Best Practices for SAS® on
EMC® VNX™ Unified Storage

⑤sas.

THE POWER TO KNOW®

Contents

Introduction

The EMC® VNX™ Unified Storage System is a scalable, highly configurable, and easy-to-manage storage subsystem answering the needs of storage performance for small to medium sized SAS® workloads. EMC® VNX™ Unified Storage Systems can be configured to cover a broad range of SAS workload types, from those that can benefit from virtualization, tiered storage, and automation; to those requiring more dedicated resources. The various workloads have specific storage needs and requirements for maximum application performance. This technical paper will outline best practices for architectural setup and tuning to maximize SAS Application performance with EMC® VNX™ Unified storage.

An overview of the storage system will be discussed first, including physical and virtual architecture, pooling and virtualization, storage tiers and management. This will be followed by a list of practical recommendations for implementation with SAS®.

Brief Overview of VNX™ Architecture

In contrast to the previous array we covered, the *EMC® SYMMETRIX VMAX™ (listed below this paragraph), the EMC®VNX™ does not provide a scale-out Virtual Matrix™ architecture to build a large holistic, shared system. Instead, the VNX sports a dual controller architecture. In dual controller systems, one controller is usually active, with the other passive. Servicing the small to mid-range market, the VNX is intended to act as a more local array versus the enterprise serving VMAX.

While still providing enterprise storage performance, the VNX systems tend to have more local roles supporting project-based and segregated-usage storage systems. Instead of a single scale up, "storage as a service" system, the VNX is more scale-out by adding additional separate VNX Systems. VNX still supports many of the popular features introduced in the VMAX, including storage virtualization, pool based storage, automated storage tier management, etc. It also integrates solid-state storage for application usage, and for FAST CACHE® where that is applicable. The EMC white paper EMC Unified VNX Storage covers best practices for VNX storage and is highly recommended for reading - http://www.emc.com/collateral/software/white-papers/h10938-vnx-best-practices-wp.pdf.

*The Best Practices for SAS on EMC SYMMETRIX VMAX Storage covers many of the original and still applicable points around pools, virtualization, tier policies, and promotion - http://support.sas.com/resources/papers/EMC_SAS_VMAX_Best_Practices.pdf.

Recently introduced to the market are new VNX models, dubbed VNX2. There are substantial improvements between VNX and the newer VNX2 models:

- The previous VNX models only utilized Single Layer Cell Enterprise Flash Drives (SLC EFDs), the VNX2 also offers Enterprise Multi-Layer Cell Enterprise Flash Drives (eMLC EFDs). More on this later….

- Controller speed is 4 times faster on the VNX2, utilizing better processors

- Logical Block Assignment moves on tiers were made at 1 GB chunks on VNX, VNX2 supports a reduced chunk size of 256K. This can make tier migrations more efficient and less disruptive.

- The storage processor cache is no longer tunable. Memory allocation amounts between Read and Write, cache page size, and cache watermarks are no longer configurable. The cache is now self-learning and auto-tunes.

- The increased capacity of the VNX2 is optimized for up to 30 GB/sec for large block, sequential, IO. It also now supports up to 32 cores on virtual machines.

- Improved flash drive handling has also been incorporated.

The VNX supported small to medium systems up to 150 drives in capacity. The VNX2 systems support medium systems up to 1500 drives in capacity, and now offer FAST (Fully Automated Storage Tiering), and supports a broader range of protocols such as FC, FCoE, and pNFS.

Physical Storage – Media Types

The underlying physical storage in the VNX™ system consists of 3 basic media types, or storage basic tiers, listed below from highest performing to lowest performing drive technologies in terms of performance:

- Tier 1 – EMC Enterprise Flash Drives(EFD) available as Single Layer Cell EFD (SLC EFD), and enhanced Multi-Layer Cell EFD (eMLC EFD)

- Tier 2 – SAS Drives (SAS 10K and SAS 15K Drives)

- Tier 3 – NL-SAS (SATA Drives Connected via a SAS Interface)

Note: ** Drive capacities may vary by installation. Check with EMC representative for your configuration

These storage tiers provide the least (Tier 3 – NL-SAS) to the highest (Tier 1 - FLASH) performance and cost per Gigabyte. NL-SAS devices are large, slower drives, with much higher response times than SAS or FLASH devices. Read Miss response times are in the 12ms range compared to Flash Drives at 1 ms or 15K FC devices at 6 ms.

Since SAS System Applications have a large-block, sequential orientation, this paper will focus on attributes to support throughput oriented performance.

Flash SLC versus eMLC

What should you use for SAS Workloads? SLC or eMLC Flash? Despite the chart above, SLC EFDs have considerably lower write service times than eMLC EFDs, and have a much longer cell life (*see the link below for further reading). Given SAS propensity for writing large amounts of data, this could be a comparative point for you. However, eMLC EFDs are now utilizing more advanced wear leveling techniques, and employing better strategies for garbage collection and reduced write amplification to give them longer cell life. Their performance has climbed significantly and has greatly reduced the gap between MLC and SLC flash. Still, a large part of the overall performance in a flash storage system lies in the controller speed and handling. So the significant cost difference between SLC EFDs and eMLC EFDS will probably be one of the deciding factors on which of these flash technologies to employ.

If a portion of your workload is Very IO intensive, and requires the fastest possible performance, regardless of price, then consider using SLC EFDs. The use of these more expensive devices may be prohibitive for deploying to your entire SAS data system, and eMLC Flash and SAS spinning drives perform quite admirably for the general workload.

* http://www.computerweekly.com/feature/MLC-vs-SLC-Which-flash-SSD-is-right-for-you

VNX Operating Environment

The VNX Operating Environment can be implemented as a File System or Block-based system. Discussions in this paper will deal only with VNX aspects running under VNX OE Block implementation. Readers are strongly encouraged to read pages 24 – 26 of the Best Practices for SAS on EMC SYMMETRIX VMAX Storage.

VNX Architecture Decisions for SAS®

Many long-term SAS® Customers have migrated through years of storage technology, across increasingly complex storage systems architectures. Over the long haul, there are standard paradigms that have not changed in terms of successful storage architecture for SAS. These paradigms still apply today, regardless of what storage technology you attempt to implement. The following paper details those considerations:
http://support.sas.com/rnd/papers/sgf07/sgf2007-iosubsystem.pdf .

There are two pertinent issues when using thin provisioning to virtualize storage for SAS® applications. The first is under-provisioning on spindle alone; the second is over-subscription of the physical capacity. Thin provisioning is often implemented on much larger, slower drives to gain capacity cheaply. Doing this reduces the actual number of disks needed. Since SAS® is I/O throughput oriented (see the paper noted in the previous paragraph), this can be detrimental. Throughput aggregation is attained by striping across many disks, and aggregating the throughput of each disk in a single "stripe" when reading or writing. When fewer disks are involved, lesser throughput can be attained. This is detrimental to large SAS® workloads. If utilizing thin pools, use wide striping to attain higher throughput.

Another goal of thin-provisioning virtualization typically includes providing thin pools of storage that multiple applications (and hence varying workloads and types) will share. The physical provisioning underneath the thin pools is not sufficient to cover the defined "logical space" on defined LUNs. Hence the term "thin". This under-provisioning of the actual total physical space available, to the sum of the Logical LUN spaces provided is called over-subscription. The paradigm banks on the fact that not all virtual customers will hit the thin pool at once, with a workload that would simultaneously demand all their defined logical LUN extents. Unfortunately that is exactly what SAS® ad hoc workload environments are easily capable of engendering. Granted this can happen in any type of array setup, thin or not, the

thin pool construction typically makes it more vulnerable for SAS if oversubscription occurs. Overhead, safety, and expansion space on the old thick LUN definition has been replaced by a pool that may be quickly become "oversubscribed". Care and attention must be paid to peak load demand, and adequate coverage provided if thin pool construction is chosen. Otherwise, you may wish to follow the route of not using thin-pool provisioning.

Some large SAS® shops, even when provisioning new virtualized storage arrays have chosen to create thin pools that were not over-subscribed, across most or all the spindles in their array. This works well as long as your demand load and performance expected fit within the architecture you create.

Other SAS shops have very heavy IO workloads, with strict performance requirements. Some of these customers have had to bypass some of the benefits of virtualized, thin-provisioned pools, and automated management, in order to guarantee fixed resource performance. In these instances only dedicated LUNs were used, outside of thin pools.

Depending on your shop's storage expertise, it will require working closely with your EMC® Storage Engineer for desired performance.

Both approaches will be covered below in "*Considerations and Best Practices for Using Thin Pools and FAST VP*", and "*Considerations and Best Practices for Using Dedicated LUNs*" below.

Disclaimer - If you are employing a VNX system for SAS, there are the general guidelines and best practices you should consider. These guidelines have been formulated based on experience with a wide variety of SAS customer sites. They are a good starting place, and do not represent a guarantee of performance for your workload, safety and recovery needs, or shop standards and economics. Use prudent judgment and depart from these guidelines where they make sense for your particular situation. There is no one-size-fits-all storage solution for SAS shops.

Given the extreme flexibility, automation, virtualization, and power of the EMC®/VNAX™ system, what are the best practices for storage support on SAS Applications? This next section will give generally recommended practices by SAS® and EMC® resulting from multiple field experiences.

Considerations and Best Practices for Using Thin Pools and FAST VP

As a general starting point, a single large thin pool should be used for all devices to be presented as FAST VP thin pool LUNs to the host server(s). This pool should be created from the SAS tier, and becomes the "bind" pool in the FAST VP process.

- Separate storage groups or management pools, are created out of the same pool of thin LUNs described above. The storage groups are used to apply FAST VP policies to differing workloads within the SAS implementation

- Utilize SAS and EFD Drives for your SASWORK, UTILLOC file systems. Use SAS and EFD drives for your Permanent SAS Data file systems that require good performance.

- Determine what portion of your workload requires the more expensive performance boost of flash. There are to trains of though, SAS Permanent Data versus SASWORK. For SAS user communities employing a typical large read at the beginning of jobs, utilizing SASWORK and UTILLOC in the middle of the job-flow, and writing out a smaller result set, flash heavily favors SAS Permanent Data Storage. Aside from that, SASWORK benefits well since it is heavily utilized. The SAS Utility space for Threaded PROCEDURES benefits from flash performance as well. These file systems typically represent close to a 50/50 read/write ratio, and large-block, sequential flash reads perform better than writes. You have to determine where your economic trade-off points are versus performance gained on allocating flash devices.

- Even on a thin pool, you may wish to dedicate resources on the LUNs to attain performance goals. In other words set the thin pool across enough spindles as to appropriately provision the entire workload underneath it from all the management pools.

- In general utilize eMLC Flash Drives for a good cost-benefit/performance ratio for any file systems residing on the Flash layer. If your workload is extreme, and you can foot the cost of SLC, consider using that instead.

- EFD drives should be targeted for a maximum capacity load of 70% of total space. This provides ample space for quick cell writes when pre-emptive garbage collection isn't effective enough, and helps avoid write cliffs resulting from write-amplification. This "down-formatting" of the file system has proven effective in maintaining effective overall flash write performance under heavy conditions.

- Spread EFD drives across all available busses.

- If you utilize NL-SAS devices at all, do so only for project storage, backups, and work that can be done with low performance and service levels.

- If you plan on using FAST VP for automated tier management in terms of promotion and demotion, the best performance is achieved by the following tier configuration:
  - Tier 1 EFD Raid-5 (4+P)
  - Tier 2 SAS Raid-5(4+1)
  - IF Used at All - Tier 3 NL-SAS Raid-6(8+2P)

- For best performance, consider using only 2 Tier Pools – eMLC and SAS 15K (same drives & sizes for each tier within a single management pool).

- A benefit of using FAST VP on VNX, even if you don't have policies for tier promotion or demotion, is to provide ongoing load balancing of LUNs across the available drives based on slice temperature. Be sure to schedule any planned data-on-tier relocations for off-hours.

- Review your FAST VP policies, and consider manually setting them in a fixed mode. You may need to consider not using FAST promotion or demotion if it has an adverse effect on your IO stream. Many customers keep the promotion or demotion activity as a manual activity (non-automated) for tier promotion.

- If availability is a high concern, consider limiting management pool sizes to the maximum number of drives recommended on initial pool creation tables. Do not expand them above this. This will help limit exposure to the MTBF rate on a very large drive stripe set. Outside availability, a large number of drives in a pool is highly desired because it can yield higher aggregate bandwidth.

- Utilize EMC Powerpath™ to balance LUNs across as many adapters as possible.

- When building a file system on a pool, 100 MB/sec throughput is the recommended minimum bandwidth, with 100 – 150 MB/sec/core as an aggregate for all SAS related file systems on the array.

- Never mix Random and Sequential workloads on the same pool of devices.

- Do not use Block Compression or Block De-duplication for SAS file systems.

- Build a minimum of 2 – 3 thin FAST VP Storage Management Pools. If you wish to use FAST VP consider the following initial pool allocations:
  - A SASWORK Pool for SAS® working space with the following initially recommended FAST Policy tier spreads:
    - 20 – 30 % Tier 1 - EFD Drives
    - 100% - Tier 2 – SAS 15K Drives
    - 0% - Tier 3 – NL-SAS
  - A SASDATA Pool for SAS® Permanent Data space with the following initially recommended FAST Policy tier spreads:
    - 20 – 30 %   Tier 1 – EFD Drives
    - 100% - Tier 2 – SAS 15K Drives
    - 30 - 60%** - Tier 3 – NL-SAS
  - A UTTILOC Pool for SAS Threaded Utility space. This would segregate the Utility space from the SASWORK file space. It would have the same initially recommended FAST Policy tier spreads as SASWORK:
    - 20 – 30 %   Tier 1 – EFD Drives
    - 100% - Tier 2 – SAS 15K Drives
    - 0% - Tier 3 – NL-SAS Drives

- Note that in the above allocations, NL-SAS is not used for SAS Working or Threaded Utility Space. It can be used for SAS® DATA in smaller SAS® workloads that do not have a high throughput requirement (e.g. >75

MB/sec). NL-SAS (SATA) is more appropriate for SAS workloads exhibiting a random access performance profile (e.g. heavy INDEX usage, SAS/OLAP™, and random traversal of data).

Considerations and Best Practices for Using Dedicated LUNs

When IO performance of very large systems becomes paramount, it may be necessary to forsake the benefits of thin pool virtualization, and automated tier management. In this case performance is so impacted by workload, that the storage architecture must be streamlined in every way to support large block, sequential, IO. In this case, the following best practices and considerations are offered:

Create standard thick (classic) LUNs across a large number of RAID Groups, on fixed tiers of storage. Dedicated LUN systems typically are "striped everything". The LUNs are striped across all the disks in the array. Do not make these LUNs part of any thin pool. Treat them as thick LUNs in the same way as EMC Clariion Arrays did.

- Raid levels should be RAID5(4+1) for SAS 15K Drives , RAID5((4+P)for EFD Drives.

- Do not mix drive sizes, types, or rotational speeds in the RAID Groups supporting common LUNs.

- Do not mix random and sequential workloads on LUNs.

- Ensure LUN resources meet minimum throughput requirements for your workload:

  o 100 MB/sec minimum per file system

  o If you are dedicating LUNs, you likely need more throughput per file system.

  o Aggregate throughput from SASWORK and Permanent SAS File systems should be >100 MB/sec per core on your server.

- Do not use NL-SAS Drives for dedicated LUNs

Throughput Testing

It is wise to perform system throughput testing with new configurations. This ensures you have a finite idea where the systems throughput boundaries are. A throughput program provided by SAS® Technical Support for UNIX/LINUX systems is here: http://support.sas.com/kb/51/660.html . The same paper for Microsoft Windows is here: http://support.sas.com/kb/51/659.html

Conclusion

Modern array technology has incorporated virtualization, enabling thin provisioning and offering automated storage tiering. These offerings are intended to combat high array costs by driving high utilization of resources, and automating much of the array Administration. Unfortunately the goal of driving high utilization is often diametric to having sufficient throughput resources quickly available for large-block I/O applications such as SAS®.

When employing such arrays, the advice in this paper, and the specific vendor papers listed in the Further Reading Section below should be carefully considered. In many instances the new technologies can be utilized effectively, and in others it must be mitigated with appropriate architecture changes, and usage.

Work with your Storage Vendor to ensure you are employing their storage technology to its best affect, and providing high performance to your SAS® applications. There is no substitute for careful planning, and testing, to ensure adequate performance will be provided.

References

Deploying EMC VNX Unified Storage Systems For Data Warehouse Applications, Best Practices for Adoption and Deployment, EMC Solutions Group, December 2011.

EMC® Technical Notes – EMC® Symmetrix® V-MAX™ Best Practices. Technical Note P/N 300-009-149 REV A02. November 18, 2009. Copyright © 2009 EMC Corporation All Rights Reserved.

EMC VNX Unified Best Practices for Performance, Applied Best Practices Guide, VNX OE for Block 5.33, EMC Enterprise and Mid-range Systems Division, 2013.

Resources

- this link contains many papers related to performance monitoring, hardware architectures and tuning, and best practices for SAS I/O Workloads.

Contact Information

Your comments and questions are valued and encouraged. Contact the authors at:

Tony Brown
SAS Institute Inc.
15455 N. Dallas Parkway
Dallas, TX 75001
United States
Work Phone: +1(214) 977-3916
Fax: +1 (214) 977-3921
E-mail: tony.brown@sas.com

Margaret Crevar
SAS Institute Inc.
100 SAS Campus Dr
Cary NC 27513-8617
United States
Work Phone: +1 (919) 531-7095
Fax: +1 919 677-4444
E-mail: margaret.crevar@sas.com

Steven Bonuchi
EMC Inc., US Central Division
United States
Work Phone: +1(913) 708-3265
E-mail: steven.bonuchi@emc.com

SAS

THE POWER TO KNOW.

To contact your local SAS office, please visit: sas.com/offices