Storage Switzerland, LLC

Flash, It's not just for Databases Anymore

Prepared by: George Crump, Lead Analyst
Prepared: August 2016

A memory-based storage system, or what we today call an all-flash storage system, is a tool IT has used in the past to accelerate database environments. The small I/O, high transaction nature of these environments are tailor-made for the attributes of flash. However, the substantial price decline of all-flash storage systems and the acceptance of data efficiency technologies has made mid-level workloads, like virtualization, ideal for all-flash storage systems. But recently, thanks to the continuing decline in flash costs, along with the increasing density of flash media, flash delivers real value for file or object-based workloads as well.

The State of File Based Data

File – and especially object storage – systems are often thought of as the digital dumping ground of the data center. But the use case for file-based storage systems has changed dramatically in recent years thanks to the phenomenon of Big Data. Many organizations have projects like Hadoop, SAS, Spark, or Splunk underway to analyze the data that Internet of Things (IoT) devices create and feed back into the organization. In addition, some organizations have the need to provide media streaming to employees or customers. Finally, as flash begins to dominate the traditional production application set, these organizations are seeing the need to provide a data protection strategy that is faster than today's hard disk or tape based configurations.

Traditionally, these secondary storage systems have been hard disk-based, but the time has come to consider moving them to all-flash storage, too.

The Case for All-Flash Secondary Storage

Secondary storage systems are designed to store primarily file-based workloads represented by big data initiatives, DevOps projects and traditional file storage. It is important to understand that secondary storage does not mean "not important." Secondary storage is vital to most organizations, but it does not need the same levels of performance and availability that primary storage workloads require. But, as the use cases for file-based storage keeps increasing, so, too, does the demand for more and more performance.

There are two primary reasons to consider flash for these workloads that traditionally had data reside on secondary storage. The first is that now, these workloads have a legitimate need for the performance inherent with a flash-based storage system.

The benefit of high performance is immediately obvious in environments like Hadoop, Spark and Splunk, where faster and more frequent data analysis leads to better decision-making. Another key driver for high performance secondary storage is personalization. Organizations are learning that if the application their customers use, customizes the experience to present the user with content they are most interested in, there is an increase in customer engagement.

So, while much of the benefit of a flash-based system is related to its reduction in latency and extremely high IOPs, a flash system that serves the secondary storage market should also deliver incredible bandwidth, which is valuable in bulk load situations or for data protection.

The second reason to consider flash is density. Flash-based storage with the correct design, even without data efficiency, can deliver far more capacity per rack unit than the typical hard disk-based storage system. While the cost of that system may be more, when factoring the cost of data center floor space, lower power requirements, savings in cooling, and the reduction in operational manpower and OPEX , the flash storage system - especially if it is particularly dense - is less expensive than a hard disk-based big data focused system.

The Requirements of an All-Flash Secondary Storage System

Extreme Density

The primary requirement of a secondary storage all-flash array is it needs to be extremely dense. IT professionals should expect 1/2 a petabyte of capacity in 3U. This is because flash modules can be very tightly packed into a storage enclosure, unlike hard disk systems which have to worry about heat and vibration caused by dozens of spinning drives. In addition, the flash form factor is smaller and does not need to be surrounded by the mechanics of the hard disk drive. Densities of these proportions should also lower the cost of powering and cooling the secondary storage tier and usually positively impact the cost of the associated server farm.

Reliable and Protected

A secondary all-flash system does not need to meet the same performance and high availability requirements of an all-flash array. But it does need to provide improved performance over a traditional hard disk drive based secondary storage system. At the same time, while it does not need to meet the 24x7x365 of a primary storage all-flash array, it does need to be very reliable and provide high data integrity.

Software Neutral

Unlike primary storage workloads where the general type of application is known, secondary storage systems need to support a variety of workloads and applications. Most of the common applications that will access data on the secondary storage system like Spark, Splunk, Hadoop, Couchbase and Cassandra were not even around five years ago. It is reasonable to expect more innovation in these platforms and to expect more new applications in the coming years.

As a result, the storage hardware should be software neutral, meaning that a variety of storage software and file systems can take advantage of it. The storage software needs to adapt so it can present both legacy protocols like NFS and emerging protocols like the Hadoop File System (HDFS) and Object Storage.

The software should also offload some of the data protection and data integrity responsibilities from the storage system by being able to ensure copies of data survive a component failure and to ensure that the data written today reads the same potentially years later.

Cost Effective

Finally, the reality is this is the secondary storage tier and, while performance is becoming more important for this tier in recent years, it is being compared to hard disk based storage. Thanks to the savings in density and power costs, as well as the significant performance increase, these systems do not need to beat the price of disk. But they do need to come close. IT professionals should expect to purchase these systems for less than $1 per GB.

It is possible to get further cost savings by not burdening the systems with "included" software. Remember, the secondary data market is so fluid that any software included might be unusable in a couple of years. Storage vendors should focus instead on providing high quality, high dense storage systems designed to complement the software, not compete with it.

Conclusion

The time has come for secondary storage systems to move to all-flash. Secondary storage systems benefit from the performance of flash as much as the traditional production applications can. The ability to instantly analyze data provides organizations with tremendous benefits. But this new flash platform has to be different. It has to be denser than its primary storage counterparts. Density not only reduces data center floor space and power costs, it also provides a much richer examination of data since a larger cross section can be analyzed in less time than if it was on hard disk storage. Finally, the secondary tier is still an evolving entity. It has to allow for the sending of source data from humans and machines. A software driven approach is ideal for this platform.