



Performance and Tuning
Considerations for SAS[®] on
Pure Storage FA-420 Flash
Array



THE POWER TO KNOW_®

Release Information

Content Version: 1.0 August 2014.

Trademarks and Patents

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Statement of Usage

This document is provided for informational purposes. This document may contain approaches, techniques and other information proprietary to SAS.

Contents

- Introduction2
- FA-420 Performance Testing2
- Test Bed Description.....2
- Data and IO Throughput.....2
- Hardware Description3
- Storage3
- Test Results4
- General Considerations and Tuning Recommendations6
 - General Notes.....6
 - Pure Storage and SAS Tuning Recommendations.....7
- Conclusion8
- Resources.....8

Introduction

The Pure Storage FA-420 flash storage device offers high performance, and excellent scalability for SAS® workloads. Pure Storage FA-420 arrays can be configured for large SAS workloads. This technical paper will outline performance test results performed by SAS, and general considerations for setup and tuning to maximize SAS Application performance with Pure Storage FA-420 arrays.

An overview of the flash testing will be discussed first, including the purpose of the testing, a detailed description of the actual test bed and workload, followed by a description of the test hardware. A report on test results will follow, accompanied by a list of tuning recommendations arising from the testing. This will be followed by a general conclusions and a list of practical recommendations for implementation with SAS®.

FA-420 Performance Testing

Performance testing was conducted with the FA-420 to garnish a relative measure of how well it performed with heavy workloads compared with high-performance spinning disk. Of particular interest was whether the flash storage would yield substantial benefits for SAS large-block, sequential IO pattern. In this section of the paper, we will describe the performance tests, the hardware used for testing and comparison, and the test results.

Test Bed Description

The test bed chosen for the flash testing was a mixed analytics SAS workload. This was a scaled workload of computation and IO oriented tests to measure concurrent, mixed job performance.

The actual workload chosen was composed of 19 individual SAS tests: 10 computation, 2 memory, and 7 IO intensive tests. Each test was composed of multiple steps, some relying on existing data stores, with others (primarily computation tests) relying on generated data. The tests were chosen as a matrix of long running and shorter-running tests (ranging in duration from approximately 5 minutes to 1 hour and 53 minutes). In some instances the same test (running against replicated data streams) is run concurrently, and/or back-to-back in a serial fashion, to achieve an average 30 simultaneous test-mix of heavy IO, computation (fed by significant IO in many cases), and Memory stress. In all, to achieve the 30-concurrent test matrix, 101 tests were launched.

Data and IO Throughput

The IO tests input an aggregate of approximately 300 Gigabytes of data, the computation over 120 Gigabytes of data – for a single instance of each test. Much more data is generated as a result of test-step activity, and threaded kernel procedures such as SORT. As stated, some of the same tests run concurrently using different data, and some of the same tests are run back to back, to garnish a total average of 30 tests running concurrently. This raises the total IO throughput of the workload significantly. In its 1 hour and 45 minute span, the workload quickly jumps to 900 MB/sec, climbs steadily to 2.0 GB/s, and achieves a peak of 2.5 GB/s throughput before declining again. This is a good average “SAS Shop” throughput characteristic for a single-instance OS (e.g. non-grid). This throughput is from all three primary SAS file systems: SASDATA, SASWORK, and UTILLOC.

SAS File Systems Utilized

There are 3 primary file systems, all XFS, involved in the flash testing:

- SAS Permanent Data File Space - SASDATA
- SAS Working Data File Space – SASWORK
- SAS Utility Data File Space – UTILLOC

For this workload's code set, data, result space, working and utility space the following space allocations were made:

- SASDATA – 3 Terabytes
- SASWORK – 3 Terabytes
- UTILLOC – 2 Terabytes

This gives you a general “size” of the application's on-storage footprint. It is important to note that throughput, not capacity, is the key factor in configuring storage for SAS performance. Fallow space was left on the file systems to facilitate write performance and avoid write issues due to garbage collection when running short on cell space.

Hardware Description

The host server information the testing was run performed on is as follows:

Host: HP DL980 G7

OS: Linux version RHEL 6.5, Kernel Version 2.6.32-431.el6.x86_64 (mockbuild@x86-023.build.eng.bos.redhat.com) (gcc version 4.4.7 20120313 (Red Hat 4.4.7-4) (GCC)) #1 SMP Sun Nov 10 22:19:54 EST 2013

Memory: 529,223,432 KB RAM,

CPU: 64 x Intel(R) Xeon(R) CPU X7560 @ 2.27GHz GenuineIntel, Model 46, CPU Family 6, Stepping 6, 2266 Mhz, 24576 KB Cache

Storage

Comparative performance testing was conducted between a high-performance spinning disk array, and the Pure Storage FA-420 . A comparison of the two arrays follows:

High-performance Spinning Disk Array Definition:

- Number and types of disks / controllers: 192 x 300GB 15K rpm FC Drives and 4 controllers, consuming 84u all together
- Raid levels: Multiple raid5 sets (3data/1parity disk per set)
- File System Type: XFS
- File System/Logical Volume Arrangement: File Systems /SASDATA, SASWORK, /UTILLOC are placed across 1 Logical Volume utilizing all 192 spindles in the array
- Fibre Channel ports: 8 x 8 Gb Adapters with ACTIVE/ACTIVE multi-pathing
- 8 GB/sec potential throughput for the array

The Pure Storage FA-420 Definition:

- Number and types of drives / controllers: 48 x 256GB SSDs and 2 controllers, consuming 8u all together
 - 2 – 2u controllers with 2 – 8 Gb/sec FC adapters, yielding 16 Gb/sec throughput per controller, 32 Gb/sec per Array
 - 2 – 2u drive shelves consisting of 11 Terabytes flash storage (2 x 5.5 TB Shelves).
- The FA-420 Array was configured into 8 luns for the /SASDATA filesystem and 8 luns for the /SASWORK file system (/UTILLOC was placed under /SASWORK)
- The file systems were striped across all of their assigned luns using lvm striping, “lvmcreate –i8”
- File System Type: XFS
- 4 GB/sec potential throughput for the array

Test Results

The Mixed Analytic Workload was run in a quiet setting (no competing activity on server or storage) for both the high-performance spinning disk storage and the Pure Storage FA-420 array. Multiple runs were committed to standardize results.

We worked with Pure Storage engineers to tune the host and IO pacing (see General Considerations and Tuning Recommendations below). The Pure firmware level used in testing was 4.0.6, the minimum level customers are recommended to use (in order to leverage automatic governing of the in-line de-duplication process). Pure Storage devices utilize multiple inline data reduction technologies including de-duplication and compression. These technologies, in combination, yielded an overall 16.9 : 1 data reduction ratio on the test data set. Depending on a customer's specific SAS data characteristics, experienced data reduction ratios will vary.

The tuning options noted below apply to LINUX operating systems for Red Hat RHEL 6.5, work with your Pure vendor for appropriate tuning mechanisms for any different OS, or the particular processors used in your system.

Table 2 below shows the performance difference between the completely tuned Pure Storage FA-420 system and the high-performance spinning disk system. This table shows an aggregate SAS FULLSTIMER Real Time, summed of all the 101 tests submitted. It also shows Summed Memory utilization, Summed User CPU Time, and Summed System CPU Time in Minutes.

Storage System	Elapsed Real Time in Minutes - Workload Aggregate	Memory Used in GB - Workload Aggregate	User CPU Time in Minutes - Workload Aggregate	System CPU Time in Minutes - Workload Aggregate
High-Performance Spinning Storage	1482.10	56.07	1750.63	282.36
8U Pure Storage FA-420	1462.15	56.08	1730.53	274.70
Aggregate DELTA	-19.95	+0.01	-20.1	-7.66

Table 2. Total Workload Elapsed Time, Memory, and User & CPU Time Reduction by using Pure Storage FA-420 Array

The first column in shows the total elapsed run time in minutes, summed from each of the jobs in the workload. It can be seen that the PURE array reduced the aggregate run time of the workload by approximately 20 minutes.

Another way to review the results is to look at the ratio of total CPU time (User + System CPU) against the total Real time. Table 3 below shows the ratio of total CPU Time to Real time. If the ratio is less than 1, then the CPU is spending time waiting on resources, usually IO. The high-performance spinning storage array is a very large, expensive, and fast array. The Real time/CPU time ratio of the workload on the Spinning Storage array was 1.157, which is excellent.

The Pure Storage FA-420 Array achieved approximately the same performance at 1.155 with an 8 rack unit (ru) single array. Both of these numbers indicate the storage was keeping up with the CPU demand and provided excellent performance. The standard deviation and spread of the ratios for spinning storage is very slightly higher than the Pure test.

The 1.155 Mean Ratio associated with the Pure Storage FA-420 Array indicates a much more CPU bound process (e.g. not spending time waiting on IO). For the IO intensive SAS Application set, this is the goal you wish to achieve! The question arises, "How can I get above a ratio of 1.0?" Because some SAS procedures are threaded, you can actually use more CPU Cycles than wall-clock, or Real time.

Storage System	Mean Ratio of CPU/Real-time - Ratio	Mean Ratio of CPU/Real-time - Standard Deviation	Mean Ratio of CPU/Real-time - Range of Values
High-Performance Spinning Storage	1.157	0.369	2.107
Pure Storage FA-420	1.155	0.362	2.039

Table 3. Frequency Mean, Standard Deviation, and Range of Values for CPU/Real Time Ratios. Less than 1 indicates IO inefficiency.

In short our test results were very pleasing. It showed that the Pure FA-420 Array, when tuned at install with recommended parameters, and utilizing firmware level 4.0.6 or greater, can provide significantly good performance for SAS workloads, in this case as good as a large, high throughput spinning array. The workload utilized was a mixed representation of what an average SAS shop may be executing at any given time. Due to workload differences your mileage may vary.

General Considerations and Tuning Recommendations

General Notes

Utilizing the Pure Storage FA-420 Flash Array can deliver significant performance for an intensive SAS IO workload. Work with your Pure Engineer on install tuning, and host CSTATE tuning (if your IT standards allow CSTATES to be altered) to maximize performance. Processor CSTATES recommendations are typically different for different processor types, and even models. They govern power states for the processor. By not allowing processors to enter a deep “sleep” state, they are more rapidly available to respond quickly to demand loads. Generally Keeping CSTATES at a 1 or 0 level works well, and for those processors supporting turbo-boost mode, it should be set to take advantage of that.

It is very helpful to utilize the SAS tuning guides for your operating system host to optimize server-side performance before FA-420 tuning, and additional host tuning is performed as noted below. See:

<http://support.sas.com/kb/42/197.html>

Some general considerations for using flash storage include leaving overhead in the flash devices, and considering which workloads if not all, to use flash for.

Reads vs. Writes. Flash devices perform much better with Reads than Writes for large-block, sequential IO (SAS). If your scale of workload dictates that you can afford flash for all your file systems that is good. If not, you may wish to bias your flash usage to file systems and data that are read-intensive to get the maximum performance for the dollar. For example, if you have a large repository that gets updated nightly, or weekly, and is queried and extracted from at a high level by users, that may be where you wish to provision your flash storage.

Pure Storage and SAS Tuning Recommendations

The following install tuning was performed on the host server during the installation of the FA-420 . These are Pure Storage and SAS recommended tuning steps.

- The following kernel boot options were used preventing cstates from going above 1:
 - intel_idle.max_cstate=0
 - processor.max_cstate=1
- Tuned Storage Profile – Enterprise Storage Profile
- Firmware Level Utilized was 4.0.6
-
- The array consisted of 4 x 8 Gb connections direct attached to HPDL980 utilizing native LINUX Multi-pathing.
- Eight luns were created each for /saswork and /sasdata file systems. The file systems were striped across all LUNs with lvm using “lvmcreate –i8”, formatted with XFS.
- The file systems were mounted with noatime, nodiratime
- We created /etc/udev/rules.d/99-pure-storage.rules containing:

```
#use noop scheduler for high-performance solid-state storage
ACTION=="add|change", SUBSYSTEM=="block",
ENV{ID_VENDOR}=="PURE", ATTR{queue/scheduler}="noop"
#Reduce CPU Overhead due to entropy collection
ACTION=="add|change", SUBSYSTEM=="block",
ENV{ID_VENDOR}=="PURE", ATTR{queue/add_random}="0"
#Schedule I/O on the core that initiated the process
ACTION=="add|change", SUBSYSTEM=="block",
ENV{ID_VENDOR}=="PURE", ATTR{queue/rq_affinity}="2"
#after creating above udev files reboot or run "udevadm contron – reload-rules ;
udevadm trigger"
We appended following lines to /etc/rc.local:
blockdev –setra 16384 /dev/mapper/vg_pure_saswork-lv_saswork
blockdev –setra 16384 /dev/mapper/vg_pure_sasdata-lv_sasdata
echo never > /sys/kernel/mm/redhat_transparent_hugepage/enabled
```
- We added following device section to /etc/multipath.conf:

```
device {
    vendor "PURE"
    user_friendly_names yes
    path_grouping_policy multibus
    path_checker tur
    rr_min_io 1
    path_selector "round-robin 0"
    no_path_retry 0
    fast_io_fail_tmo 3
    dev_loss_tmo 30
}
```
- We set fc hba queue depths to 128 in modprobe files:

```
#qllogic /etc/modprobe.d/qla2xxx.conf:
Options qla2xxx ql2xmaxdeth=128
#emulex /etc/modprobe.d/lpfc.conf:
Options lpfc lpfc_devloss_tmo=10
Options lpfc lpfc_hba_queue_depth=128
```

Conclusion

The Pure Storage FA-420 array can be extremely beneficial for many SAS Workloads. Testing has shown it can significantly eliminate application IO latency, providing improved performance. It is crucial to work with your Pure Storage engineer to plan, install, and tune your host utilizing the FA-420 array to get maximum performance. The guidelines listed in this paper are beneficial and recommended. Your individual experience may require additional guidance by Pure Storage depending on your host system, and workload characteristics.

Resources

SAS Papers on Performance Best Practices and Tuning Guides: <http://support.sas.com/kb/42/197.html>

Contact Information:

Name: Geoff Hough, Director, Strategic Alliances
Enterprise: Pure Storage
Address: 650 Castro Street, Suite 200
City, State: Mountain View, CA
United States
Work Phone: +1 (800) 379-7973
Email: geoff@purestorage.com

Name: Radha Manga, Director, Customer Solutions Group
Enterprise: Pure Storage
Address: 650 Castro Street, Suite 200
City, State: Mountain View, CA
United States
Work Phone: +1 (408) 857-8901
Email: radha@purestorage.com

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

Name: Margaret Crevar
Enterprise: SAS Institute Inc.
Address: 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



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

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2014, SAS Institute Inc. All rights reserved.
