

The HC3[®] Storage Advantage



When it comes to data access for users, faster is better for productivity. Many factors affect how efficiently the data is delivered to users including networking, processing, and maybe the most important, storage architecture efficiency. Many storage architectures are complex, layering multiple storage protocols along with virtual appliances simply to present block storage to virtual machines. Scale Computing has created a storage architecture within HC3 to deliver high efficiency along with streamlined simplicity.

SCRIBE

The Scale Computing Reliable Independent Block Engine is the core component of the HC3 system, combining the storage drives from each HC3 node into a single, logical, system-wide storage pool. The pooling occurs automatically with no user configuration required. Blocks are stored redundantly across the system to allow for the loss of individual drives or an entire system node.

The SCRIBE storage pool is available to all nodes of the system and presented without any file systems, protocols, or virtual appliances. SCRIBE is embedded directly in the HC3 HyperCore operating system. When a VM is created on HC3, the virtual disks provide direct block access between the virtual machines and the SCRIBE storage pool. The only file systems created in HC3 are the file systems used by the guest operating systems in the VMs to address the virtual disks.

Other storage architectures tend to emulate SAN or NAS devices that were traditionally used in virtualization. These start with a storage pool on the lowest level with a file system layered on, then presented to the hypervisor where another file system is layer on and managed by a virtual storage appliance (VSA), and finally presented to a VM where another file system is layer on. Aside from the multiple levels of protocols that must be traversed for each I/O operation, the VSAs managing the storage can consume a large portion of the RAM and CPU that would otherwise be used for creating more VMs.

As a comparison, the route of an I/O operation in a VSA architecture may look something like:

Application -> RAM -> Hypervisor -> RAM -> VSA -> RAM -> Hypervisor -> RAM -> Write-cache SSD -> Erasure Code(SW R5/6) -> Disk -> Network to next node -> RAM -> Hypervisor -> RAM -> VSA -> RAM -> Hypervisor -> RAM -> Write-cache SSD -> Erasure code (SW R5/6) -> Disk

On HC3, the same I/O operation route would look more like:

Application -> RAM -> Disk -> RAM -> Network to next node -> RAM -> Disk

The example above includes an SSD cache as part of the VSA architecture. SCRIBE can also incorporate SSD but instead of using it merely as a cache, it is used as a storage tier within the SCRIBE storage pool. As a cache, the higher speed of SSD helps mask the inefficient design in VSA architecture. In HC3, as a storage tier, SSD is used for data storage, increasing both the size and the overall speed of the storage pool access. With SCRIBE, a hybrid storage architecture allows for data to be moved dynamically between SSD and HDD tiers using HyperCore Enhanced Automated Tiering (HEAT) technology.

HEAT

Data sets are most commonly characterized by only a small percentage of data being actively accessed. In a hybrid, tiered storage system, the most benefit to efficiency comes from the active data residing on SSD and the inactive data being stored on slower HDD storage. The HEAT technology in SCRIBE monitors data access and creates a dynamic mapping of active data blocks and moves these blocks to the SSD storage tier while moving inactive storage blocks off of SSD onto HDD.

This movement of data blocks between tiers is transparent to users and even administrators, happening automatically. By default, each disk is actively trying to achieve an equal level of efficiency by moving data blocks between SSD and HDD tiers. Sometimes, one disk requires an even greater level of efficiency than another disk; in these cases HC3 administrators can give those virtual disks priority.

For each virtual disk, the relative priority of SSD utilization can be adjusted on a scale from 0-11. The default for all disks is 4 and if no disk is ever changed, no disk has priority over any other. Any disk can be dynamically adjusted in priority, up or down, to increase or decrease the SSD utilization on that disk. Priority 0 will completely bypass SSD and only utilize HDD for that disk. Priority 11 will attempt to put the data for that disk completely on SSD, or at least as possible given the available SSD on the system, taking into account all other disk priorities.



The screenshot shows a dialog box titled "Update Block Device" with a close button (X) in the top right corner. Inside the dialog, there is a "Type" dropdown menu currently showing "VIRTIO". Below it is a "Flash Priority" slider ranging from 0 to 11, with a blue circle indicating the current value is 11. Underneath the slider, the text "These go to 11" is visible. At the bottom of the dialog are two buttons: "Update" and "Delete".

HC3 has built-in, real-time IOPS meters for each virtual disk so the results of changing the priority on a particular disk can begin to be seen immediately. Because the virtual disks can be changed dynamically, the priority can be adjusted as much as necessary to achieve the correct balance of IOPS efficiency across virtual disks and VMs.

Scaling Out

SCRIBE and HEAT are both designed to automatically incorporate new storage from additional HC3 nodes in a system. When an HC3 system is scaled out with new nodes, the storage from the new nodes is seamlessly added to the SCRIBE pool and immediately visible and usable from every node in the system. There is no storage configuration required by the administrator, only the option to adjust SSD priority as needed.

The Advantage

By design, embedding the storage system in the hypervisor allows HC3 to have extreme efficiency with absolute simplicity. With HC3, users experience the benefits of SSD as a tier, and have the edge of clustered redundancy that absorbs the loss of failed disks or entire nodes.

This is the advantage of HC3. A storage system that is designed specifically to work with modern virtualization. None of the complexity of SAN and NAS that was designed to work with physical servers. Efficiency, convergence, and user experience have been simplified in one system.



Corporate Headquarters
525 S. Meridian Street
Indianapolis, IN 46225
P. +1 317-856-9959

www.scalecomputing.com

West Coast Office
360 Ritch Street
Suite 300
San Francisco, CA 94107

1-877-SCALE-59 (877-722-5359)

EMEA Office
Saunders House
52-53 The Mall
London
W5 3TA
United Kingdom