LOAD DYNAMIX

White Paper

Everything you need to know about flash storage performance

The unique characteristics of flash make performance validation testing immensely challenging and critically important; follow these best practices

May 2015

Table of Contents

Introduction

All flash storage arrays, aka AFAs, are receiving lots of attention these days and for good reason. Compared to spinning disks, all flash arrays provide dramatically better performance, take up less floor space, and even offer overall cost of ownership advantages. Nevertheless, the cost per GB is still relatively expensive, which means that all flash arrays should not be deployed for every application workload. The obvious questions are which workloads are most cost-effectively deployed on flash storage and which vendor has the optimal flash storage product for your workloads?

We at Load DynamiX, a leading provider of storage performance validation solutions, have invested a great deal of time and energy on these questions. Our focus has been to develop advanced workload modeling and load generation solutions for both storage technology vendors and IT organizations.

Load DynamiX combines an intuitive storage workload modeling application – Load DynamiX Enterprise – with a purpose-built load generation appliance. The solution generates massive, highly realistic loads that stress networked storage infrastructure to its limits and beyond, helping storage architects and engineers fully understand storage system behavior and performance characteristics before purchase and deployment decisions are made.

Based on deep experiences with global 2000 companies, and in collaboration with flash storage visionaries within the industry (both leading vendors and well-known analysts), Load Dynamix has created an all flash array performance validation methodology. Below, I'll share some of the foundational aspects of Load DynamiX's flash performance testing and describe two specific methodologies for understanding the performance of all flash arrays. A more detailed all-flash array testing methodology can be found at the Load DynamiX website.

All flash array performance testing is important if you have these kinds of questions:

- Can I improve application performance with flash? If so, by how much?
- Can I afford the performance improvement? Will dedupe / compression reduce the effective $/GB without substantially impacting performance?
- How do I select the best vendor or product?
- Which of my workloads will run best on all flash arrays?
- How can I optimize all flash storage configurations?
- How much does my performance degrade with dedupe, compression, snapshots, etc.?
- Where are the performance limits of potential configurations?
- How will flash storage behave when it reaches its performance limits?
- Does flash performance degrade over time?
- For which workloads should I use an all flash array or a hybrid flash array?

While everyone agrees that the most accurate way to test performance is in a production environment, it's simply not possible. The next best thing is a realistic, scalable test in a lab environment. Storage engineers have had decades to refine HDD-based array testing, but only a short time to learn about flash storage.

How Flash Storage is Different

All flash arrays differ from traditional spinning disk arrays in behavior, performance, and often durability. For example, SSDs write and read in blocks, and they are limited in the number of writes that can be performed on a particular block. Sophisticated data reduction and array-wide wear leveling techniques can dramatically increase SSD durability, in some cases beyond the expected life of a spinning disk drive. While the performance advantages of all flash arrays are well documented, much effort has gone into the design of modern all flash arrays to attack the price premium of flash over disk. Sophisticated efficiency techniques promise, for a few workloads (such as full clone VDI), to bring the effective cost of all flash arrays within striking distance of disk arrays, and in some cases below the effective cost of an all-disk implementation.

Below are some of the biggest ways flash arrays differ from disk arrays and how they affect performance testing and evaluation.

Data compression and data deduplication.

These data reduction techniques reduce both data storage footprints and transmission loads (bandwidth requirements). But because deduped and compressed data must be decompressed to use, it imposes additional computational costs and can therefore have a significant impact on application performance. Algorithms can vary greatly, and their differences can significantly affect performance. Because the economic payoff of flash may heavily rely on reduced storage capacity requirements, and because different vendors handle data reduction techniques differently, the performance of a given all flash array may differ widely depending on data type. Your test method and load generator must be able to be extremely configurable for compression and dedupe.

Metadata.

A great deal of the internal management of flash-based arrays is meant to optimize the performance and reliability of the media. Array performance and scale is greatly affected by where metadata is stored and how it is used. This is a big reason to precondition a flash array properly (i.e., write to each flash cell) before testing, to avoid artificially fast read results.

Workload profiles and scale.

Hard disk arrays are capable of IOPS in the range of many thousands. Flash-based arrays can support IOPS in the hundreds of thousands. Workload profiles for which flash-based arrays are generally deployed are very different from the classic workloads of the past. The mixed virtualized workloads for which flash-based arrays can be deployed exhibit much more variability than traditional workloads. They include both extremely random and sequential data streams; a wide mix of block sizes and read/write ratios; compressible, dedupable, and non-compressible/dedupable blocks; and hot spots. To test flash-based arrays to performance saturation points, you must be able to generate workloads rarely if ever seen on disk-based systems. And you must be able to reproduce the right I/O and data profile at that scale. Your load generator must be both powerful and flexible.

Overprovisioning.

To improve the probability that a write operation arriving from the host has immediate access to a pre-erased block, most but not all flash products contain extra capacity. Overprovisioning is common because it can help flash designers mitigate various performance challenges that result from garbage collection and wear leveling, among other flash management activities. It also increases the longevity of flash arrays. You should test at near the maximum usable capacity recommended by the all flash array vendor to assess the performance benefit of overprovisioning. Typical recommendations are 90, 95, and 99 percent of capacity.

Hotspots.

Most real-world workloads exhibit hotspots (i.e. the characteristics of temporal and spatial locality). Garbage collection, which proactively eliminates the need for whole block erasures prior to every write operation, may exacerbate hotspots. Methodologies differ. . And testing hotspots is advised, but importance may vary by array vendor.

Protocols.

You may have to throw out some preconceptions learned from decades of HDD system testing. Storage protocols often create quite different performance levels with flash. Factors such as block sizes and error correction overhead can make a big difference in throughput and IOPS. You should test all of your file and block protocols, because the rules have changed.

Software services.

Replication, snapshots, clones, and thin provisioning can be very useful for improving utilization, recovery options, failover, provisioning, and disaster recovery. However, implementation may have big performance impacts and must be accounted for in the testing methodology. Their effects may be different than what you find in HDD systems. It's important to run workloads on newly created clones, and not simply create clones while workloads are present.

QoS at scale.

Quality of Service affects both infrastructure and application performance. Build and run your tests with QoS configured for how you plan to use it. As your load increases, measure the ability to deliver expected performance in mixed workload environments.

Effective cost of storage. Looking at just cost per gigabyte ($/GB) is not a good way to compare storage costs. A good question to ask is, how much is usable? Arrays vary widely in their conversion from raw storage to usable storage. For instance, due to the inherent speed of flash, you can effectively use deduplication and compression to fit substantially more data on a given amount of raw storage. Also, it's common to have to overprovision HDD storage aggressively to get the number of spindles necessary to deliver the performance required (a strategy called "short stroking"). Further, disk arrays often have to make extensive use of really expensive cache memory in order to achieve performance SLAs. Finally, you must consider factors like power and space requirements. Flash typically takes a fraction of the power and space of a traditional HDD-based array. Of course, you need to ensure that your data reduction assumptions are realistic. Talk with your application vendors and storage vendors. Storage vendors have storage efficiency estimation tools that will give you an accurate idea of what to expect from their particular storage platforms. If you want to get a feel for how compressible your files will be, zip them and compare with unzipped sizes.

Storage engineers and architects considering all flash arrays for their workloads must explore the behavior of these products, and as far as possible, assess their performance in the context of their expected workloads. With

a robust validation process in place, storage engineers and architects can select and configure flash storage solutions for their workloads with a clear idea of their impact on both performance and cost in production.

Performance profiling and workload modeling

To tune its performance validation methodology for all flash arrays, Load Dynamix recommends paying specific attention to the following three areas:

1. Specific pre-conditioning of the array to create a state that has characteristics similar to an aged flash storage array, prior to applying load.
2. Stressing of specific all flash array behaviors, such as data reduction techniques, clones, snapshots, failover, replication, backups, and other enterprise features that affect performance and cost.
3. Stressing the array with realistic emulations of typical supported workloads.

There are two primary methodologies for storage performance validation: performance profiling and workload modeling.

Performance profiling is sometimes called “performance corners testing” or “multi-dimensional benchmarking.” It provides a very useful outline of the workload-to-performance relationship, and in some cases is sufficient to support the engineer’s decisions. The objective of intelligent performance profiling is to characterize the behavior of a storage system under a large set of workload conditions. Doing so provides the storage engineer with a map of the behavior of the storage system, making it easy to understand where sweet spots or bottlenecks may be, or which workload attributes most directly affect the performance of the system. Engineers can then use this information to optimally match their workloads to storage systems.

This methodology is characterized by an iteration workflow that allows the user to iterate on any of the many workload characterization attributes (load profile, block size, command mix, etc.) to stress the storage system under dozens, hundreds, or even thousands of workload configurations, with automated test execution, aggregation of data, and presentation of results. This can be accomplished with custom scripting, or with off-the-shelf test products like Load DynamiX Enterprise. For example, Figure 1 below shows the input screen of the Iterator function in Load DynamiX Enterprise. It’s configured to run 18 sequential tests, without scripting, to test the effect of compression ratios, number of workers, and block size on three KPIs.

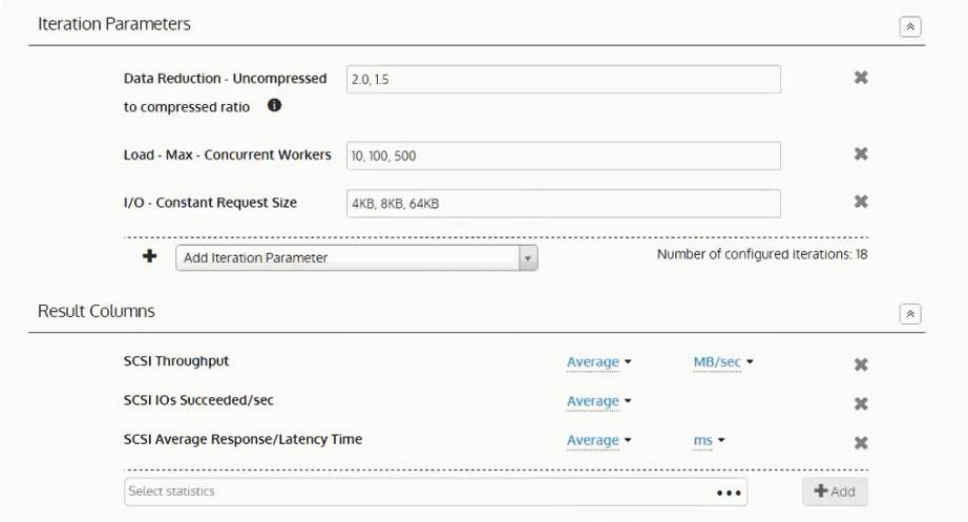

Figure 1: Load DynamiX Enterprise input screen, demonstrating how to profile AFA performance by running 18 tests with an automated, off-the-shelf approach.

In Figure 2 below, we see the results. In this figure, we've sorted on the IOPS column to find the configuration that results in the greatest IOPS (approximately 22,014). Sorting by latency would quickly show figures exceeding 6ms for 500 or more concurrent workers.

Iteration Results Export to CSV

#	Status	Duration	Data Reduction - Uncompressed to compressed ratio	Load - Max - Concurrent Workers	I/O - Constant Request Size	SCSI Throughput (average)	SCSI IOs Succeeded/sec (average)	SCSI Average Response/Latency Time (average)
12	Finished	01:31	1.5	10	64KB	1,376.1 MB/sec	22014.239	3.8 ms
15	Finished	01:31	1.5	100	64KB	1,399.4 MB/sec	22372.652	40.0 ms
18	Finished	01:32	1.5	500	64KB	1,408.9 MB/sec	22551.616	44.8 ms
9	Finished	01:31	2	500	64KB	1,682.8 MB/sec	26904.297	37.5 ms
6	Finished	01:31	2	100	64KB	1,800.0 MB/sec	28800.75	31.3 ms
3	Finished	01:31	2	10	64KB	1,885.5 MB/sec	30162.428	3.2 ms
10	Finished	01:31	1.5	10	4KB	365.8 MB/sec	93275.342	1.0 ms
11	Finished	01:31	1.5	10	8KB	759.9 MB/sec	97084.138	1.0 ms
1	Finished	01:31	2	10	4KB	413.3 MB/sec	105396.173	0.8 ms
16	Finished	01:31	1.5	500	4KB	463.3 MB/sec	118103.395	8.6 ms
13	Finished	01:31	1.5	100	4KB	482.6 MB/sec	123085.863	6.7 ms
2	Finished	01:31	2	10	8KB	1,021.5 MB/sec	130491.623	0.7 ms
7	Finished	01:31	2	500	4KB	539.1 MB/sec	137498.286	7.3 ms
17	Finished	01:31	1.5	500	8KB	1,088.3 MB/sec	139044.464	7.3 ms
14	Finished	01:31	1.5	100	8KB	1,105.7 MB/sec	141265.633	6.0 ms
4	Finished	01:31	2	100	4KB	563.4 MB/sec	143688.891	5.5 ms
8	Finished	01:31	2	500	8KB	1,277.2 MB/sec	163183.3	6.2 ms
5	Finished	01:31	2	100	8KB	1,310.9 MB/sec	167472.132	4.9 ms

Figure 2: Extract of output report of Iterator function, showing effect of changing parameters on performance.

Workload modeling goes to a greater level of detail. Whereas the goal of performance profiling is to test under a wide range of workload conditions, the objective of workload modeling is to stress the storage system under a realistic simulation of the workloads it will actually be supporting in production. Workload modeling requires a prerequisite knowledge of the characterization of the workloads, usually based on the storage engineer's knowledge of the application and data typically provided by storage monitoring utilities.

The workload models should allow users to characterize access patterns with as much detail as needed. For example, block size can be represented as a realistic distribution of values, not just a single value. Different workloads should also be combined into a single "composite workload" that stresses different areas of the storage system. Testing a storage system with a workload simulation that is sufficiently realistic allows storage engineers to develop a great deal of confidence in their decisions about product selection and configuration prior to deployment in production.

These two methodologies represent the core of performance validation. They can be used to support several typical storage testing approaches. These include limits finding, or determining the workload conditions that drive performance below minimal thresholds, and the documenting of storage behavior at failure point; functional testing, the investigation under simulated load of various functions of the storage system (backup, replication, etc.); error injection, the investigation under simulated load of specific failure scenarios; and soak testing, the observation of the storage system under load sustained over significant time (e.g., two days, one week). In short, the performance profiling and workload modeling test methodologies can be used to answer virtually any question one might think to ask about an all flash array.

Flash-based storage arrays offer some tremendous advantages over disk-based arrays, but their fundamental differences make the storage buying decision more complex than ever. By following testing best practices in the lab, including performance profiling and workload modeling, all flash array product selection and configuration become a mathematical exercise, not a guessing game.