HyperCore Enhanced Automated Tiering (HEAT)

Feature Note

This document provides information around the native HEAT feature built into the HyperCore HC3 system. Check for additional information on this topic in the support Knowledge base available in the Scale Computing Portal here: http://www.scalecomputing.com/support/login/.

03/2018

You can find the most up-to-date technical documentation on the Scale Computing Portal at:
http://www.scalecomputing.com/support/login/

The Scale Computing site also provides the latest product updates and information:
http://www.scalecomputing.com/

Provide feedback on this document at:
documentation@scalecomputing.com

Contact support at:
support@scalecomputing.com
http://www.scalecomputing.com/support/login/
US: +1-877-SCALE-59 (877-722-5359)
Europe: 0808 234 0699

Document Version 2.1: 03/2018

Corporate Headquarters
The Union 525
525 South Meridian Street
Indianapolis, IN 46225
P. +1 877-722-5359

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

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

www.scalecomputing.com

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

Contents

Introduction

This guide provides information on HyperCore Enhanced Automated Tiering (HEAT) as a part of the SCRIBE (Scale Computing Reliable Independent Block Engine) storage layer and HEAT's use within Scale Computing HC3 systems.

HEAT is the intelligent placement and management of data on the SSD and HDD tiers in HC3 nodes. HEAT includes features such as configurable SSD priority allocation at the individual VM virtual disk level through an easy-to-use slide bar in the HC3 web interface, and intelligent data block priority based on block I/O heat mapping assessed utilizing I/O information on each virtual disk.

Requirements

In order to utilize HEAT in the HC3 system, there are a few criteria to meet:

- In a clustered HC3 system (3+ nodes in one cluster), there must be a minimum of 3 Tiered HC3 nodes (nodes that utilize SSD drives); see the latest HC3 support matrix for supported configurations.
- 1 GbE or 10 GbE network capabilities (10 GbE is strongly recommended)
- System software version is 7.0 and up

HEAT Overview

An overview of key features and principles of the HEAT feature set.

Core Principles

HEAT is an extension of the SCRIBE storage layer and as such is designed to be fast, flexible, and simple—the core design principles of the HC3 system. However, making something simple should never mean making something less, which is where HEAT and SCRIBE shine.

As HEAT was designed as an extension of SCRIBE, it uses the existing SCRIBE and HyperCore architecture as the groundwork for simplicity and high availability, shown in the four HEAT principles below.

1. All VMs have access to SSDs, no matter what node the VM may actually be running on.
2. SSDs are additional capacity for VM disks (**subvirtual tiering**), not a cache for system data.
3. Administrators have granular control of SSD access at the VM virtual disk level.
4. Administrators are able to mix and match Tiered HC3 nodes with standard HC3 nodes and Storage Only nodes without any extra work or requirements.

Integration

Utilizing the HyperCore hypervisor, the SCRIBE data management layer, and the system state machine, HEAT is compatible with existing infrastructure as it is scaled out to include Tiered HC3 nodes (nodes utilizing SSDs) and supporting firmwares.

All that is needed is a minimum of three Tiered HC3 nodes and the minimum 7.0 HCOS software version. A 10 GbE network base is strongly recommended for the proper performance on the Tiered HC3 nodes, although it is possible to use 1 GbE in certain configuration scenarios. Having a minimum of three Tiered HC3 nodes ensures that there is always a redundant place to write SSD data copies for the system wide striping, even while a Tiered HC3 node may be unavailable (during a HyperCore rolling software update, for example).

HEAT is a key part of SCRIBE for SSD functionality. Utilizing SCRIBE, HEAT allows virtual disk level, priority

data placement for allocated data (actual consumed capacity on a VM virtual disk). This is accomplished through a real-time heat map of virtual disk I/O in order to "tier" the data. Data blocks that are "hot"—accessed regularly by the virtual disk—are stored at the SSD level while "colder" data blocks are stored at the HDD level. This heat mapping capability combined with the intelligence of the HyperCore hypervisor and the real-time monitoring capabilities of the state machine offer a highly efficient, reliable, and capable data tiering design. See the **HEAT Data Management** section for more information on how HEAT works within SCRIBE to place data and the impact HEAT may have on VM and system performance.

Intelligence

HEAT utilizes custom algorithms and a heat map of data within SCRIBE to weight data placement between SSD and HDD tiers in the HC3 system. These algorithms utilize the assigned VM virtual disk priority from the HC3 web interface, the allocated (actual consumed) VM virtual disk capacity, and the overall virtual disk pool priority and capacity in conjunction with the I/O heat mapping to determine data placement.

It is easy to adjust priority placement for VM virtual disk data utilizing the slide bar for the particular VM virtual disk in the HC3 web interface. Even one number's adjustment can bring significant changes in data placement, making it faster and easier to customize VM settings at the virtual disk level. If you have a heavily utilized database drive and a more static, file serving drive all on one VM, simply set the slide bar appropriately for SSD tier access independently on each virtual disk and the HyperCore hypervisor and SCRIBE data management layer will handle the rest.

WARNING

Review the HEAT Best Practices section before any large data transfers to the HC3 system or before altering any of the HEAT slide bars in order to maintain proper VM performance and stability during heavy I/O transitions.

HEAT Data Management

HEAT is a part of the SCRIBE data management layer and works together with the HyperCore hypervisor to provide SSD data management features. Utilizing the SCRIBE and HyperCore software, HEAT crafts a heat map of blocks for each individual VM virtual disk within SCRIBE that can be used to determine a block's tier level (SSD or HDD).

HEAT and SCRIBE will set the overall priority each VM virtual disk should be given to the SSD tier from the configured setting in the HC3 Web Interface and the VM virtual disk's own heat map when determining tiering. HEAT will also work with SCRIBE to ensure that all data written to the system, whether on SSD or HDD, is stored safely and redundantly using SCRIBE's system wide striping.

HEAT Mapping

HEAT utilizes a heat map of blocks within SCRIBE to record I/O on a VM virtual disk. Using this heat map and the virtual disk's set HEAT priority level, individual data blocks can be properly assigned to either the SSD or HDD tier on the HC3 system inside of the virtual disk depending on their activity.

Heat maps utilize I/O iterations to determine the tiering information for a given data block on a VM virtual disk. This ensures that data tiering is fluid when necessary. As access changes for a given VM virtual disk—user access dwindles or grows depending on company operations over the course of the fiscal year, for example—I/O is always tiered appropriately.

More information on data priority and striping can be found in the **Priority and Data Striping** section below. See the **SSD Data Overpopulation** section for best practices.

Priority and Data Striping

The HC3 web interface provides an easy-to-use slide bar in order to set the priority of a VM's virtual disk data for HEAT, explained in the **Customizing HEAT Priority Level** section.

The slide bar in the HC3 web interface operates at the individual virtual disk level for each VM, and provides an easy way to assign priority to a VM virtual disk for SSD access in HEAT. The **Flash Priority** in the HC3 web interface is calculated with the heat map of all of the virtual disks in the HC3 system to determine the overall tiering of all data blocks in the system. Please note that larger and/or more active allocated drives will have higher priority access to the SSD tier when all drives have equal access.

Find more details around best practices for configuring HEAT **Flash Priority** in the HC3 web interface in the **SSD Data Overpopulation** section.

As HEAT is an extension of SCRIBE, data blocks assigned to the SSD tier will be wide striped for data redundancy within the SSD tier, and data blocks assigned to the HDD tier will be wide striped for data redundancy within the HDD tier.

If there are ever any questions on the status of a VM virtual disk's access to the SSD tier, please contact ScaleCare Support to open a case to discuss virtual disk performance and data placement in detail.

Customizing HEAT Priority Level

It is easy to customize HEAT access for VM virtual disks through the HC3 web interface.

Below is a chart of the available number settings and their corresponding descriptions as a measure of relative "I/O speed"—the SSD priority that will be assigned for any data placed on the VM virtual disk, ranging from 0 to 11. By default, new virtual disks will be set to a HEAT access level of 4.

HEAT Priority Level	HC3 Web Interface Description
0	Off
1	Minimum
2	Very Low
3	Low
4 (Default)	Normal
5	High
6	Very High
7	Extreme
8	Absurd
9	Hyperspeed
10	Ludicrous Speed
11	These go to 11

Each step up from one through 10 doubles the virtual disk's priority for the SSD tier. 0 and 11 are special cases. 0 turns off all SSD access for the VM virtual disk, and 11 multiplies the priority by 10. This makes it easy to customize virtual disk settings without spending hours or even days "dialing in" on the perfect setting.

WARNING

Make HEAT priority changes slowly and with caution; altering HEAT priority will effect ALL VM virtual disks on the system. Each increase in flash priority access will dedicate roughly TWICE as much flash capacity for the VM virtual disk—and consequently reduce the flash capacity available for other VM virtual disks on the system.

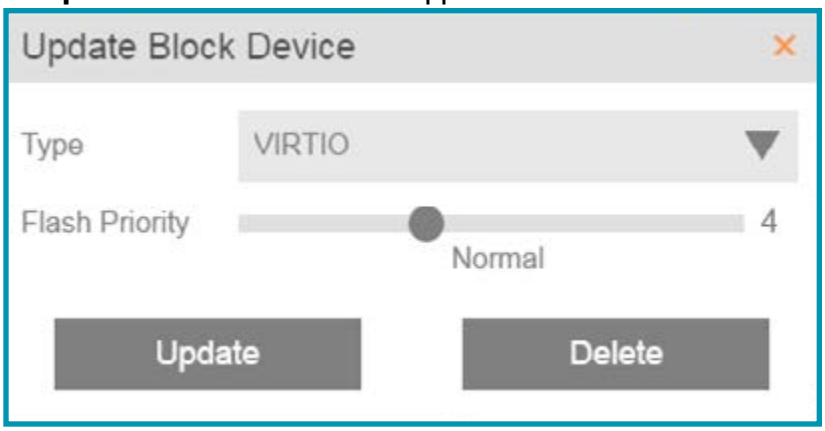
Changes to HEAT priority will trigger a large redistribution of data between the spinning disk and flash disk tiers that may detrimentally impact system and VM performance for the duration of the operation. Changing more than one virtual disk priority level at a time can extend the performance impact of the redistribution operation. Due to the performance impact of data redistribution, any performance assessment will be invalid until the operation is complete.

We recommend moving the slider of one VM virtual disk one level up or down once every 24 hours. Only a single virtual disk's flash priority level should be changed in a 24 hour window. After waiting 24 hours for the data to be redistributed, check performance and repeat as necessary using the same "one level every 24 hours" rule of thumb.

If you are experiencing performance problems, or if you have questions about your VM virtual disk heat maps, please contact ScaleCare Support to review your HC3 environment in detail.

In order to make a change to the HEAT priority for the VM virtual disk, follow the steps below. The HEAT priority adjustment can be made at any time, whether the VM is powered on or powered off. Utilize the **HyperCore User Guide** under **Resources** if any assistance is needed navigating the HC3 web interface.

1. Click the **edit** button in the **Devices** screen of the VM card next to the VM virtual disk the priority adjustment is required on.
2. A dialog box titled **Update Block Device** will appear.



3. Use the **Flash Priority** slide bar to adjust the SSD access for the virtual disk to the desired setting.
4. Click **Update** to confirm the change.
5. The new HEAT priority setting will be immediately activated on the VM virtual disk.

HEAT Best Practices

The following sections detail HEAT recommendations and best practices for the best experience with the HEAT feature, and to further a working knowledge of the HEAT capabilities.

SSD Data Overpopulation

Due to the limited amount of capacity the SSD tier provides in comparison to the overall system capacity, it is expected that the SSD tier will be in high demand in most environments (at the very least the SSD space will be in demand initially on the system until more regular VM virtual disk patterns are established by HEAT and SCRIBE). The concept of "SSD overpopulation" is HEAT's way within the SCRIBE storage layer of dealing with this highly demanded but limited space.

Utilizing the VM virtual disk heat map, data will be relocated between SSD and HDD storage if it is determined that a more active, higher priority drive is in need of the SSD capacity. As mentioned, this is most common in newly installed systems as data blocks are still being prioritized and assessed as patterns are determined. This can happen in a variety of other scenarios, however, some of which may be: a VM virtual disk priority being increased or decreased, data traffic patterns changing as users are added or removed, quarterly or yearly tasks being in demand, large data migrations to or from the HC3 system, or workloads shifting as VM virtual drives are combined, expanded, or deleted.

The concept of overpopulation allows HEAT to track assigned data blocks through SCRIBE in order to assess the "population" of the SSD tier. When it is determined that any given block of data has fallen below or above the population threshold, a change is triggered to allow data flow between the SSD and HDD tiers as needed. This actively provides an intelligent performance experience for any VM virtual disk with SSD access.

Understanding how overpopulation works is important. If a change is made to a VM virtual disk's priority designation for the SSD tier, it is not an immediate change. The data heat map for the VM virtual disk must be assessed by HEAT for the appropriate changes between the SSD and HDD tier, and then the data must be distributed onto the new tier.

WARNING

Make HEAT priority changes slowly and with caution; altering HEAT priority will effect ALL VM virtual disks on the system. Each increase in flash priority access will dedicate roughly TWICE as much flash capacity for the VM virtual disk—and consequently reduce the flash capacity available for other VM virtual disks on the system.

Changes to HEAT priority will trigger a large redistribution of data between the spinning disk and flash disk tiers that may detrimentally impact system and VM performance for the duration of the operation. Changing more than one virtual disk priority level at a time can extend the performance impact of the redistribution operation. Due to the performance impact of data redistribution, any performance assessment will be invalid until the operation is complete.

We recommend moving the slider of one VM virtual disk one level up or down once every 24 hours. Only a single virtual disk's flash priority level should be changed in a 24 hour window. After waiting 24 hours for the data to be redistributed, check performance and repeat as necessary using the same "one level every 24 hours" rule of thumb.

Virtual Machine Cloning and Replication

It is possible to replicate VMs between a system containing Tiered HC3 nodes (nodes containing SSDs) and a system that does not have any Tiered HC3 nodes (only HC3 nodes). It should be noted that a VM restored on a system with HC3 nodes only and that has been replicated from a system with Tiered HC3 nodes will not have access to the customized performance facilitated by HEAT, SCRIBE, and the SSD tier, and may not perform as smoothly as on a system with HEAT capabilities.

When using HC3 replication between two systems with Tiered HC3 nodes, HEAT information is not passed forward to the replica VM on the target system. This means that any **Flash Priority** setting configured for the VM virtual disks will not carry over when a VM is cloned on the target system.

NOTE: It will be necessary to reconfigure Flash Priority settings on a restored VM's virtual disks. VM virtual disks restored or cloned from a snapshot image will revert to the default 4 HEAT level.

In the same way, clone VMs do not possess the HEAT information of the parent VM. When cloning a VM or restoring a VM on a target system, HEAT information is assessed independently for the cloned or restored VM.

For more information on replication, cloning, and capacity management on the HyperCore HC3 system, see the **Resources** section.

HEAT and System Scaling

As discussed in the HEAT Overview section, a core design feature of HEAT within SCRIBE was the ability to support both Tiered HC3 nodes and standard HC3 nodes together in one system. In order to incorporate Tiered HC3 nodes into an existing HyperCore HC3 system, check the following requirements.

- **Tiering Specific** - There must be a minimum of three Tiered HC3 nodes in a system for proper support and failover at the SSD tier.
- **Networking** - Different Tiered HC3 node lines offer both 1 GbE and 10 GbE networking in order to support environmental needs. Tiered HC3 nodes perform optimally with 10 GbE. This should be taken into consideration for the networking environment. Contact ScaleCare Support for further details on combining nodes.
- **General** - HC3 systems are currently supported up to 8 nodes total. If more nodes are required please contact ScaleCare Support to discuss expansion options.

Resources

Additional informational and technical resources from Scale. Some documents may require a **Customer Portal** or **Partner Portal** login.

- HC3, SCRIBE, and HyperCore Theory of Operations
- HyperCore HC3 Capacity, Clone, and Snapshot Management Feature Note
- HyperCore HC3 to HC3 Native Replication Feature Note
- Latest Scale Computing User Guide

Feedback and Support

Document Feedback

Scale Computing welcomes your suggestions for improving our documentation. If you have any comments or suggestions please send your feedback to documentation@scalecomputing.com.

Technical Support and Resources

There are many technical support resources available for use. Access this document, and many others, at http://www.scalecomputing.com/support/login/.

Online Community Scale Computing has an online forum and community! This is a great medium with which to solicit the advice of your peers, benefit from their experience, find and discuss documentation, and participate in ongoing conversations. Please note this community is not intended to provide ScaleCare Support assistance or replace other Scale Computing communication channels. Find the **Scale Legion HC3 Discussion Forum** at https://scalelegion.community

Online Support You can submit support cases and view account information online through the Portal at http://www.scalecomputing.com/support/login/. You can also Live Chat with support through www.scalecomputing.com during standard hours Monday-Friday 8-8.

Telephone Support Support is available for critical issues 24/7 by phone at 1-877-SCALE-59 (877-722-5359) in the US and at 0808 234 0699 in Europe. Telephone support is recommended for the fastest response on priority issues, and the only response after standard Support hours.

Professional Resources Scale Computing offers many professional service options for both remote and on-site assistance in getting your new system off the ground quickly and knowledgeably. Contact your Scale Computing sales representative today to discuss our offerings that include options like the HC3 Premium Installation.

Disclaimer

Any information listed here is not a substitute for the product's user guide or support and is not covered under the support and warranty for the HC3 system. Scale Computing is not responsible for any issues or damages arising out of the use of this reference guide.