Data protection in two dimensions

IBM FlashSystem with IBM Variable Stripe RAID delivers two dimensions of data protection

Highlights

- Add a new dimension of data protection with IBM® Variable Stripe RAID™

- Automatically address flash failures with no loss of data or storage capacity

- Eliminate the data protection, performance and resource utilization inefficiencies caused by commodity solid-state drives (SSDs)

- Increase the granularity of flash monitoring, management and provisioning down to the individual flash block level and below

- Maximize utilization of flash storage resources

Most enterprise all-flash storage arrays available in the marketplace today are designed using proprietary software integrated with essentially commodity hardware. This architecture is intended to reduce both the cost of the product and its time to market. A prime example of this approach is the use of commodity SSDs as the flash storage medium. To provide enterprise-grade data protection, these products implement various RAID regimes across the SSDs to reduce the chance of data loss from a flash failure.

A key liability of this system-level RAID approach to data protection is that when a single flash media failure is detected in an individual SSD, the entire SSD must be taken offline and replaced before robust data protection is re-established. This leads to higher maintenance costs, lower system availability and times of greater vulnerability for business-critical applications.

The IBM FlashSystem® family of industry-leading all-flash storage arrays takes a different approach. Instead of commodity hardware and SSDs, IBM FlashSystem leverages the purpose-engineered advantages of IBM FlashCore™ technology to provide two independent dimensions of data protection. System-level RAID is implemented across all the IBM MicroLatency® modules in every IBM FlashSystem array.

In addition, patented Variable Stripe RAID is implemented within each MicroLatency module. Variable Stripe RAID monitors the health of the flash media, detects and manages flash failures very efficiently and optimizes utilization of all flash resources.

Two-dimensional RAID in IBM FlashSystem

IBM FlashCore technology is the DNA of all IBM FlashSystem family members. The term "FlashCore" refers to the IBM innovations that enable IBM FlashSystem storage to deliver extreme performance, ultra-low latency, enterprise-grade reliability, and a wide range of operational and cost efficiencies. These technologies and innovations are realized in the IBM FlashSystem hardware-accelerated architecture, MicroLatency modules, and many other advanced flash management features and capabilities. IBM FlashCore technology truly defines IBM FlashSystem storage and differentiates it from competing all-flash systems.

IBM FlashSystem 900

One of the many advantages IBM FlashCore technology brings to IBM FlashSystem storage customers is the implementation of two separate levels of data protection, one very granular level or dimension at the MicroLatency module level, and the other at the system-wide level across all the MicroLatency modules. IBM calls this innovation "two-dimensional RAID." It is designed to protect data from both types of failures that occur within the flash storage medium—failures that are expected to happen and those that are not. The first type of error is handled automatically by the flash controller using Variable Stripe RAID. The process isolates these failures, reduces them to the absolute minimum affected area, and then bypasses them.

Flash failures addressed in this way are described as "self-healing." Within the individual MicroLatency module, the failed area of flash media is handled by Variable Stripe RAID and the array keeps running without a loss in performance or data protection. With Variable Stripe RAID, a flash failure does not result in the need for a maintenance event. No flash module replacement is necessary. The IBM FlashSystem data protection regime is not degraded in any way. And there is no need for a lengthy hard disk drive rebuild. The vast majority of flash failures are detected and automatically addressed by the Variable Stripe RAID technology at the individual MicroLatency module level with no effect to system performance or behavior.

It is possible, however, for a second type of error in which the system experiences a failure of the entire MicroLatency module. This is the unexpected type of failure. These failures can't be self-healed. Thanks to the two dimensions of data protection provided by IBM FlashCore technology, this very unexpected type of failure is also easily and automatically addressed.

A failure of this type triggers existing system thresholds that fail the entire module immediately and automatically engage the second RAID dimension, system-level RAID 5. This data protection technology will rebuild the RAID regime using the system's online spare MicroLatency module and send out a maintenance alert so the failed module can be replaced.

The first dimension: Variable Stripe RAID

Expected flash failures result from the fact that NAND flash memory is susceptible to failure modes that are uniquely associated with the storage media technology. Many common flash failure modes come from localized issues that take only a single flash plane or die offline, leaving the remainder of the chip fully functional. These failures include block wear-out, disturb errors, plane failures and even whole chip failures. IBM does not use DIMMS or SSDs; the individual flash chips are integrated directly into the MicroLatency module design. There are multiple flash controllers (field-programmable gate arrays [FPGAs]) per module (either one, two or four depending on the module capacity), with each controller being responsible for a set group of flash chips (anywhere from 13 to 20 flash chips, depending on module size and system model). This design enables a higher degree of parallelism for maintaining very consistent response times over ever-increasing workloads.

IBM FlashSystem arrays have extra flash capacity beyond the standard amount of overprovisioned capacity. Overprovisioning flash capacity is not novel. All enterprise-grade flash products have a certain amount of extra or overprovisioned flash storage capacity beyond what the vendor advertises as usable storage.

This extra flash capacity is used for routine flash management functions such as garbage collection and enhancing write performance. Even SSDs have some extra capacity for use in this way. But SSDs handle the flash management tasks internally. This means that the results and effectiveness of the SSD flash management are mostly obscured from the system using the SSD. In addition to using overprovisioning for garbage collection, wear leveling and write performance, IBM FlashCore technology with Variable Stripe RAID specifically leverages the extra capacity to mitigate the expected flash failures. When a failure occurs, it will knock out some portion of the storage capacity, but thanks to the remarkable capabilities of Variable Stripe RAID, the unusable capacity involved in the failure will be automatically replaced with some of the extra flash capacity waiting for just this scenario.

Variable Stripe RAID does not simply create an array effect across flash chips the way standard RAID does across disks (though this would still be an improvement over almost all SSDs). Instead, Variable Stripe RAID was designed specifically for flash. A flash chip is composed of many sub-components. Each chip has multiple die. Each die is made from multiple planes. Each plane has multiple blocks. Each block is divided into pages. Variable Stripe RAID stripes data across flash planes. This level of granularity gives IBM FlashSystem MicroLatency modules the ability to survive sub-chip failures without impacting usable capacity or system performance. This is important because such sub-chip failures are not rare occurrences.

When a plane fails, Variable Stripe RAID technology allows that plane to be removed from use without impacting the available capacity of other flash devices within the RAID stripe. Upon detection of a failure, Variable Stripe RAID removes the failing plane from use (no further writes are allowed) and all used pages within the affected stripe are marked as "critical to move." Valid data from the affected stripe is then gradually relocated in the background to known good stripes. Once the data has been safely relocated, the failed plane is removed from the original stripe, allowing the stripe to be used for new data. While the failed plane cannot be reused, the remaining good planes are retained, and additional data can be striped across them. Furthermore, the new, smaller stripe can still survive another plane failure through the same process.

The gradual losses introduced by these common, expected flash failures eat into the IBM FlashSystem overprovisioned capacity. IBM FlashCore technology monitors and reports on the amount of overprovisioned capacity available and reports this status as module board health stated as a percentage from 100 percent (perfectly healthy) to 0 percent (any further failure will result in loss of usable capacity). IBM has set thresholds in this board health that trigger a transition between three different states: GOOD, where health is acceptable with little or no chance of immediate failure; WARNING, where the system issues notifications of a worsening state via email, SNMP traps, GUI and call home; and CRITICAL, where the system fails the module and initiates system-level protections.

The second dimension: System-level RAID
Unexpected flash failures are extremely unlikely and therefore don't warrant the extra expense of attempting to address them via self-healing mechanisms. These include failures of integrated circuit components such as the FPGA implementing the flash controller, the embedded processor handling long-term monitoring and management, or the handful of other components on each MicroLatency module board such as resistor and capacitor packs. These components all have extremely high mean time between failure rates.

These types of issues are no different from those faced by any IBM FlashSystem competitor. If the same type of component fails within one of their SSDs, the competing system doesn't even have the information about what exactly failed. Instead, that entire SSD would simply go dead and the competitor's system would need to engage its system-level protection. In the case of IBM FlashSystem, the system-level RAID 5 kicks in and the system automatically detects the failure and immediately begins to restripe data across the hot-spare module in the system (as opposed to competing systems that often lack hot-spare flash modules).

The IBM difference
The most important difference between competing systems and IBM FlashSystem is that IBM provides Variable Stripe RAID, resulting in two dimensions of data protection. Most other enterprise all-flash array vendors do not have an equivalent technology. In those competing products, all components work until the smallest of errors, at which point none of them work.

By purpose-engineering IBM FlashSystem MicroLatency modules and flash controllers instead of using commodity components, IBM creates granular insight into exactly what the flash is doing all the way down to the individual cell level. This enables IBM FlashSystem arrays to handle the many different types of all-too-common errors inline, isolate the failures and simply map around them to keep running. New developments in IBM FlashCore technology even allow IBM systems to monitor and predict upcoming failures and dynamically adjust flash programming voltage thresholds to minimize the likelihood of uncorrectable errors. This can significantly extend the time before such errors occur.

IBM systems also monitor the health of each individual flash block and have the ability to place the most frequently accessed data on the healthiest blocks to provide adaptive, self-optimizing wear leveling. Most competing products using SSDs do not have this capability. They are at the mercy of the SSD vendor. SSDs are commodity and designed as such. They feature the most rudimentary bad block management and data protection, so all failures result in failure of the entire disk/SSD.

Most other all-flash array vendors provide only one dimension of data protection. IBM FlashSystem provides two. Other vendors use commodity hardware such as SSDs. IBM purpose-engineers every 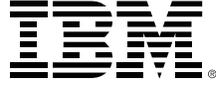component of IBM FlashSystem, from the chip to the front bezel. Variable Stripe RAID within IBM FlashCore technology monitors flash health with a granularity unmatched by competing systems, and leverages this detailed information about real-time flash health to increase storage efficiency and endurance and ultimately lower costs.

The IBM difference is clear. Variable Stripe RAID and the two-dimensional data protection provided by IBM FlashCore technology is the level of protection your applications and data should not operate without.

For more information

To learn more about the IBM FlashSystem family, please contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/systems/storage/flash/

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

IBM