFDs per engine:
 - ▶ Spares have to be added

EMC²

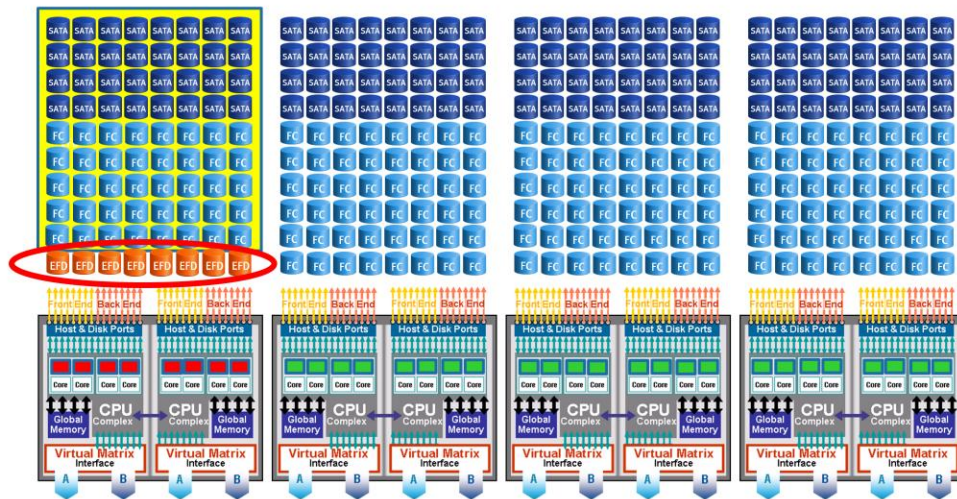
Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

54

To reinforce the point, and for reference, the recommendation for a two engine VMAX is multiples of 16 EFDs, so in this example 16 EFDs is the minimum that is recommended, to provide a good foundation for the deployment of VP and FAST VP on a two engine VMAX, to ensure even utilization of the DAs.

Example of a Poor EFD Configuration – 1



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

55

This example shows a four engine VMAX, which has 32 DAs but only 8 EFDs.

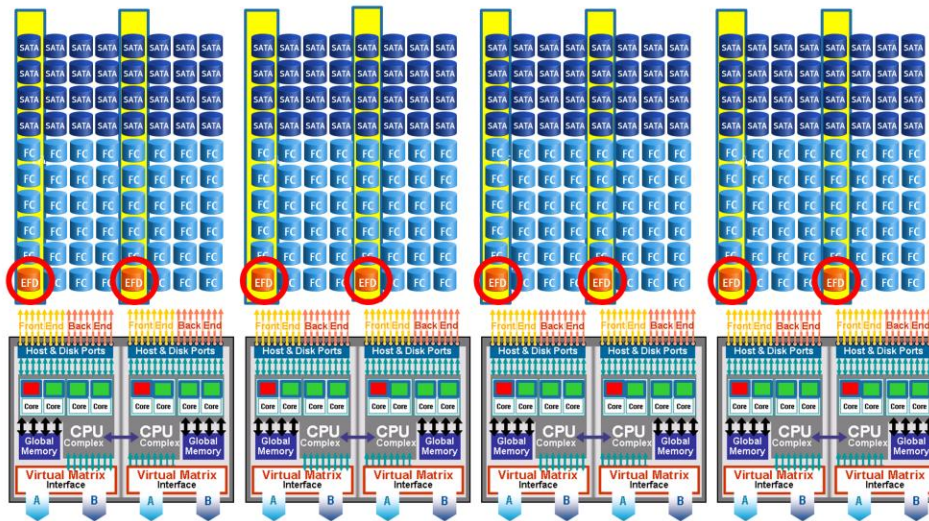
This layout is shown from a graphical perspective for ease of illustration. The following slide shows how the EFDs would be typically laid out in a production deployment.

FAST VP will promote the most active extents from all of the FC and SATA disks to the 8 EFDs.

Based on EMC internal testing and customer testing the 8 EFDs in a 4 engine VMAX, which might represent less than 1% of the total useable capacity, will have a very good chance of being saturated with I/O and FAST VP will be busy promoting and demoting data.

Also the DAs to which the 8 EFDs are configured will have a high utilization, while the remaining 24 DAs will have a low utilization, meaning that the response times will be impacted for the 8 EFDs and FC and SATA disks attached to the same DAs as the EFDs.

Example of a Poor EFD Configuration – 2



EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

56

This example shows an unbalanced EFD configuration on a four engine VMAX, with 32 DAs, but only 8 EFDs.

This example shows a one to one relationship between the DA and the EFD. Other DAs with the same engine or in other engines cannot service any I/O to that EFD, which means that if the EFD is busy, the DA will be busy.

Choose 200 GB EFDs over 400 GB EFDs in Most Cases

- 200 GB EFDs support the same number of IOs/sec as 400 GB EFDs:
 - ▶ Using 200 GBs will double the IOs/sec capacity
 - ▶ Less chance of EFDs causing a performance bottleneck
 - ▶ By using more drives, it is easier to meet the “multiple of 8” rule for configurations using small amounts of EFD storage
- Consider using 400GB EFDs when you are configuring a large percentage of EFD capacity, when the required number of drive slots becomes large, and you are drive slot constrained.
- For existing VMAX without EFDs, all existing and new engines should be configured with EFDs:
 - ▶ Do not just add EFDs to the new engines and leave the existing engines without EFDs



Copyright © 2011 EMC Corporation. All Rights Reserved.

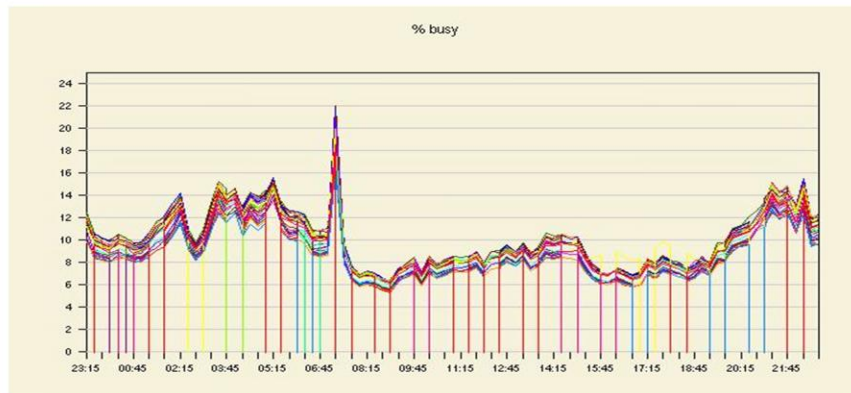
Module 2: Best Practices for Symmetrix VMAX

57

2 x 200 GB EFDs are better than 1 x 400 GB EFD. The former will provide double the potential performance capability of one 400 GB EFD. Both EFDs are able to support the same IOPS, so the 200 GB EFD is able to support double the I/O density, where I/O density is measured in IOPS/GB. Also the 200 GB EFDs will provide better capacity granularity, enabling the requirement to have at least one EFD per DA to be more easily met, due to the high cost of EFDs.

This is not to say that the 400 GB EFDs are bad or that a solution using them will fail. There are several implementations with 400 GB EFDs. But for new VMAX configurations it makes technical and commercial sense to use 200 GB EFDs where possible.

Example of Balanced DA Utilization – Prior to Enabling FAST VP



- This chart shows balanced DA utilization for a 4 Engine VMAX prior to FAST VP been enabled.

EMC²

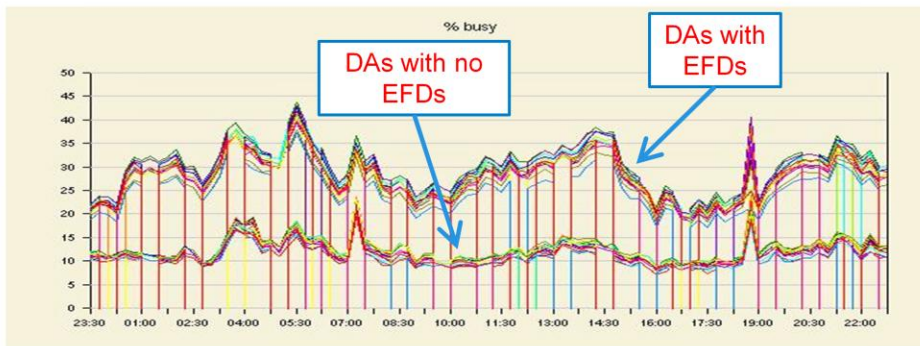
Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

58

For most customers this would have been a 2 Engine VMAX, but the customer deliberately configures and runs at a low average utilization to be able to support workload spikes.

Example of Imbalanced DA Utilization – FAST VP Enabled with DA/EFD Imbalance



- This 4 Engine VMAX had 16 EFDs configured. i.e. 16 DAs had EFDs, 16 DAs had no EFDs. This clearly shows the two bands of DA utilization, with DA utilization now significantly higher for those DAs which have EFDs configured.

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

59

As the DA utilization was previously low prior to enabling FAST VP, the resulting DA utilization imbalance is not severe and is acceptable. However if the DA utilization had been at 50% prior to enabling FAST VP, enabling FAST VP would have created severe DA utilization imbalance, resulting in performance problems.

The I/O load on this specific day was significantly higher than the day before FAST VP was enabled.

In this example 50%+ of the disk I/O is been handled by the EFDs.

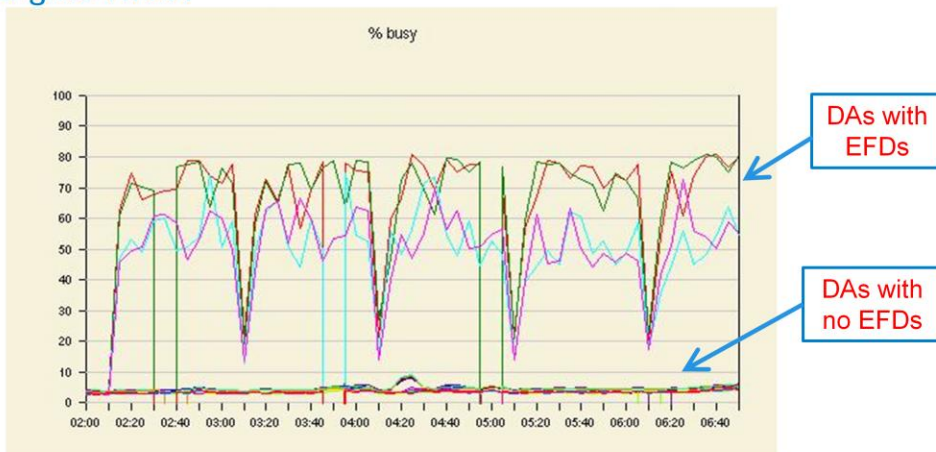
Impact of EFD Imbalance

- Consider 3% of the capacity on EFD doing 50% of the system IO:
 - ▶ A system sized for 50% utilization overall (engines)
 - ▶ Only $\frac{1}{2}$ of the engines have EFDs
 - ▶ Those DAs with EFDs will then have 50% of the workload plus another 25%
 - ▶ So $\frac{1}{2}$ of the DAs will be 75% utilized while the others are 25% utilized
 - ▶ If only $\frac{1}{4}$ of the engines have EFDs the problem is even worse



This slide explains the reason for recommending balanced EFD distribution.

Example of Imbalanced DA Utilization with 4 EFDs in a 2 Engine VMAX



- This example is taken from a customer with a 2 Engine VMAX with 4 EFD testing FAST VP.

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

61

This clearly shows the two bands of DA utilization, with DA utilization peaking up to 80% for the 4 DAs with EFDs, and the 12 DAs without EFDs are at approximately 5% utilization.

Generally Recommended RAID Levels for FAST VP

- EFD Tier RAID 5 (3+1) recommended for EFD Tier:
 - ▶ RAID 1 is better for performance but costs more
 - ▶ Other RAID 5 and RAID 6 levels cause the DAs to get very busy because of the number of backend IOs and parity calculations
- RAID 1 for FC Tier offers the optimal balance between cost, performance and reliability:
 - ▶ No parity calculations needed by the DAs
- RAID 6 (14+2) should be considered as default SATA Tier:
 - ▶ RAID 6(6+2) can be used as it provides better granularity for capacity expansion – in groups of 8 drives at a time
 - ▶ A disadvantage of RAID 6(14+2) is the need to add 16 drives at a time for balanced configurations



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

62

For the EFD Tier: RAID 5 offers a good balance between cost and performance. Although RAID 1 is better and may occasionally be right for 400 GB EFDs, it requires twice the amount of needed capacity, which represents a large expense for this class of storage. RAID 6 for the EFD tier should be avoided unless the customer insists on it and understands its impact. Each host write and FAST VP promotion will create 3 reads and 3 writes, significantly lower the host I/O load the EFD tier can support. Also the load on the DAs will be significantly higher, meaning more VMAX engines will be required to support a given host I/O load.

For the FC Tier: RAID 1 offers the right balance between performance, cost and reliability. Using RAID 1 relieves the DAs of parity calculations and extra IOs during writes. RAID 5(3+1) can also be considered. RAID 6 for the High Performance tier should be avoided unless the customer insists on it and understands its performance and cost impact.

For the SATA Tier: RAID 6 is recommended, because it offers the best reliability. Since the SATA tier is expected to host storage which is rarely accessed, the high demands on the DAs caused by RAID 6 is not expected to create an unduly heavy load. If 14+2 protection is chosen, 16 drives will need to be added to maintain DAs symmetry, whenever storage needs to be added to the SATA tier. With a single engine VMAX RAID 6 (6+2) should be used. In this scenario RAID 6 (14+2) will result in 2 disk members configured on the same DA slice.

Choosing FAST Policies

- A FAST policy associates a storage group with up to 3 tiers.
- Individually, each FAST tier is allowed to have a maximum participation of 100% thus allowing for an aggregate of 300%:
 - ▶ The effect of a 100%+ allocation permits the Symmetrix to optimize data placement dynamically across tiers
 - ▶ While this may be good for application performance it is difficult for planning and charge back
- The percentage of storage specified for each FAST tier in the policy, when aggregated, must total at least 100%:
 - ▶ This means that the allocations within tiers is static and a FAST related promotion must be matched with a corresponding demotion within the storage group
- Use Tier Advisor to select an optimum mix if 100%/100%/100% is not practical.



Copyright © 2011 EMC Corporation. All Rights Reserved.

Module 2: Best Practices for Symmetrix VMAX

63

A FAST policy associates a storage group with up to three tiers. Each tier may have a maximum participation of 100%. At a minimum, the percentage of storage specified for each tier in the policy when aggregated must total at least 100 percent.

If the sum of all tier Allocations in a policy is 300%, then it means that at any given time each associated storage group can be placed entirely on any tier and FAST will locate the devices based on performance criteria. In this case all storage of the tiers is shared dynamically by all storage groups associated with that FAST policy.

While this is good from a performance perspective, it is unpractical for planning purposes, because you don't want to have to provision 300% of the needed storage capacity in your array. In addition, it is difficult for chargeback, because an application can be using different amounts of storage from different tiers at different times.

Consider a different case where a policy is configured with three tiers and the three tiers have a maximum allocation of $i\%$, $j\%$ and $k\%$ respectively. When the sum of the tier allocations ($i\% + j\% + k\%$) is 100%, there can be no flexibility of storage allocations between the different storage groups.

This policy definition restricts FAST movements such that any promotion of a device or a thin device extent must be matched by a corresponding demotion of a device or a thin device extent from the same storage group. This method facilitates chargeback because we can be certain that a storage group always uses the pre-assigned percentages of storage from each tier.

The second one is the process by which Tier Advisor makes its recommendations. Of course the sum of $i\%$, $j\%$ and $k\%$ can be higher than 100 but not as much as 300. This would give the Symmetrix some flexibility while not requiring that 300% of needed storage capacity be provisioned.

FAST VP Control Parameters

- FAST VP Operating Mode:
 - ▶ Automatic: FAST VP moves data during inclusive movement window
 - ▶ Off: No data movements are performed for thin devices
- Workload Analysis Period:
 - ▶ Specifies the maximum amount of time over which long term activity rates will decay
- Initial Analysis Period:
 - ▶ Defines the amount of time FAST VP should wait after a thin device is brought under FAST VP control before performing a data movement
- Pool Reserved Capacity (PRC):
 - ▶ Reserves a percentage of each pool to be excluded from FAST VP data movements
- FAST VP Relocation Rate (FRR):
 - ▶ Determines the “aggressiveness” of the microcode/FAST controller in creating data movement plans for thin device extents



The above are the control parameters used to govern the behavior of FAST VP.

Suggested FAST VP Configuration Parameters – 1

- Initial Analysis Period:
 - ▶ At the initial deployment of FAST VP, set the initial analysis period to at least 24 hours to ensure, at a minimum, that a typical daily workload cycle is seen.
- FAST VP Relocation Rate:
 - ▶ Start conservatively at 7 or 8 and increase gradually to a more aggressive 2 or 3
- Pool Reserved Capacity (PRC):
 - ▶ For individual pools with bound thin devices set the PRC based on the lowest allocation warning level for that thin pool.
 - ▶ For pools with no bound thin devices, set the PRC to 1 percent.



When configuring FAST VP, it is good to use the default performance time window to collect performance data round the clock.

Unlike FAST, sub-LUN data movements do not take as much of the Symmetrix's resources as full device movements. Hence, the data movement window should be open as long as the data collection window.

A default Workload Analysis period of 7 days is recommended.

Suggested FAST VP Configuration Parameters – 2

- Initial Analysis Period:
 - ▶ At the initial deployment of FAST VP, set the initial analysis period to at least 24 hours to ensure, at a minimum, that a typical daily workload cycle is seen.
- FAST VP Relocation Rate:
 - ▶ Start conservatively at 7 or 8 and increase gradually to a more aggressive 2 or 3
- Pool Reserved Capacity (PRC):
 - ▶ For individual pools with bound thin devices set the PRC based on the lowest allocation warning level for that thin pool.
 - ▶ For pools with no bound thin devices, set the PRC to 1 percent.



At the initial deployment of FAST VP, set the initial analysis period to at least 24 hours to ensure, at a minimum, that a typical daily workload cycle is seen. For the initial deployment of FAST VP, start with a conservative value for the relocation rate, perhaps 7 or 8. At a later date the FRR can be gradually lowered to a more aggressive level, possibly 2 or 3.

For individual pools with bound thin devices set the PRC based on the lowest allocation warning level for that thin pool. This means that if warnings start being issued when the pool is 60% full, set the pool reserve capacity to 40%.

For pools with no bound thin devices, set the PRC to 1 percent.

Recommendations for Thin Devices

- Thin device binding:
 - ▶ Bind all thin devices to a pool within the second highest tier
 - ▶▶ It is not recommended to bind thin devices, under FAST VP control, to EFD pools
- Do not Pre-allocate:
 - ▶ Pre-allocated, but unwritten, extents will be demoted to lowest tier



While configuring thin devices, it is best to bind them to an FC pool. Binding them to an EFD tier is cost prohibitive.

System Write Pending Limit Consideration

- FAST VP throttles back promotions and demotions as write pending count approaches System Write Pending limit:
 - ▶ Host I/O takes priority
 - ▶ FAST VP currently stops at 60% of System WP
- In a badly configured Symmetrix from a physical or logical perspective FAST VP will do nothing.
- A very busy workload running on SATA disks, with SATA disks at 100% utilization causing a high % of write pendings, will impact FAST VP from promoting active extents to FC and/or EFD tiers.
- Ensure a busy workload is initially placed on the right tier.



FAST and FAST VP take into consideration the effect that promotion or demotion will have on system performance. Consequently, FAST will be very slow to act if any Symmetrix tier is over-extended.

SRDF/S and SRDF/A and FAST VP

- FAST VP works in isolation at a physical VMAX level, with no coordination of promotions and demotions between the R1's and R2's.
- At the Primary Site FAST VP will promote and demote extents, based on the read and write activity the R1 it experiences.
- At the Secondary Site , FAST VP will promote the extents for R2's which experience write activity on the R1.



FAST VP works in isolation at a physical VMAX level, with no coordination of promotions and demotions for the R1's and R2's. FAST VP will promote and demote extents at the source site based on the read and write activity the R1 experiences. FAST VP will promote and demote extents on the R1 box as normal. The R2 devices typically experience writes directed at the R1 but no reads. At the target site , FAST VP functions totally independently of the Primary Site.

Module 2: Summary

Key points covered in this module:

- Configuring for capacity and performance
- Symmetrix Priority Controls and Dynamic Cache Partitioning
- Virtual Provisioning
- FAST and FAST VP



This module covered best practices for Symmetrix VMAX.

Check Your Knowledge

- What is the recommended resource utilization level when configuring a Symmetrix VMAX for performance?
 - ▶ 25%
 - ▶ 50%
 - ▶ 75%
 - ▶ 100%
- What is the maximum recommended stripe size when considering host-based striping for a Symmetrix VMAX array?
 - ▶ 32 KB
 - ▶ 64 KB
 - ▶ 128 KB
 - ▶ 256 KB
- Which two [2] (select two) tools can be used to manage preferential disk resource allocation for different application workloads?
 - ▶ SPC
 - ▶ DCP
 - ▶ QoS
 - ▶ FRR



Check Your Knowledge

- Why is a balanced EFD configuration recommended when designing FAST VP?
- What is the optimal recommended RAID protection for Fibre Channel Tier when deploying FAST VP?
 - ▶ RAID 1
 - ▶ RAID 5 (7+1)
 - ▶ RAID 6 (6+2)
 - ▶ RAID 5 (3+1)

