STORAGE SWITZERLAND

OVERCOMING CONDITIONAL SSD PERFORMANCE

Storage Switzerland, LLC

George Crump, Senior Analyst

Flash based solid state disk (SSD) improves storage I/O response times. However, the challenge with memory-based storage, especially PCIe devices, is that most solid state products require that the right conditions be in place or be "arranged" so that maximum performance can be obtained. The management of the NAND flash itself is often the cause of most of this performance variability and it is up to the flash vendors to provide systems that overcome these challenges.

What causes Performance Conditions?

Understanding the differences in write patterns is important because flash is a write limited technology, meaning that it can only sustain a fixed number of write cycles. After the flash drive has been completely filled, each new write operation requires that a prior block be erased before the new data is placed on it, which also slows down write performance...significantly.

To get around this, most vendors have developed a 'garbage collection' routine that will pre-clear old data areas so that when new writes come in they don't have to wait for data to be erased. The problem is that this requires the attention of the flash controller and may hurt performance especially when high storage I/O activity is occurring or when the drive is nearly full.

This significant gap in write performance vs. read performance, (reads are roughly 10X faster) creates a second problem. High write activity may block read requests. This means that writes need to be managed so that they don't impact read response times.

These reasons make the design of the flash controller and how that controller technology is applied to the flash system the most important elements in determining consistent performance, no matter the condition of the solid state media. The flash controller has to expand beyond being a component that simply maintains an even wear of the NAND flash cells. The controller also needs to make sure that performance stays consistent over the entire life of the device.

Where to Develop the Controller

Development of the controller becomes a key issue for solid state providers. There are some that have designed and manufactured their own silicon, which takes a significant investment in capital and inventory. The changing nature of the flash controller makes design flexibility critical. Silicon vendors that don't have the volume of sales may not be able to cycle through their inventory to keep up with improvements in controller code. In other words creating specific silicon to solve problems in a dynamically evolving market like flash storage can impact development flexibility.

Field Programmable Gate Arrays (FPGAs) are another option and essentially programmable silicon. They allow for greater programing flexibility without having to create millions of hard programmed pieces of silicon. They are essentially silicon chips that can be programmed after the manufacturing process. They do cost more than the equivalent purpose built silicon and may not perform as well as specifically designed chips but for smaller vendors they are a viable compromise.

Another option, especially for PCIe SSD vendors, is to use host resources. Companies like Virident are leveraging the CPU and memory of the server in which they're installed in conjunction with FPGAs. While this does consume a small amount of CPU and RAM resources it can yield very high performance since it provides a scalable, ever-growing engine and allows for rapid development of feature rich code.

What is Conditional Performance?

Conditional performance means that the only way to achieve maximum or better than average, solid state storage performance is to make sure that the environment is favorable to the type of solid state technology being used. The first condition is the requirement that a certain amount of free space be available to perform flash management functions like garbage collection and bad block replacement. While most solid state vendors reserve a certain percentage of capacity for these processes the performance of those functions will degrade under certain conditions.

The Condition of Freshness

Another condition is a requirement that the drive be completely written to one time. In other words it is no longer "fresh out of the box". While there may be plenty of free space on the drive all the cells have been written to at least once, meaning that each additional write needs to first trigger a program/erase cycle. Again the reserve area is there to sustain performance by pre-clearing old cells but unless the flash management scheduler is efficient, performance is often impacted.

The Condition of Fullness

Performance can also be impacted by the drive being full or near full of active data. This is a common reality for flash, especially when it's used for a cache storage area since most cache algorithms will try to keep the high-speed storage area as full as possible. Full utilization of SSD is a wise investment for premium priced storage but it can stress the garbage collection processes, which needs free space to operate efficiently.

In this situation sheer processing power is required to quickly reclaim written cells since there may be a limited number of pre-cleaned cells. Again, in a cache environment a sudden burst of write activity caused by a data set becoming active may be the event that pushes the garbage collection to its maximum rate. This can in turn trigger a hesitation in solid state performance which impacts application response time causing storage managers and application owners to develop workarounds for the device, like provisioning less of the available flash on the solid state device.

The Condition of Bad Cells

Another condition that impacts performance is in the area of bad flash cell management. Most solid state storage devices have the ability to identify degrading or bad cells and map those out so user data is not lost. These bad cells are replaced with good cells from the reserve area again via mapping. There are two problems with this technique that become apparent over time. First, the more bad cells there are the more complex the remapping process becomes and, again, raw horsepower is needed to help the I/O request find the right location. Second, as new cells are mapped out of the reserve area less of that area is available for garbage collection, making the importance of garbage collection processing power more critical.

The Condition of Bad Cells

Another condition that impacts performance is in the area of bad flash cell management. Most solid state storage devices have the ability to identify degrading or bad cells and map those out so user data is not lost. These bad cells are replaced with good cells from the reserve area again via mapping. There are two problems with this technique that become apparent over time. First, the more bad cells there are the more complex the remapping process becomes and, again, raw horsepower is needed to help the I/O request find the right location. Second, as new cells are mapped out of the reserve area less of that area is available for garbage collection, making the importance of garbage collection processing power more critical.

The Application Condition

From the application side many solid state devices tend to be good at some types of I/O and not others. For example peak I/O with large blocks will perform entirely differently on some solid state devices than the random I/O common in databases and virtual environments. This means that customers will have to either modify applications or buy specific PCIe SSD boards for different use cases.

The application and server hardware changes are typically one-time events that have to be worked through. What will change however is the type of application or the nature of the same application that’s being used with the solid state technology. No data center is static. More concerning may be the other conditions like flash management and garbage collection issues that may not show themselves in the initial evaluation cycle, instead appearing well after the evaluation has ended and the product is moved to introduction. It is important to look for solid state storage devices that can handle a mixed variety of workloads as well as provide performance that won’t change as the product is used for a longer period of time.

Dealing with Conditional Performance Issues

The most common workaround for conditional solid state storage performance is to reserve even more of the solid state capacity than what the vendor sets aside for the reserve area. This is comparable to short stroking mechanical hard drives, in that a set-aside is created so that garbage collection and bad block management have more room in which to operate.

Providing Unconditional Performance

The key with any solid state storage device, but especially one that is being offered as an alternative to expensive DRAM, is to make sure that a consistently high level of performance can be maintained over the entire life of the device and under any I/O circumstance.

To do this requires that multiple controllers be virtualized in the same way that a compute cluster of processors is virtualized to act as one entity on the problem at hand. The problem, in this case, is making sure that write I/O is consistently handled and that it doesn’t interfere with overall application performance. Having multiple flash controllers that are virtualized and managed by a single master controller handles this problem very efficiently by allowing operations to be spread out amongst multiple individual controllers, while maintaining the global view of the flash media to ensure even wear leveling across all of the media on the device. The result is consistent write performance, even under high write conditions and consistent read performance under all conditions.

Summary

There are too many “what if” scenarios when implementing solid state storage, something which has been a big inhibitor to its adoption in the enterprise. IT professionals don’t have the time or desire to carefully manage I/O patterns and they certainly don’t want to under provision something that is already priced at a premium. They simply want a device that delivers the same performance on day one as it does on day 300. Companies like Virident are delivering this capability be leveraging virtualization to integrate multiple controllers on a single PCIe SSD to deliver that reality.